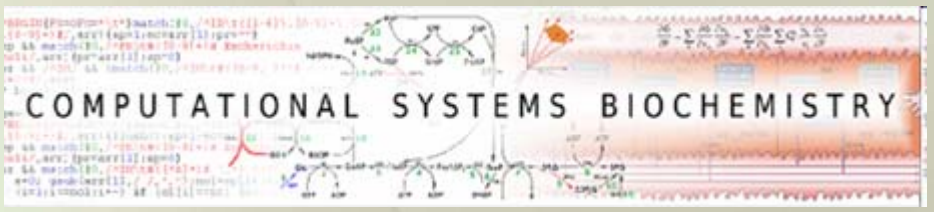
virtual liver network

CHARITÉ
UNIVERSITÄTSMEDIZIN BERLIN

On the adequate use of microarray data

About semi-quantitative analysis

Andreas Hoppe, Charité Universtitätsmedizin Berlin
Computational systems biochemistry group



Contents

- Introduction
- Gene array accuracy
- What is semi-quantitative?
- Guidelines for semi-quantitative analysis
- Microarray preparation in ModeScore

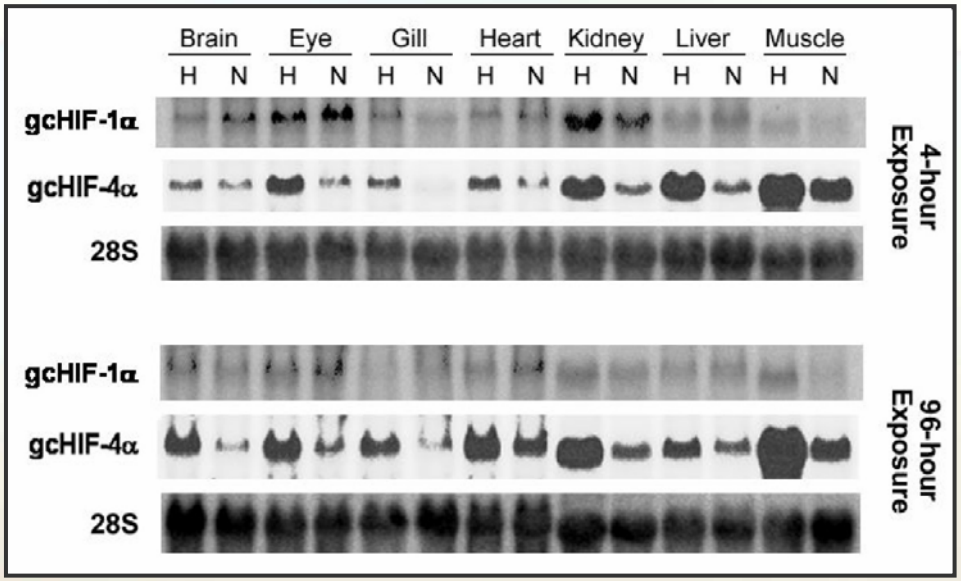
Contents

- Introduction
- Gene array accuracy
- What is semi-quantitative?
- Guidelines for semi-quantitative analysis
- Microarray preparation in ModeScore

Introduction

- Matthias König: Microarrays are
  - adequate to detect the direction of change
  - not adequate to quantify absolute differences in RNA
  - not adequate to compare between genes

Introducing example

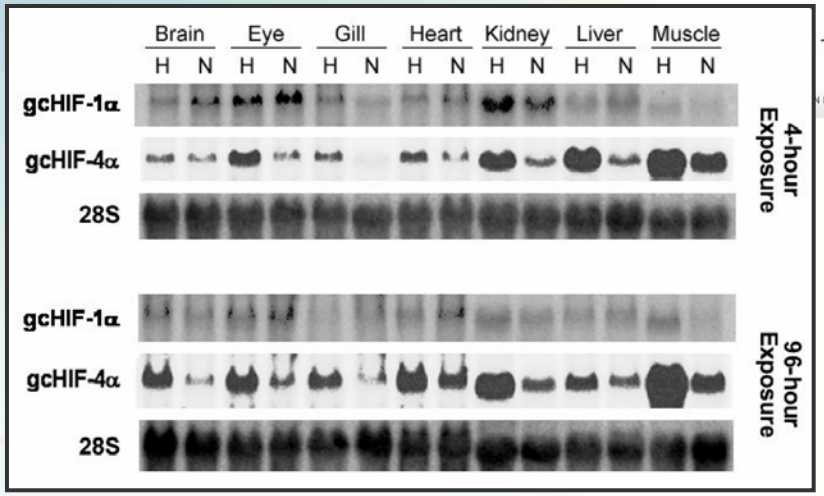


Northern blot analysis of gcHIF-1α and gcHIF-4α

Law et. al, 2006, *BMC Molecular Biology* 2006, 7:15.

virtual liver network

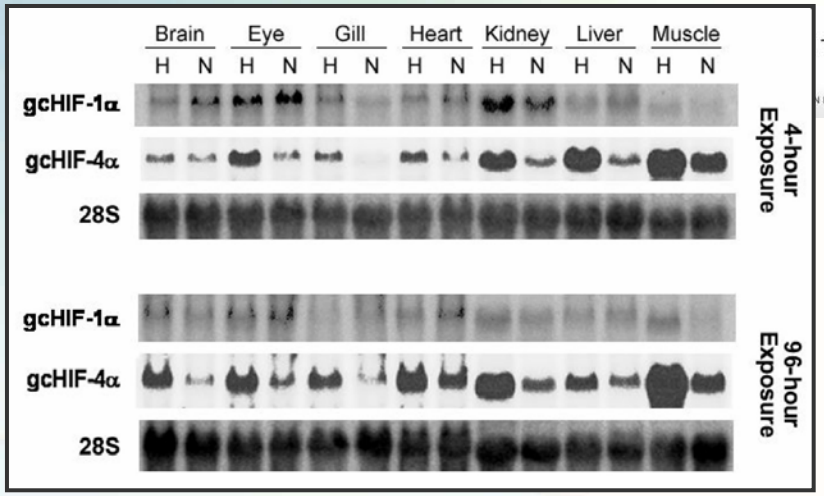
Introducing example



"Overall, normoxic expression and hypoxic induction patterns of the *gcHIF-1α* and *gcHIF-4α* genes were consistent amongst the replicate blots derived from different fish tissues, and a representative autoradiogram is shown. Under normoxic conditions, the 3.9-kb *gcHIF-1α* mRNA transcript was expressed most abundantly in eye and kidney; with lower expression levels being detected in brain, gill, heart and liver; and negligible expression in muscle. In contrast, the 3.7-kb *gcHIF-4α* transcript was expressed at comparatively higher levels in brain, heart, kidney, liver and muscle relative to *gcHIF-1α* under similar conditions. A marked increase in *gcHIF-1α* expression was observed in gill and kidney after exposure to hypoxia for 4 h (but not at 96 h); while *gcHIF-1α* expression was seemingly downregulated in brain, heart and liver, and appeared unchanged in eye. (…) Interestingly, *gcHIF-4α* was markedly upregulated following exposure to hypoxia for 4 and 96 h in eye, gill, heart, kidney, liver and muscle. Curiously, although the less abundant 2.1-kb *gcHIF-4α* transcript showed prominent expression and hypoxic up-regulation (ca. 5-fold) in kidney; expression of this smaller *gcHIF-4α* transcript was barely detectable in all other tissues examined under both normoxic and hypoxic conditions."

virtual liver network

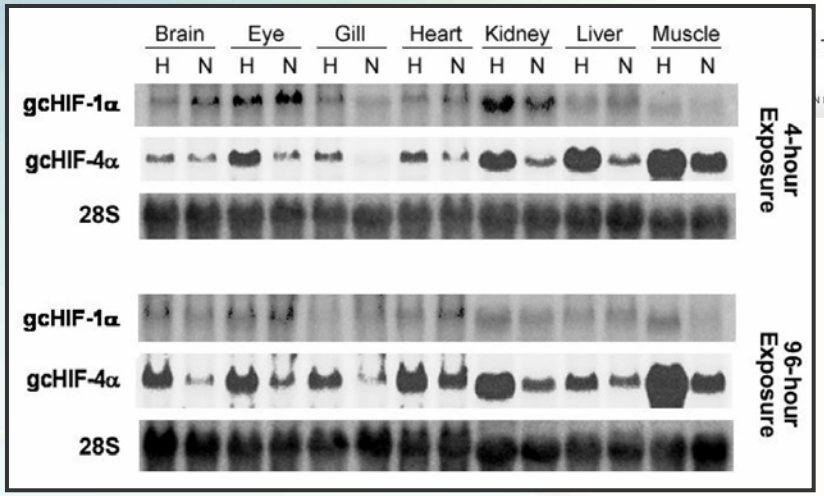
Introducing example



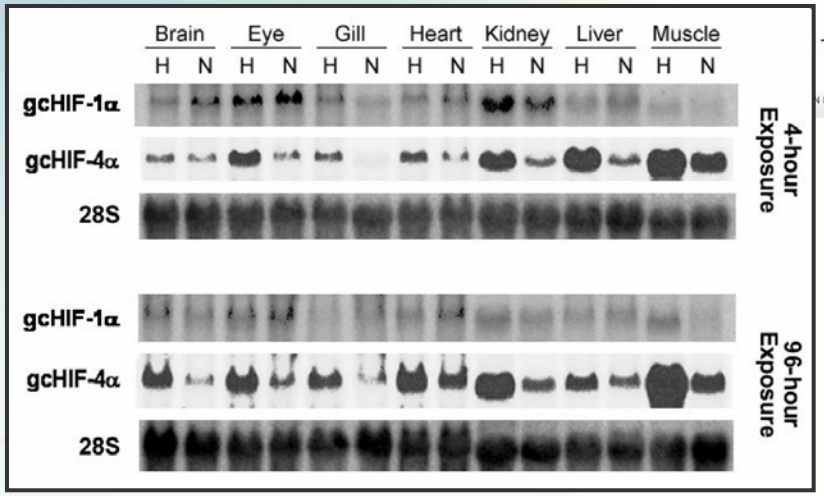
"Overall, normoxic expression and hypoxic induction patterns of the *gcHIF-1α* and *gcHIF-4α* genes were consistent amongst the replicate blots derived from different fish tissues, and a representative autoradiogram is shown. Under normoxic conditions, the 3.9-kb *gcHIF-1α* mRNA transcript was expressed most abundantly in eye and kidney; with lower expression levels being detected in brain, gill, heart and liver; and negligible expression in muscle. In contrast, the 3.7-kb *gcHIF-4α* transcript was expressed at comparatively higher levels in brain, heart, kidney, liver and muscle relative to *gcHIF-1α* under similar conditions. A marked increase in *gcHIF-1α* expression was observed in gill and kidney after exposure to hypoxia for 4 h (but not at 96 h); while *gcHIF-1α* expression was seemingly downregulated in brain, heart and liver, and appeared unchanged in eye. (…) Interestingly, *gcHIF-4α* was markedly upregulated following exposure to hypoxia for 4 and 96 h in eye, gill, heart, kidney, liver and muscle. Curiously, although the less abundant 2.1-kb *gcHIF-4α* transcript showed prominent expression and hypoxic up-regulation (ca. 5-fold) in kidney; expression of this smaller *gcHIF-4α* transcript was barely detectable in all other tissues examined under both normoxic and hypoxic conditions."

virtual liver network

Introducing example



downregulated, up-regulation — direction of regulation
less abundant, comparatively higher levels — compare across tissues and genes
most abundantly — top ranker
unchanged — expression equality (at which ε ?)
ca. 5-fold — numeric relations
barely detectable, negligible expression — different levels of zero
marked increase, markedly upregulated, prominent expression — emphasis on strength of expression
consistent patterns — summarizing evaluation of the relations of a 84-vector



Introducing example

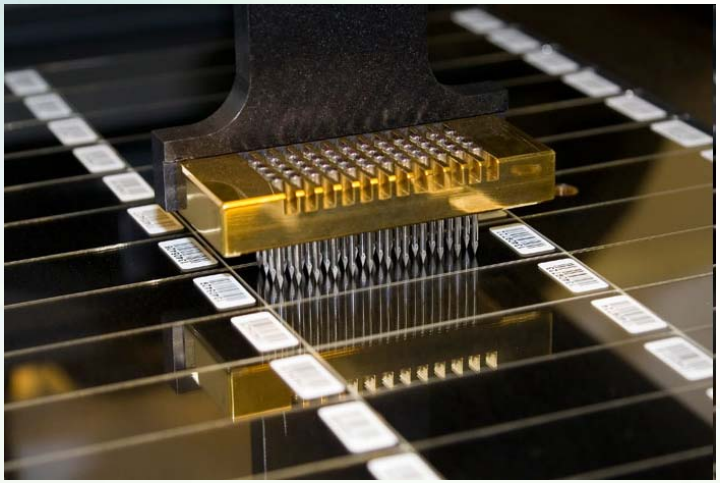
- Authors aware that
  - no linear relation: blot intensity — RNA count
  - unknown if affinity gcHIF1α and gcHIF4α equal
- but still
  - draw comparisons which distinguish ≈20 expression levels
  - confident on relation of gcHIF1α and gcHIF4α
  - deduce the “general pattern”
- *“What is acceptable to the molecular biologist must be acceptable to the systems biologist!”*

Contents

- Introduction
- Gene array accuracy
- What is semi-quantitative?
- Guidelines for semi-quantitative analysis
- Microarray preparation in ModeScore

RNA gene array

- Highly parallelized and automatized Northern blot

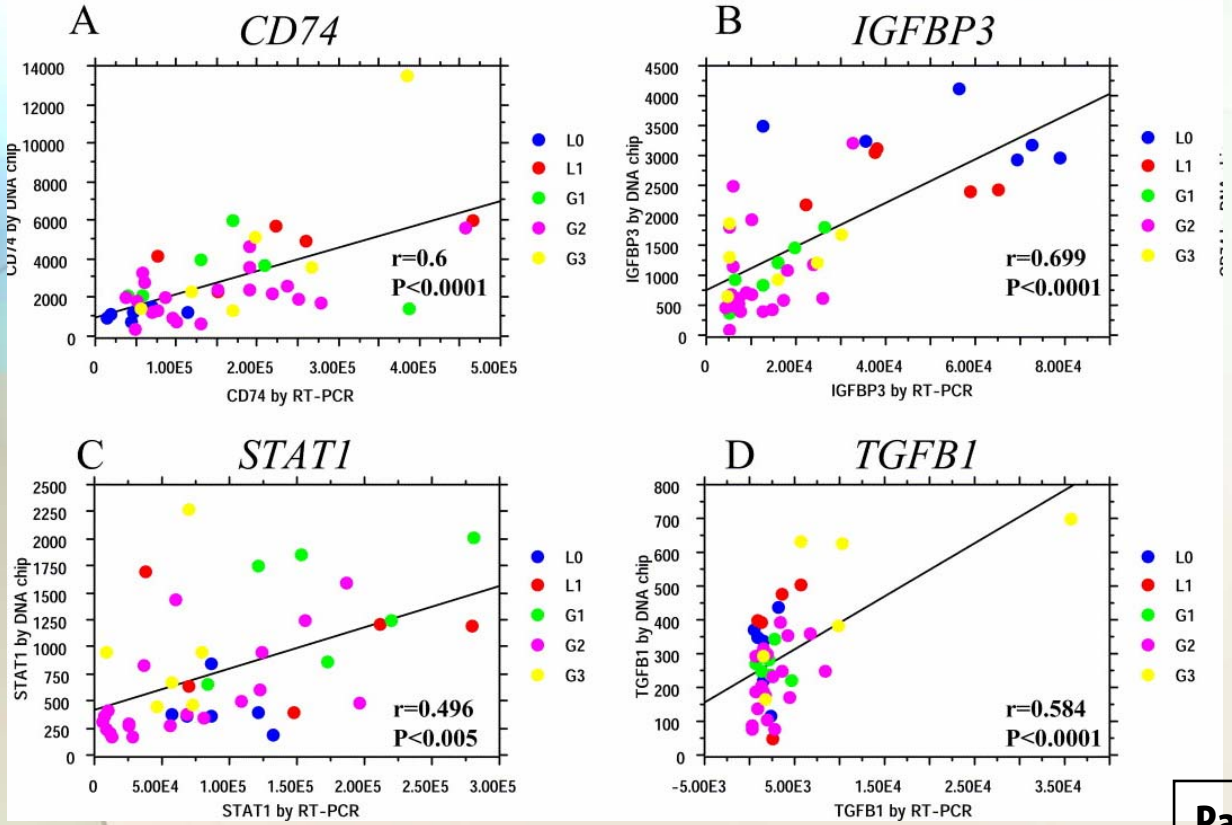


Microarray printer *BRI microarray lab*

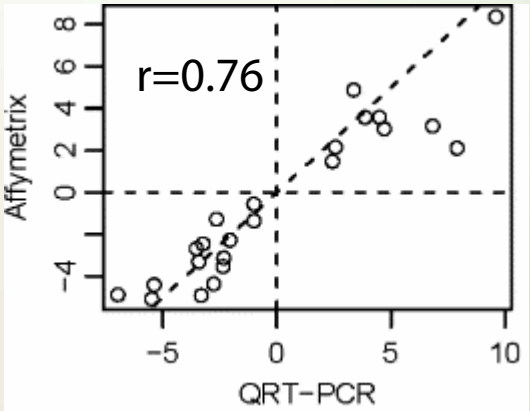
- Standardization at a higher level than individual blot experiments
  - millions of copies of a specific chip
  - community analyzing properties of the chips
  - sequence of probes known, chip definition can develop

Comparison with rt-PCR

Figure S2



Iizuka. al, *FEBS letters* 2005, 579(5):1089-1100.



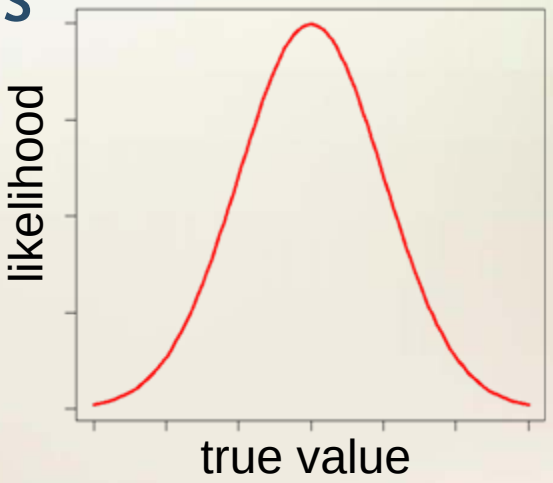
Park. al, *Journal of Biotechnology*, 2004, 112(3):225-245.

Note: rt-PCR closer to RNA count but far from accurate

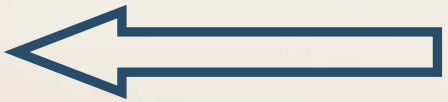
Gene array accuracy

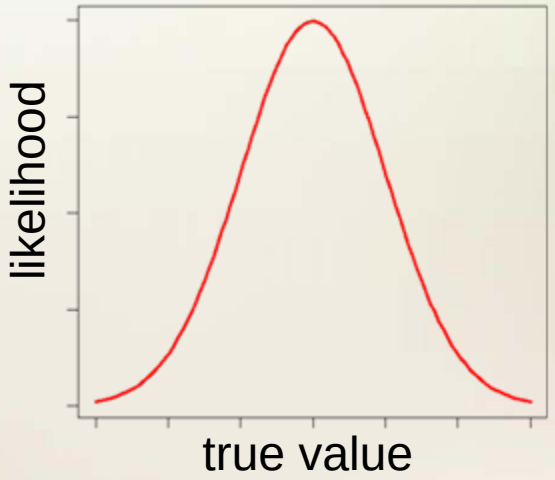
- Pearson correlations: 0.5 … 0.75
- Luminance/RNA only approximately linear
- Slope different for genes
- Slope unknown for most genes

Thus, given a luminance value, and an average slope



True for any experimental technique!
Just deviation differs.





Contents

- Introduction
- Gene array accuracy
- What is semi-quantitative?
- Guidelines for semi-quantitative analysis
- Microarray preparation in ModeScore

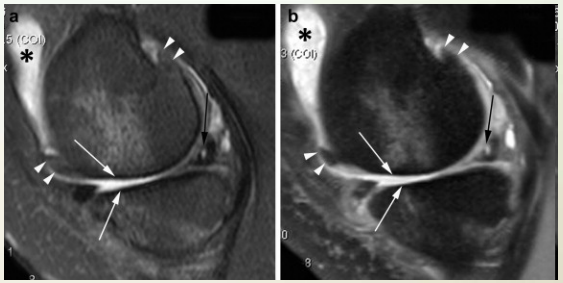
What is semi-quantitative?

- In mass-spectrometry:
  - measurements of accuracy ±30% … ±50%
  - quantitative: accuracy at least ±10%

  Amarasiriwardena et al, 1997, Microchem J., 56 (3) 352ff

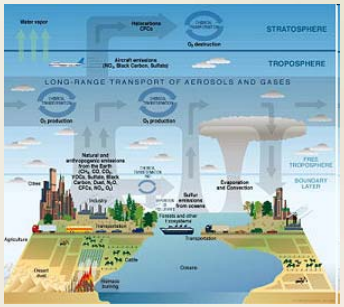
- In medical imaging:
  - Discrete scores of image properties (≤8 levels) combined (8 scores)

  Crema et al, 2013, Osteoarthritis & Cartilage, 21(3) 428ff



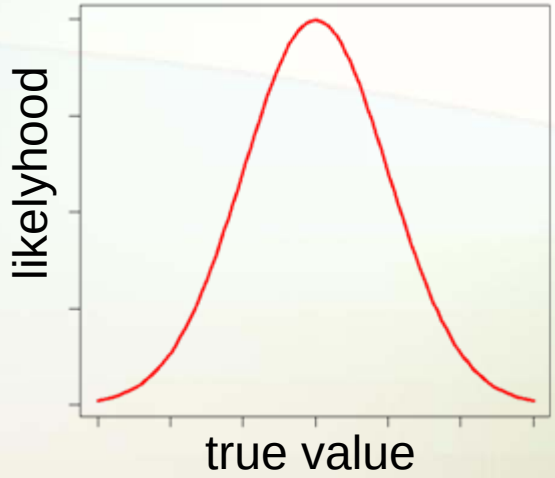
- In environmental modeling:
  - ranked observations

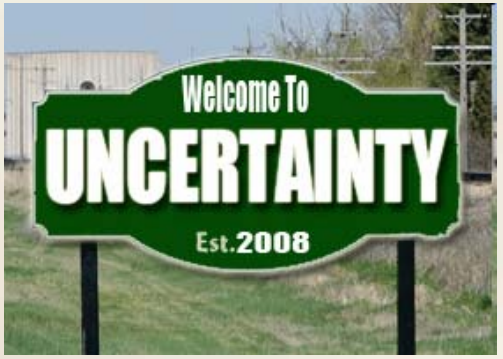
  Legendre et al., 1998, Dev in Eviron Mod 20, 185ff



Semi-quantitative analysis

- Using data with
  - large deviations from true values
  - Unknown data distribution
  - Unknown error distribution
  - Systematic bias possible
  - Significance calculations unsafe
- Cope with a higher level of





Contents

- Introduction
- Gene array accuracy
- What is semi-quantitative?
- Guidelines for semi-quantitative analysis
- Microarray preparation in ModeScore

Guidelines/rules for SQA (summary)

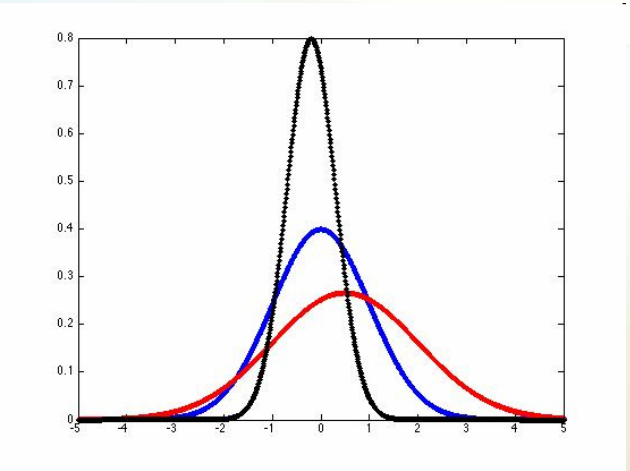
- Identify **levels of accuracy**
- Little **data derivation**
- **Re-check** at raw data
- Focus on **strong effects**
- **Robust** calculations
- **Robust** reasoning

Identify levels of accuracy

- Same **gene** comparisons > across genes
- Same **batch** comparisons > across batches
- Same **chip type** comparisons > across chips
- Same **lab** comparisons > across labs
- Median expression **values** > off and highly abundant genes

Little data derivation

"Every data processing step (potentially) *widens* the error distribution"



- Expression difference
- Averaging
  - Pathway averaging
  - Biological repeats averaging
- Thresholding
- Expression change correlation

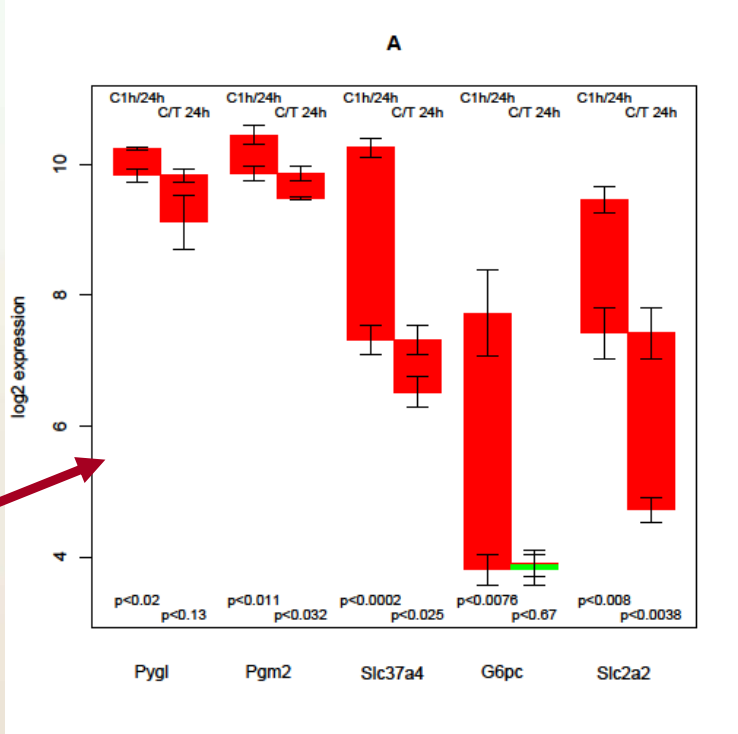
Re-check at raw data

"Did data derivation introduce artifacts?"

"Could the microbiologist follow the argument at the blots?"

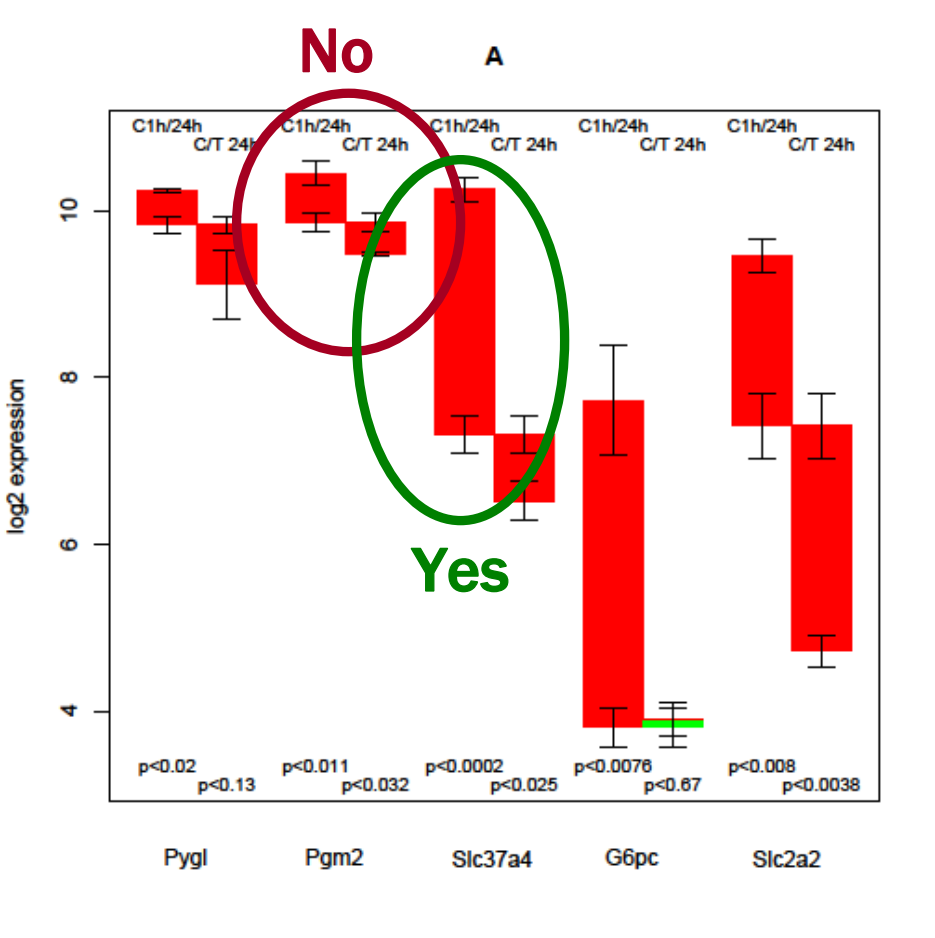
In ModeScore, 3 steps

- rank functions
- rank genes in a function (selected functions)
- evaluate regulation pattern (selected genes)



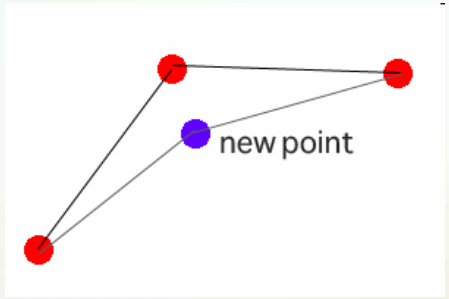
Focus on strong effects

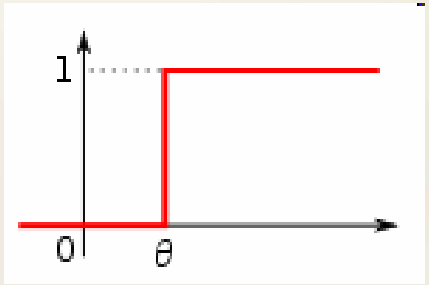
- Huge impact on significance
- Likelihood of artifacts lower
- Mechanistic cause more likely
- Higher possible relevance



Robust calculations

- Sensitivity to small changes
- Stable calculations
  - Averaging
  - Ranking
- Unstable calculations
  - Clustering
  - Thresholding





Robust reasoning

- Calculation result: **evidence**, not proof
- Focus on results consistent with
  - literature knowledge
  - published regulation mechanisms
- Contradicting results
  - needs additional evidence
- Final aim: give **hypotheses**

Contents

- Introduction
- Gene array accuracy
- What is semi-quantitative?
- Guidelines for semi-quantitative analysis
- Microarray preparation in ModeScore

Aroma project

- Open-source R framework for microarray analysis
- www.aroma-project.org, Henrik Bengtsson
- Fast, stable, flexible
- Workflow
  - RMA background correction
  - RMA quantile normalization
  - RMA probe summarization

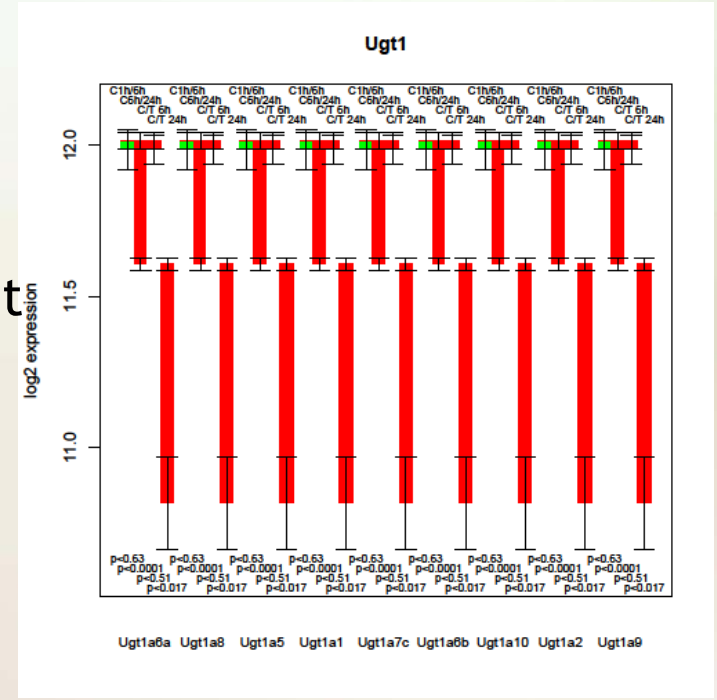
Custom chip definitions *critical!*

Genes encoding enzymes/transporters

- better annotated
- minor differences in probe set definitions

CDF problem

- Mouse 430a_2 chip
- UDP glucuronosyl transferase 1
  - Brainarray CDF: all isoforms absent
  - Affymetrix CDF: present, all isoforms same value
- ModeScore: coverage more important
- ModeScore function ranking, *nearly identical*



Summary

Microarray experiments *can*

approximately estimate

- absolute expression
- expression changes
- expression change correlation

allow the comparison of gene changes

if principles of semi-quantitative reasoning are applied

virtual liver network

CHARITÉ
UNIVERSITÄTSMEDIZIN BERLIN

The end