Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

www.nr.no

Limitations of mRNA amplification from small-size cell samples

Anders Løland

Norsk Regnesentral

EMS 2005

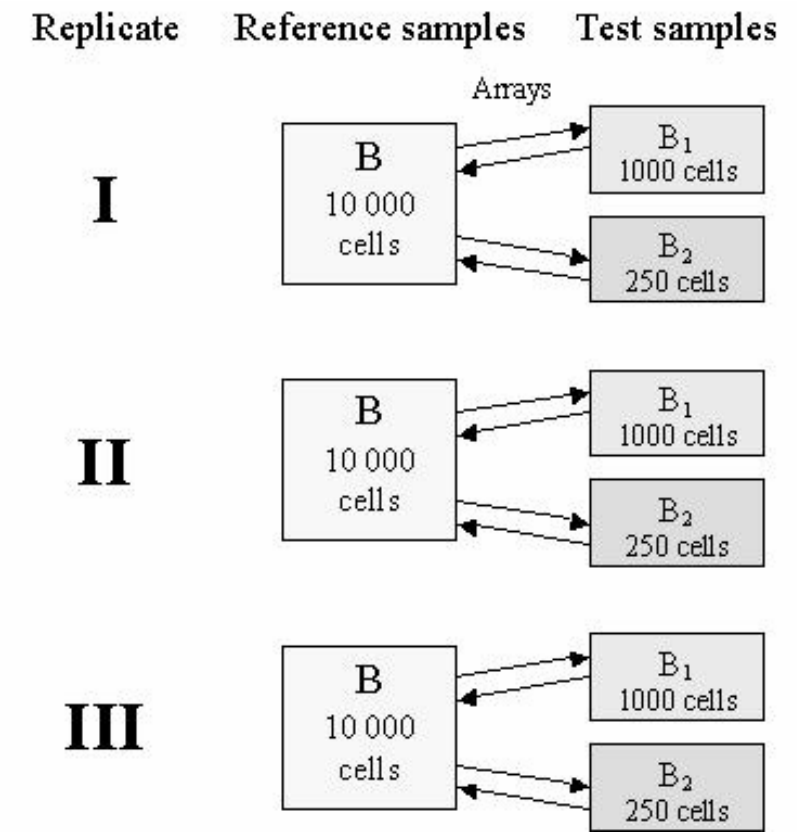
Background

- Microarrays represent a high-throughput technology for studying gene expression.
- The amount of biological material (total RNA) needed for a microarray experiment is often not available.
- A widely used approach to obtain gene expression profiles from limited material, is to **amplify** the starting material.
  - How **reliable** is the reflection of the starting material in the results obtained from the microarray experiments?
  - Especially important with **extremely low** quantities of input RNA.
  - A detailed **quantitative investigation** is **currently lacking** in the literature.
- The results presented here are described in the submitted paper:

  "Limitations of mRNA amplification from small-size cell samples".
  Vigdis Nygaard, Marit Holden, Anders Løland, Mette Langaas, Ola Myklebost, Øystein Fodstad, Eivind Hovig.

Experimental design

- Aim of the paper: Investigate the effect of different amounts of starting material.
  - Will a certain amount of starting material reliably reflect the starting material in the results obtained?
  - In addition, is the reliability of the results gene/concentration specific?
- Three cell sample sizes were compared.
  - **10 000** cell samples (**reference**).
    - Represent samples sufficiently many for reliably reflecting the starting material.
  - **250** and **1000** cell samples.
    - Represent limited material for investigating the effect of reduced starting material.
  - Three batches amplified for each cell sample size.



Investigating the effect of reduced starting material

- Analysis of variance (ANOVA; a mixed effects model) was used to isolate the effect of reduced starting material.
  - This analysis clearly showed that the results are dependent on the size of the starting material.
- Only the overall effect of different amounts of starting material was investigated.
  - We could conclude that (parts of) the results obtained from 250 and 1000 cell samples were not sufficiently well reflecting the starting material, but we could not conclude which parts.
- This conclusion was supported by multiple hypothesis testing.

Investigating the effect of reduced starting material (cont.)

- In the previous talk “Model-based estimation of transcript concentrations from spotted microarray data”, Marit Holden presented the **Transcount** method.
- From expression data, TransCount provides
  - estimates of absolute transcript concentrations for each gene and examined sample and
  - the highly multivariate joint posterior probability distribution of all concentrations obtained via MCMC.
- We used this information to find which parts of the data obtained from limited material that are unreliable for the different sample sizes.

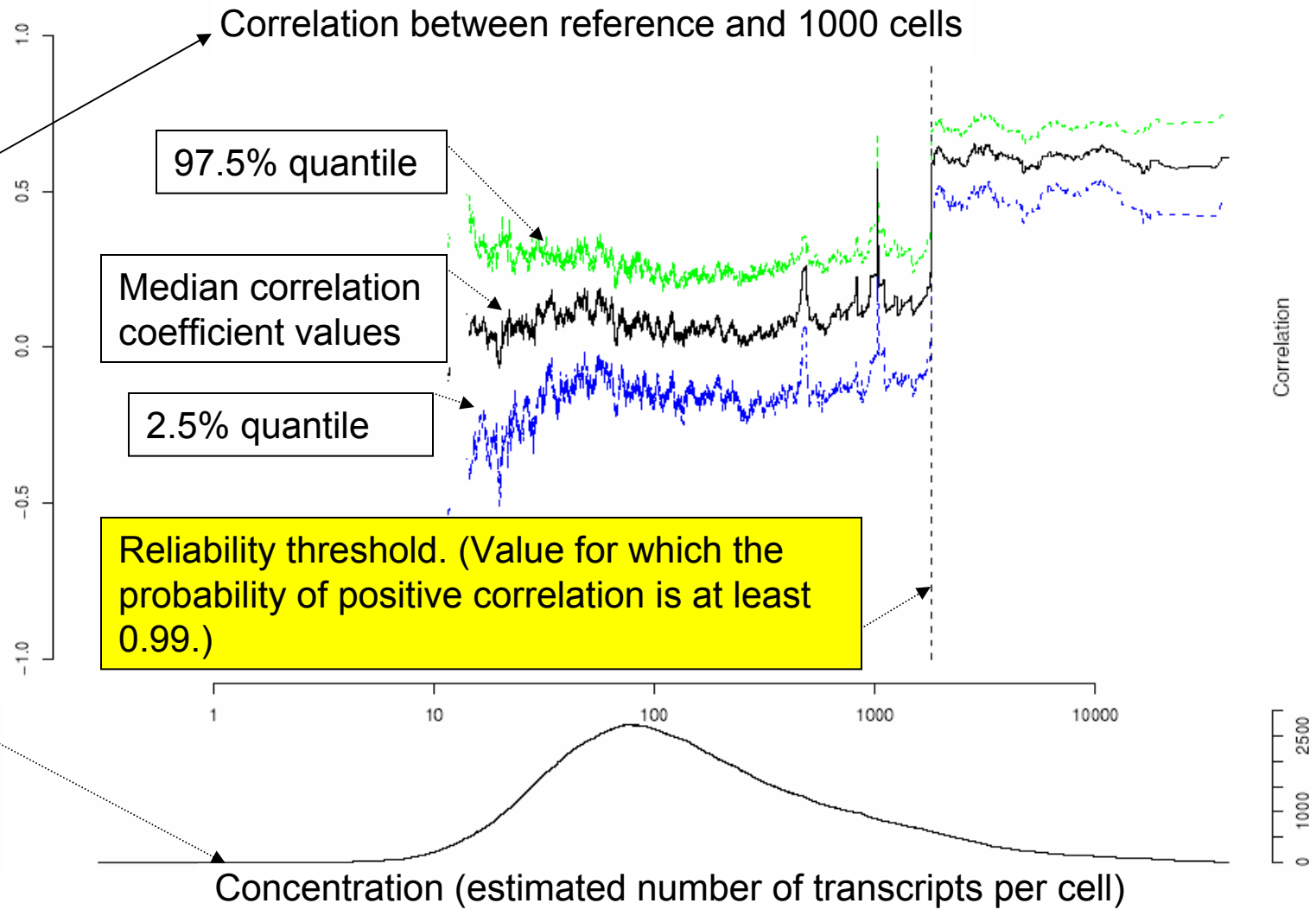
Using the output from Transcount

- We used the posterior probability distribution of all transcript concentrations for calculating, for each concentration, the distribution of the Pearson correlation coefficient ρ between the reference and 250 (1000) cell sample transcript concentrations.
- The correlation for a certain concentration *c* was computed from genes with concentrations in a small interval around *c*.
- The correlations were clearly transcript copy number dependent. High copy numbers yielded high correlation coefficients, and a critical level was observed.
- In the plots on the next two slides, a reliability threshold is found for each of the cell sample sizes 250 and 1000.

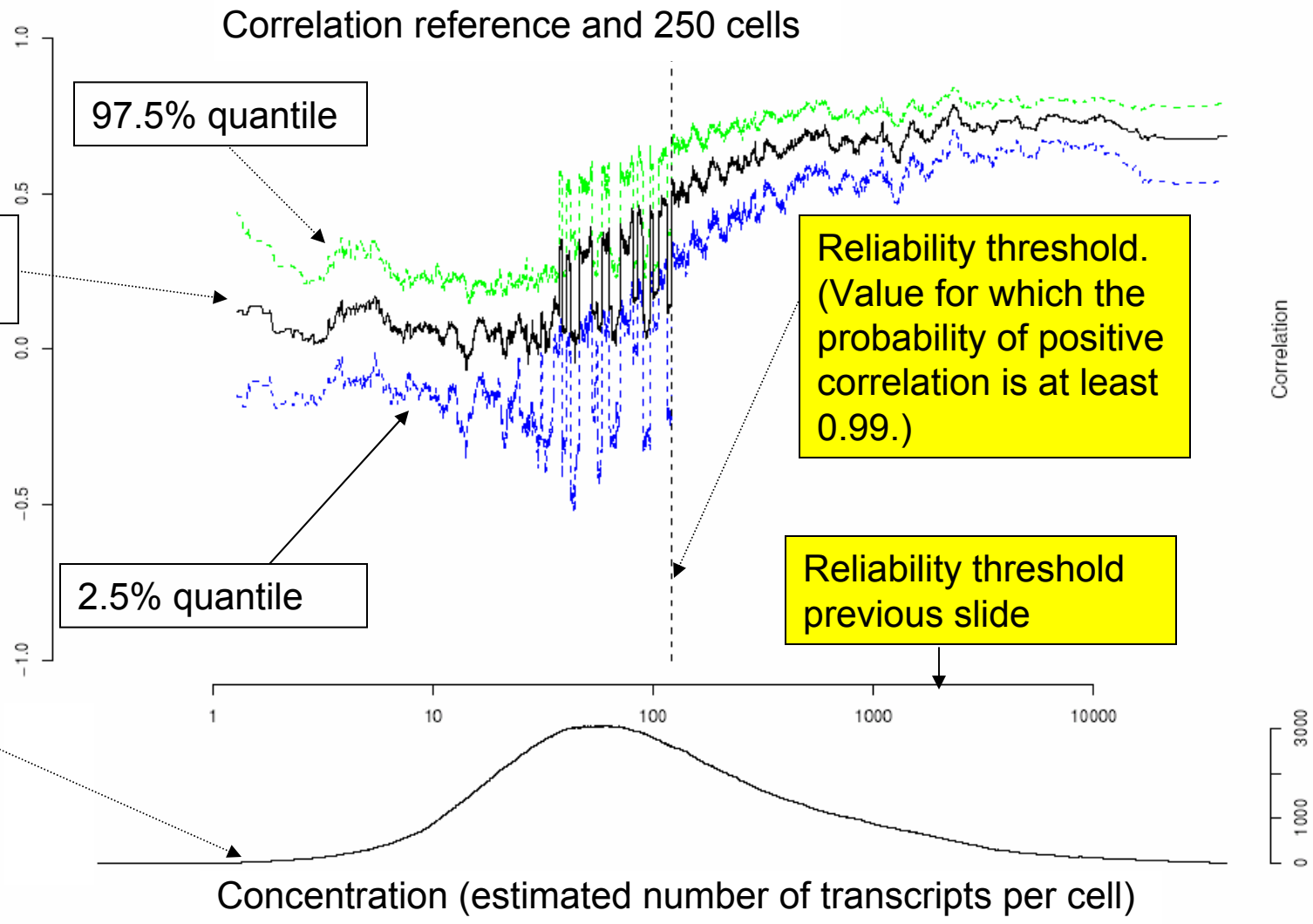
Summary values for the 1000 cell samples computed from the MCMC traces

The correlation for a certain concentration *c* was computed from genes with concentrations in a small interval around *c*

Number of genes per concentration. For a certain concentration *c*, the number of genes is counted in the interval (*c/2*, *2c*).



Summary values for the 250 cell samples computed from the MCMC traces



Conclusions

- Correlations were clearly transcript copy number dependent.
  - High copy numbers yielded high correlation coefficients, and a critical level was observed.
- We can define limits with respect to number of transcripts necessary for meaningful interpretation of expression data from conditions using reduced input RNA.
  - As the input cell number decreases, the necessary number of transcripts per cell increases.
- Based on our findings regarding the sensitivity limit of our amplification protocol in use we conclude that using positive correlation as a measure of reliability,
  - moderate to high expressing genes can be regarded in experiments with <1000 cells and
  - all data from low expressing genes should be disregarded due to inaccuracy.
- Caution is warranted when extrapolating biological relevance from the increasing number of expression profiling results published based on extremely low cell numbers.