

Microarray Data Analysis: Beyond Differential Expression and Clustering

Korbinian Strimmer

Department of Statistics, University of Munich, Germany.

Munich, 19 February 2004



State of Microarray Analysis

In the workshop we have covered so far the three main pillars of microarray data analysis:

- i) Calibration and normalization (data adjustments and transformations)
- ii) Differential expression (t-tests, ANOVA, multiple testing)
- iii) Multivariate analysis (MANOVA, clustering, classification)

The nature of microarray data (high dimensionality, small sample size) typically requires specially tuned methods.

However, it turns out that (contrary to the assumption in the early hype) there was no need to invent radically “new statistics” and many adhoc algorithms proposed by computer scientists have also turned out to be good old friends indeed...

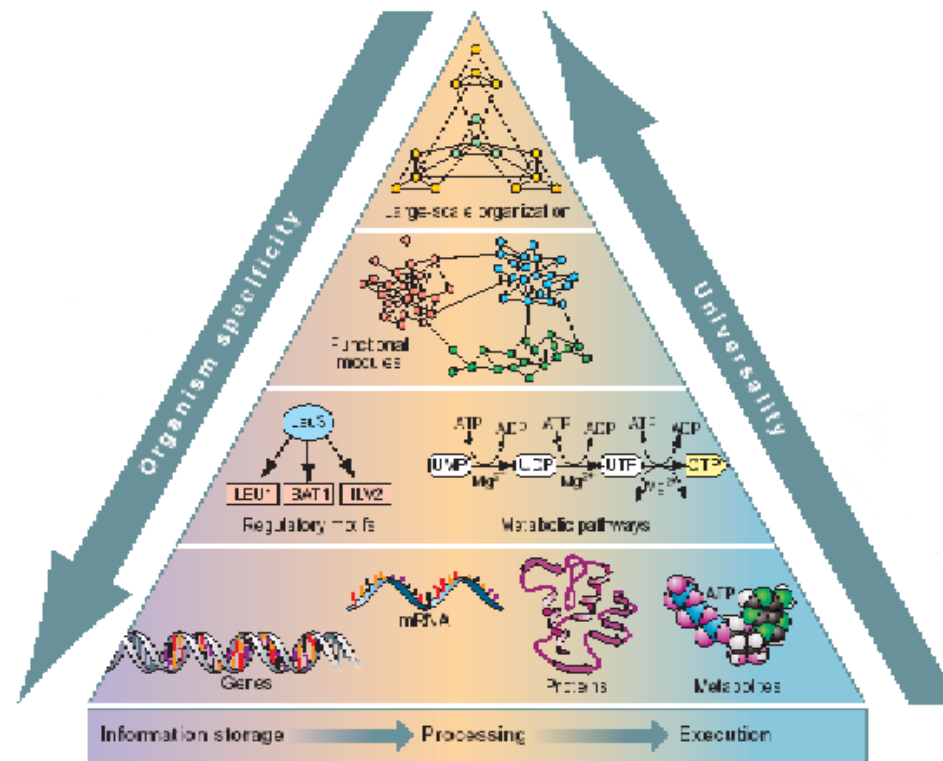
Thanks to concentrated effort in research in the period 1999-2003 (and several hundred papers later...) a more or less **standard set of tools** for all three of the above tasks **has emerged**.

So what next?

One lesson to learn is that proper statistical modeling, as opposed to simple algorithmic analysis, is extremely important when analyzing noisy genomic data.

But what are new questions for biologists and statisticians alike in the current post-genomics era?

Life's Complexity Pyramid

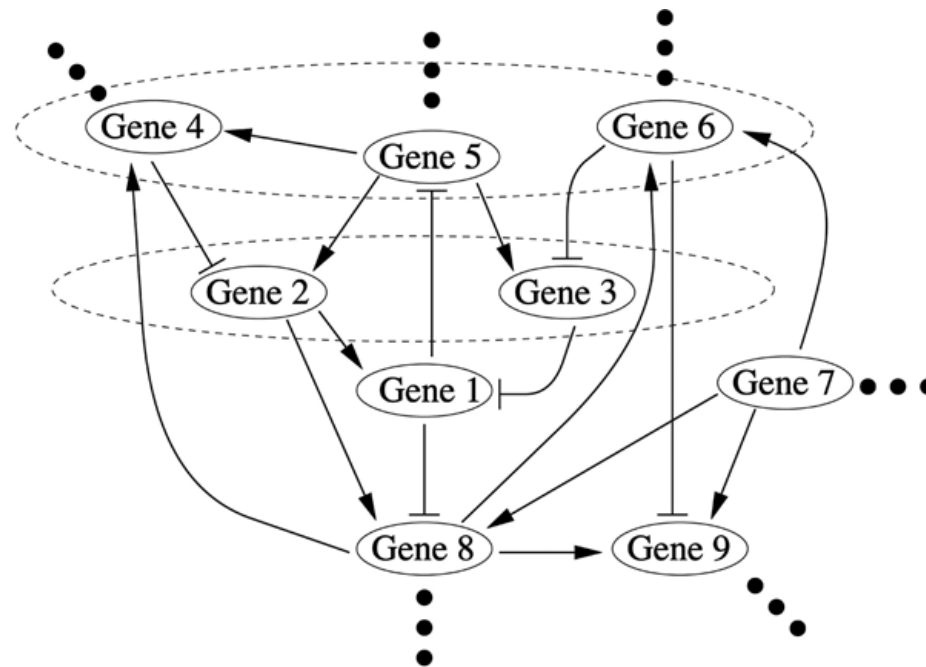


(figure from Science 2002)

Understanding the Mechanisms in a Cell

The new paradigm is the **network**:

→ network biology, systems biology



Challenge

We want/need to study interaction among hundreds or even thousands of genes simultaneously.

Challenge: create dynamic stochastic interpretable models for prediction and inferring of large scale gene interaction

Data type: mostly multiple time series data

Methods used: classification, graphical models, time series analysis

Funny: similar problems (though on much smaller scale with more data) arise in computational economics (models of stock markets) and mathematical biology (ecology).

Today's Program in the Workshop

On this last day we aim to give you a perspective on our current research in system-biology related methods for gene expression analysis:

iv) Analysis of gene expression time series data

1. Analysis of cell cycle data
→ Korbinian Strimmer
2. Gene interaction and emerging patterns
→ Anne-Laure Boulesteix
3. Graphical models for genetic networks
→ Juliane Schäfer

Practical exercises (using GeneTS and graph R packages) in the afternoon!