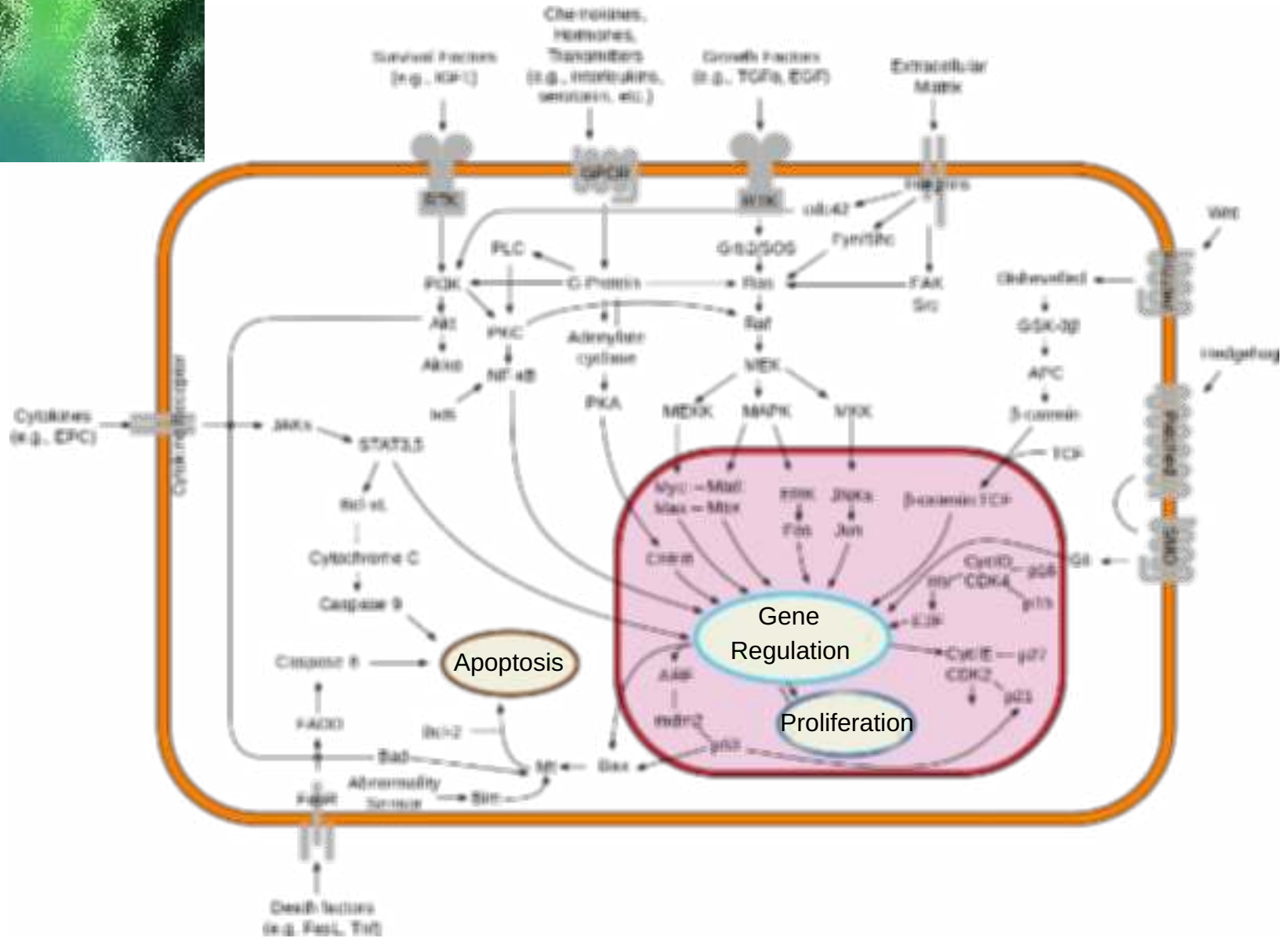
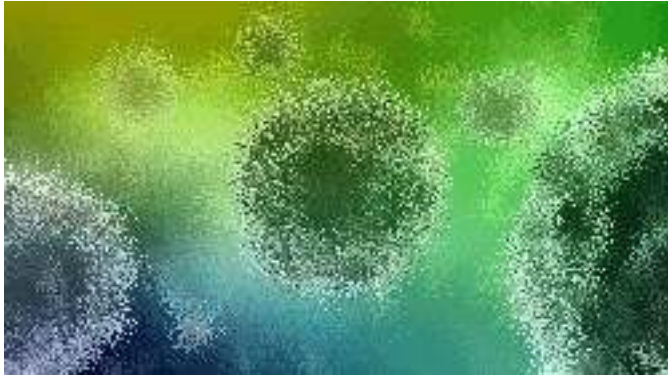


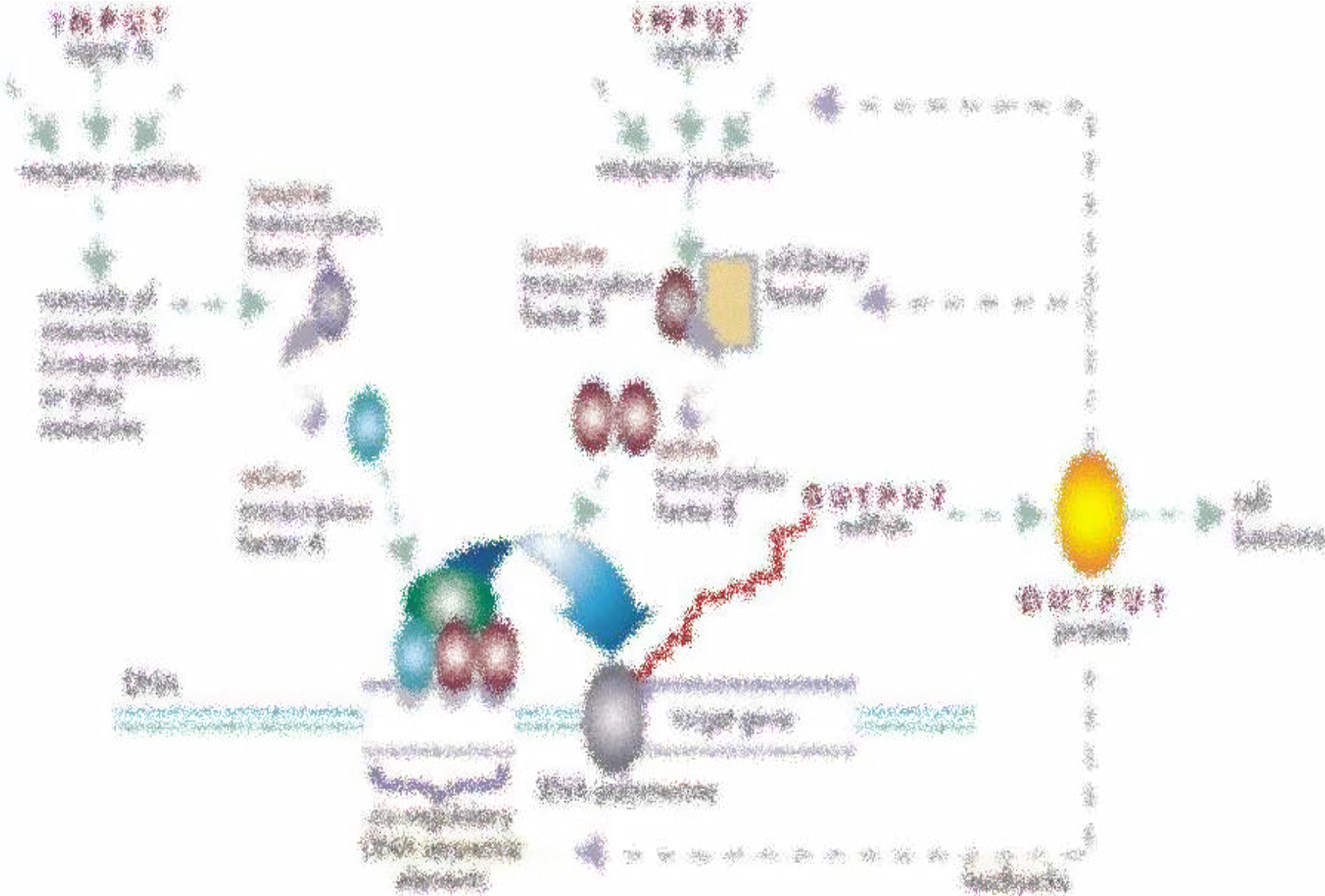
Methods to analyze transcriptome data in view of gene regulation and signaling pathways

**Prof. Dr. Tim Beißbarth
Institute of Medical Statistics
Statistical Bioinformatics Group**

We want to understand the molecular workings of a living cell



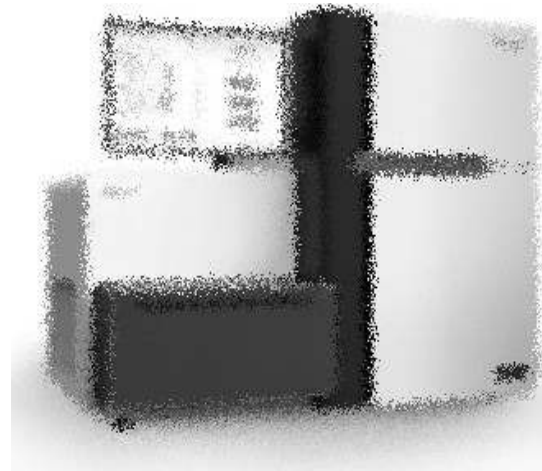
We want to understand the molecular workings of a living cell



Most of the time we measure only transcriptome levels



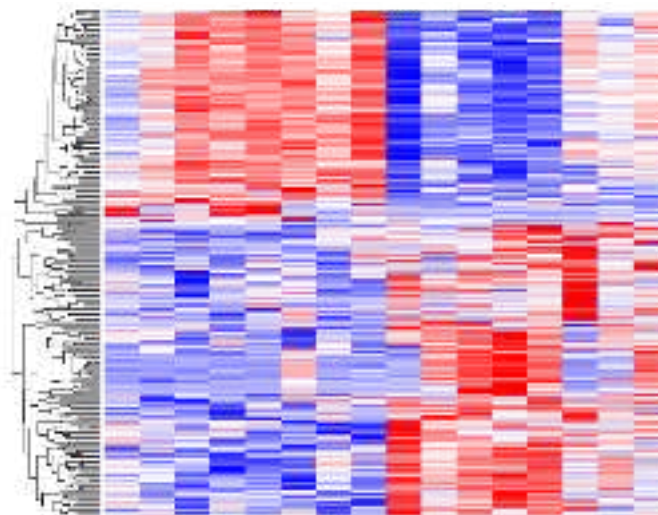
Microarrays
since about 1990s



RNA-Seq
since about 2010s



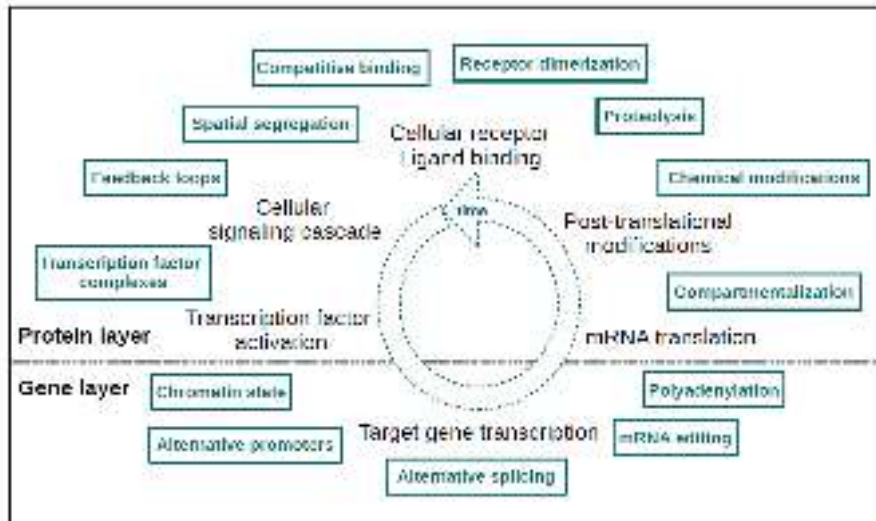
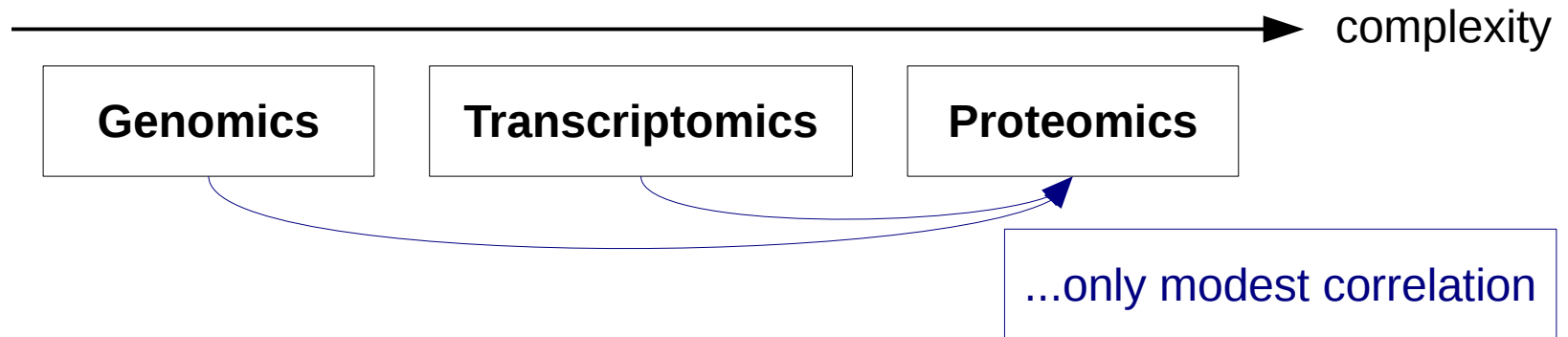
matrix of
gene expression
levels



different cellular conditions

almost all gene transcripts

Can we learn about the workings of a cell based on transcriptomics data?



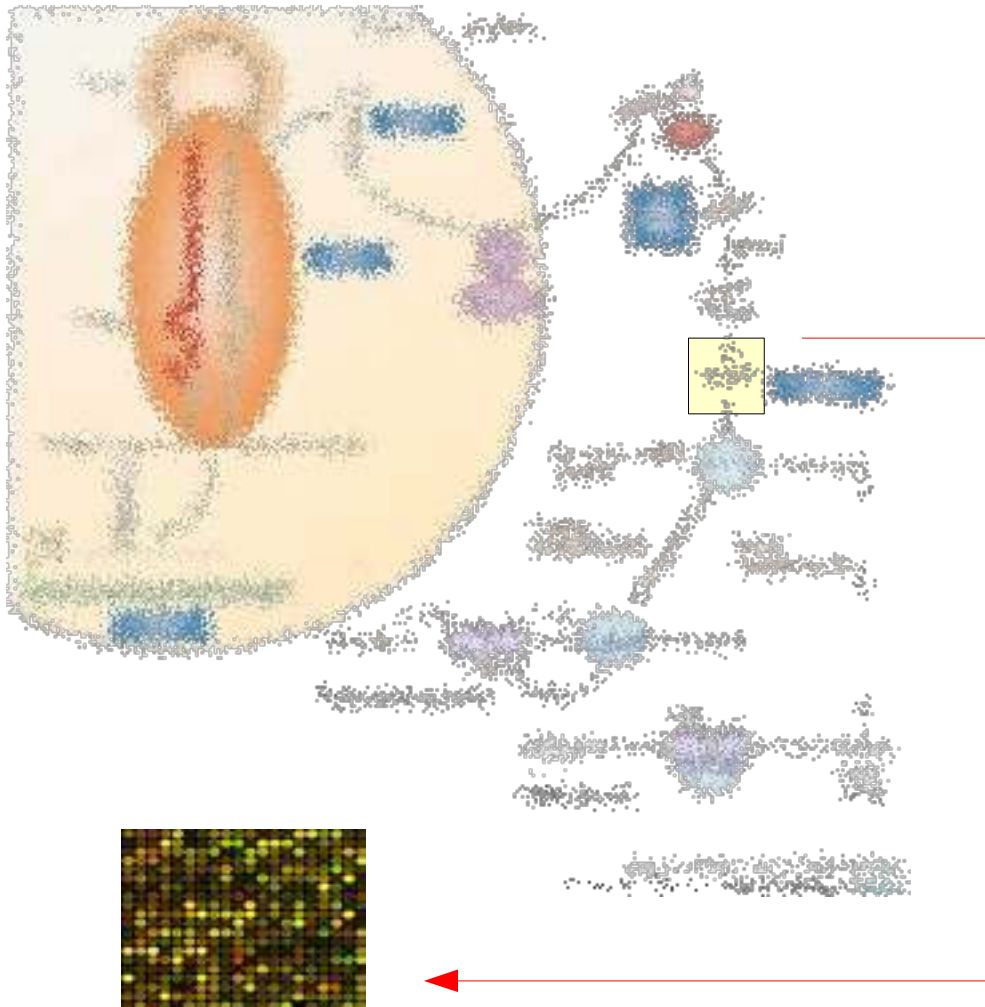
regulatory control on different cellular layer:

- protein layer
- protein activation layer
- transcription factor layer
- miRNA layer
- transcript/mRNA layer
- gene layer
- ...

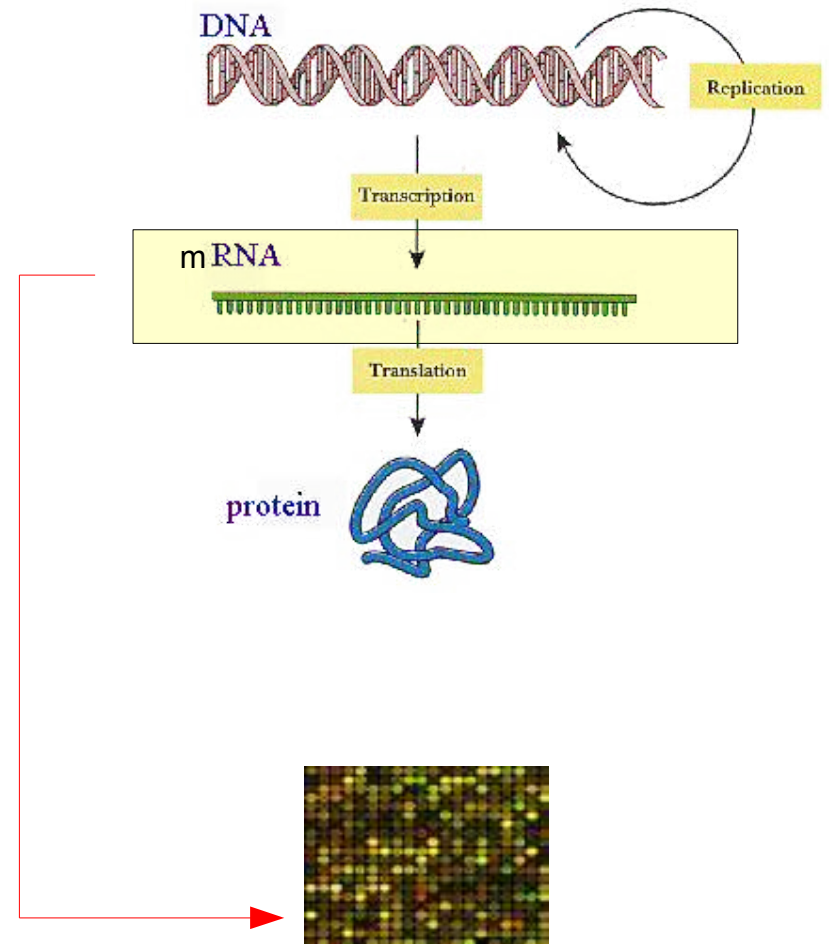
Can we estimate miRNA activity from gene expression data?

- miRNAs are important regulators of gene expression
- often Gene Expression Microarrays and miRNA-Microarrays are performed in parallel

• miRNA Expression



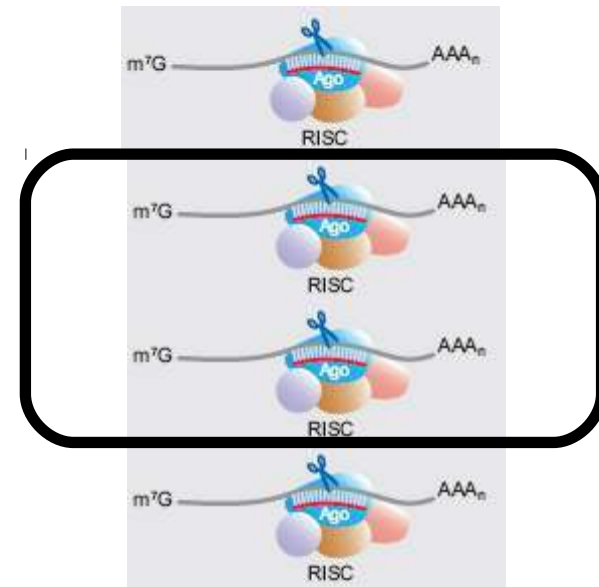
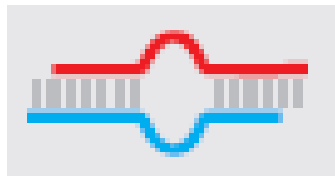
• Gene Expression



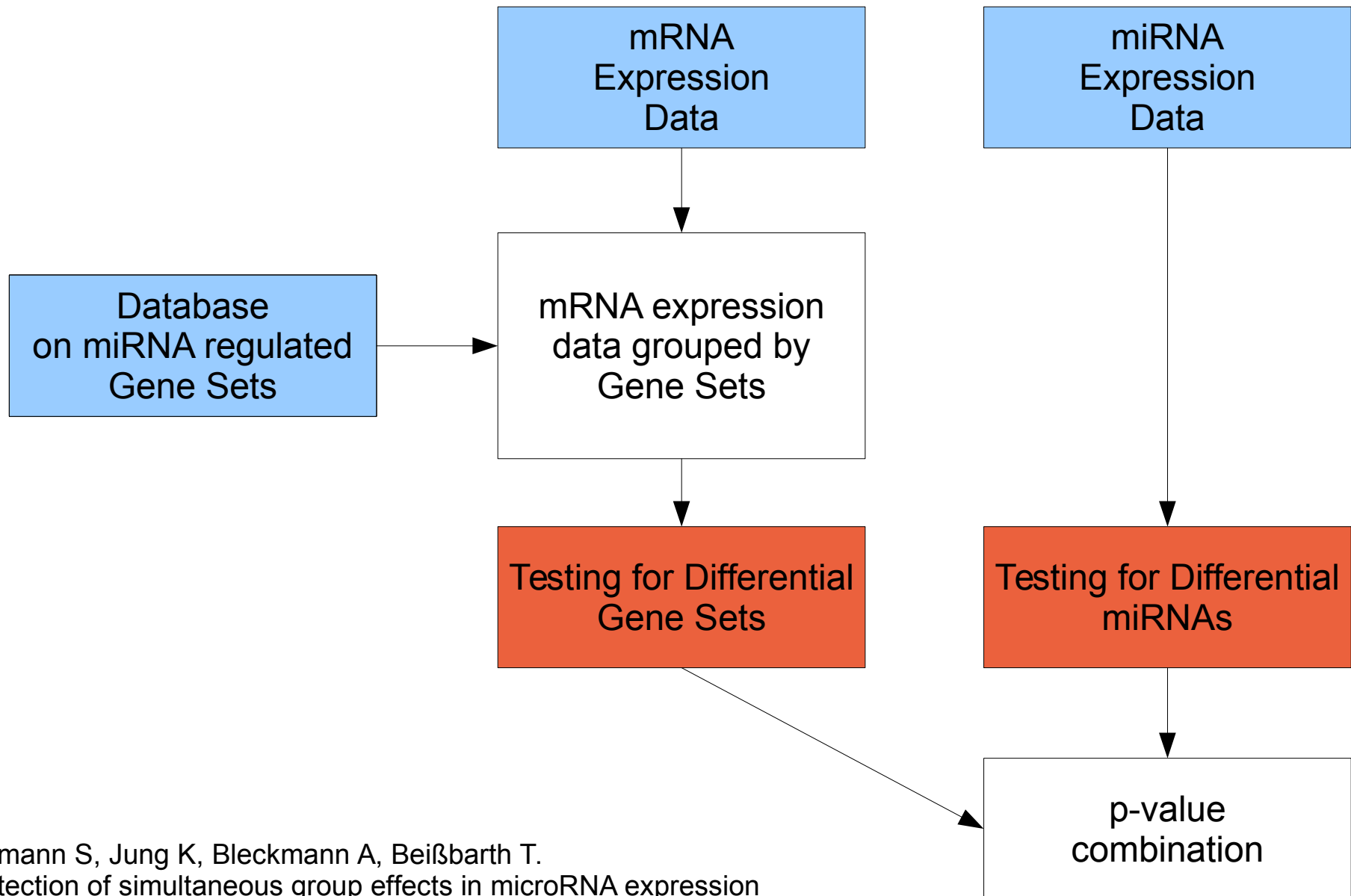
Different sources of information

- Expression of miRNAs
- Expression of mRNAs
- Target Prediction: which miRNA influences which mRNA?

e.g. MicroCosm (Griffiths-Jones et al, Nucleic Acids Res, 2008)



Combination of Test Results in order to find differential miRNAs



Artmann S, Jung K, Bleckmann A, Beißbarth T.
Detection of simultaneous group effects in microRNA expression
and related target gene sets.
PLoS One. 2012;7(6):e38365.

R – Package: *mirTest*

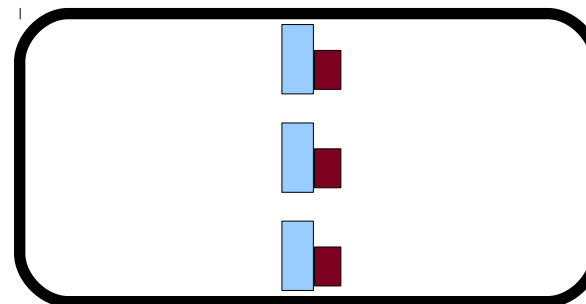
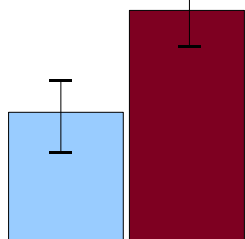
Use Gene-Set Enrichment Tests

miRNA Expression

mRNA Expression

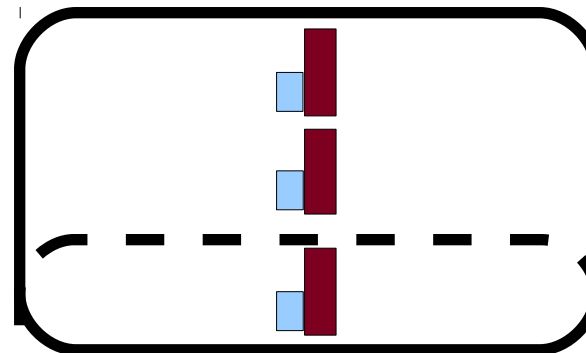
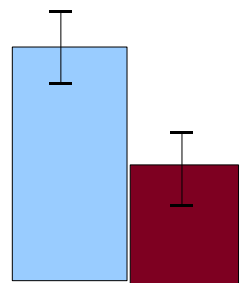
Mean in Group 1 Mean in Group 2

miR-1



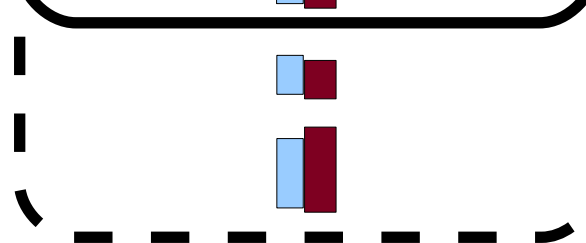
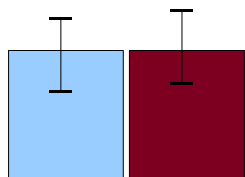
GS-1

miR-2



GS-2

miR-3



GS-3

LIMMA
(Smyth et al. 2004)

**Gene Set Enrichment /
Globaltest**

Global vs. Enrichment tests

- Enrichment Tests

competitive Null-Hypothesis

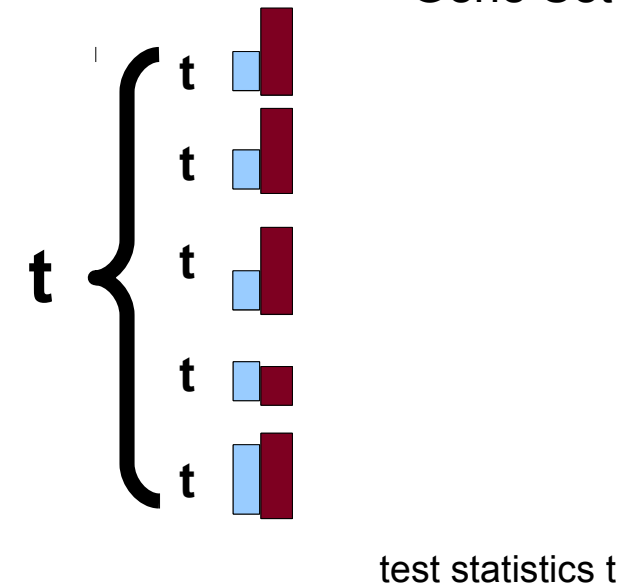
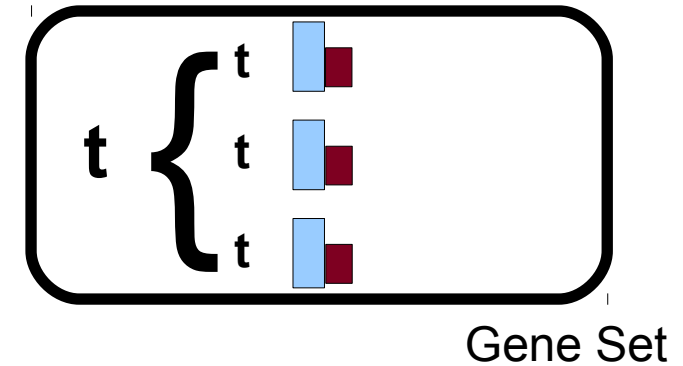
(e.g. Fisher Test, Wilcoxon, Kolm.-Smirnov Test)

- Globaltests

self contained
Null-Hypothesis

(e.g. GlobalTest, GlobalAncova, RepeatedHighDim)

mRNA Expression



Beissbarth T, Speed TP.

GOstat: find statistically overrepresented Gene Ontologies within a group of genes. Bioinformatics. 2004 Jun 12;20(9):1464-5.

Jung K, Becker B, Brunner E, Beissbarth T.

Comparison of global tests for functional gene sets in two-group designs and selection of potentially effect-causing genes. Bioinformatics. 2011 May 15;27(10):1377-83.

Combination of P-values using a meta-analytic approach

miR-1 p-Value  p-Value 1  p-Value GS-1

miR-2 p-Value  p-Value 2  p-Value GS-2

miR-3 p-Value  p-Value 2  p-Value GS-3

P-value combinations with method of Fisher or Stouffer.

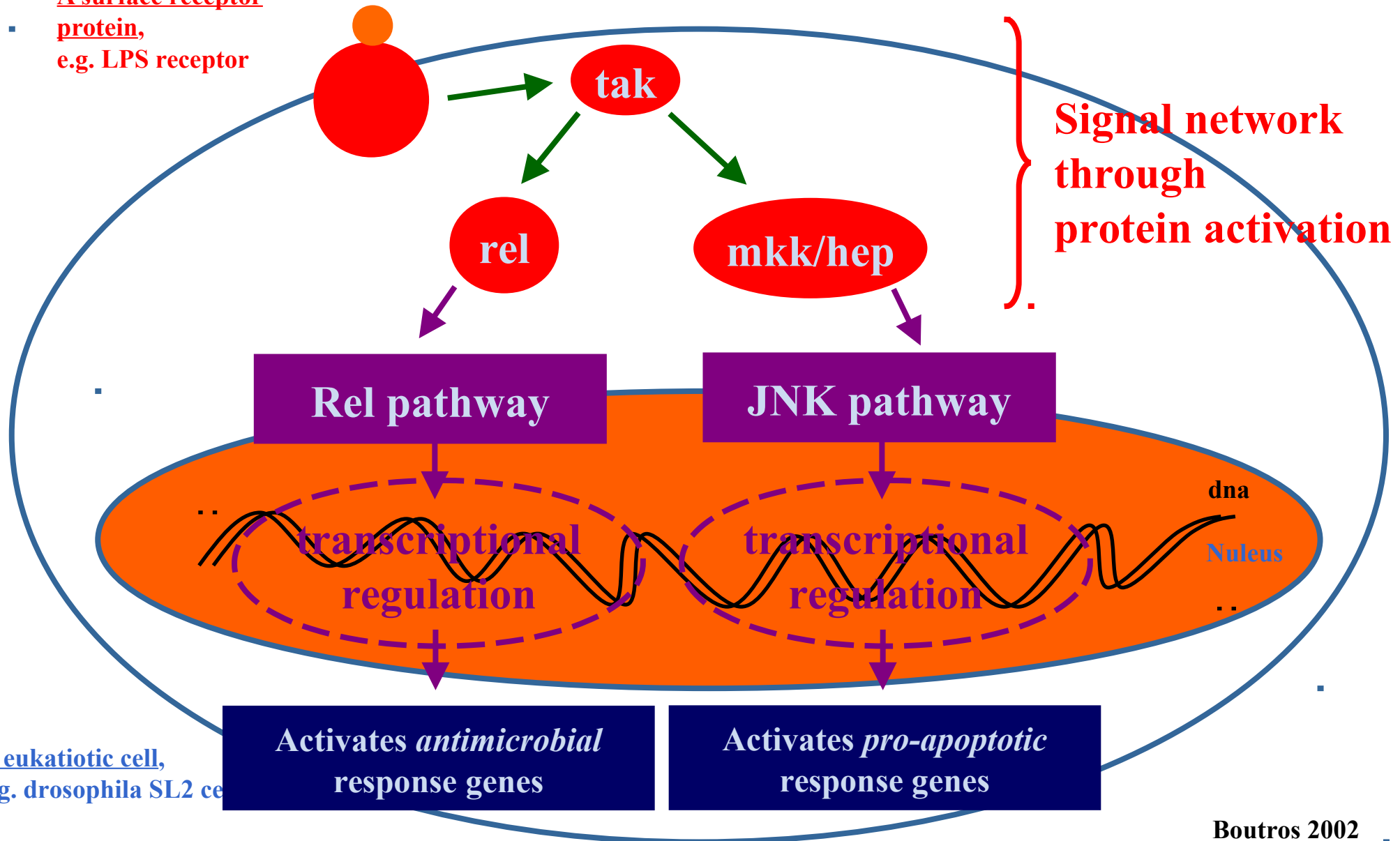
Results of simulation

Test	FDR	Power
Globaltests		
<i>Globaltest</i>	≥ 0.05	Limma < GST < Combi.
<i>GlobalAncova</i>	≥ 0.05	Limma < GST < Combi.
<i>RepeatedHighDim</i>	$\gg 0.05$	Limma < GST < Combi.
Enrichment Tests		
<i>Kolm. Smirnov</i>	± 0.05	Limma < GST < Combi.
<i>Wilcoxon</i>	± 0.05	Limma < GST < Combi.
<i>Fisher</i>	$\ll 0.05$	Limma < GST < Combi.
Rotation Tests		
<i>ROAST</i>	± 0.05	Limma < Combi. < GST
<i>Romer</i>	± 0.05	Limma < Combi. < GST

Can we learn signaling pathways based on transcriptome data?

An external stimulus,
e.g. LPS - a principal cell wall component of bacteria

A surface receptor
protein,
e.g. LPS receptor



Experimental data

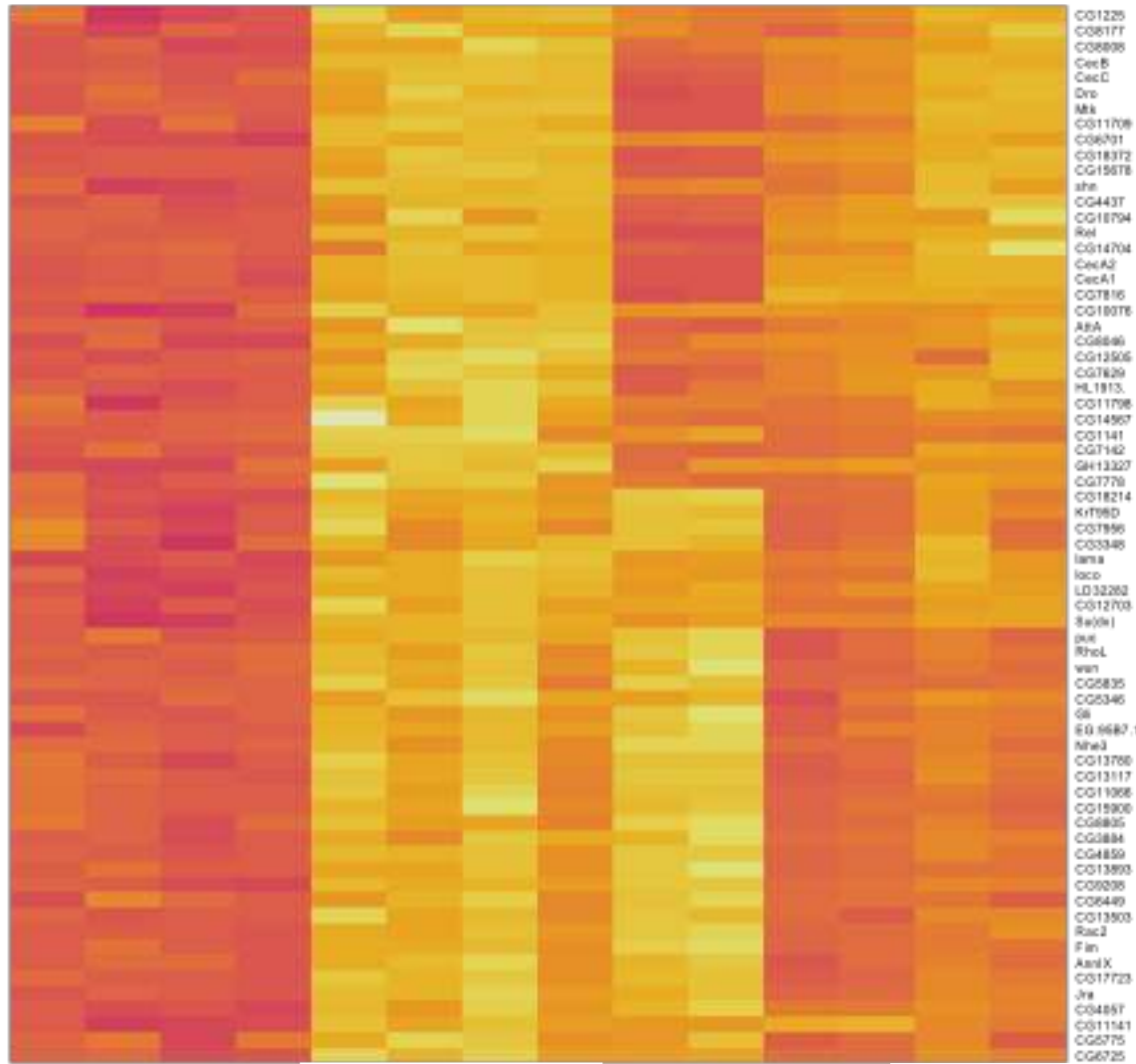
Microarray measurements are used to measure gene expression of response genes.

Genes are selectively silenced using siRNA.

Data of intervention effects can be used to reconstruct signal network.

Selected differential genes

Microarray experiments



CG1229
CG8177
CG8896
CecB
CecE
Dre
Mts
CG11709
CG8691
CG18372
CG15878
shn
CG4437
CG18794
Rel
CG14704
CecA2
CecA1
CG7816
CG18076
ABA
CG8846
CG13506
CG7829
HL1813
CG11708
CG14967
CG1141
CG7142
GH13207
CG7776
CG18214
Kit98D
CG7896
CG3348
lama
loco
LD32282
CG12703
3a00l
pmr
RhoL
van
CG5835
CG3346
G8
E0.6687.1
Nhe3
CG13760
CG13117
CG11986
CG15900
CG8835
CG2884
CG4859
CG13893
CG9236
CG5449
CG13503
Rac2
Fin
AmkX
CG17723
Jra
CG4867
CG11141
CG5775
CG6725

rel targets

∩

tak targets

∪

mkk/hep targets

controls

LPS treat.

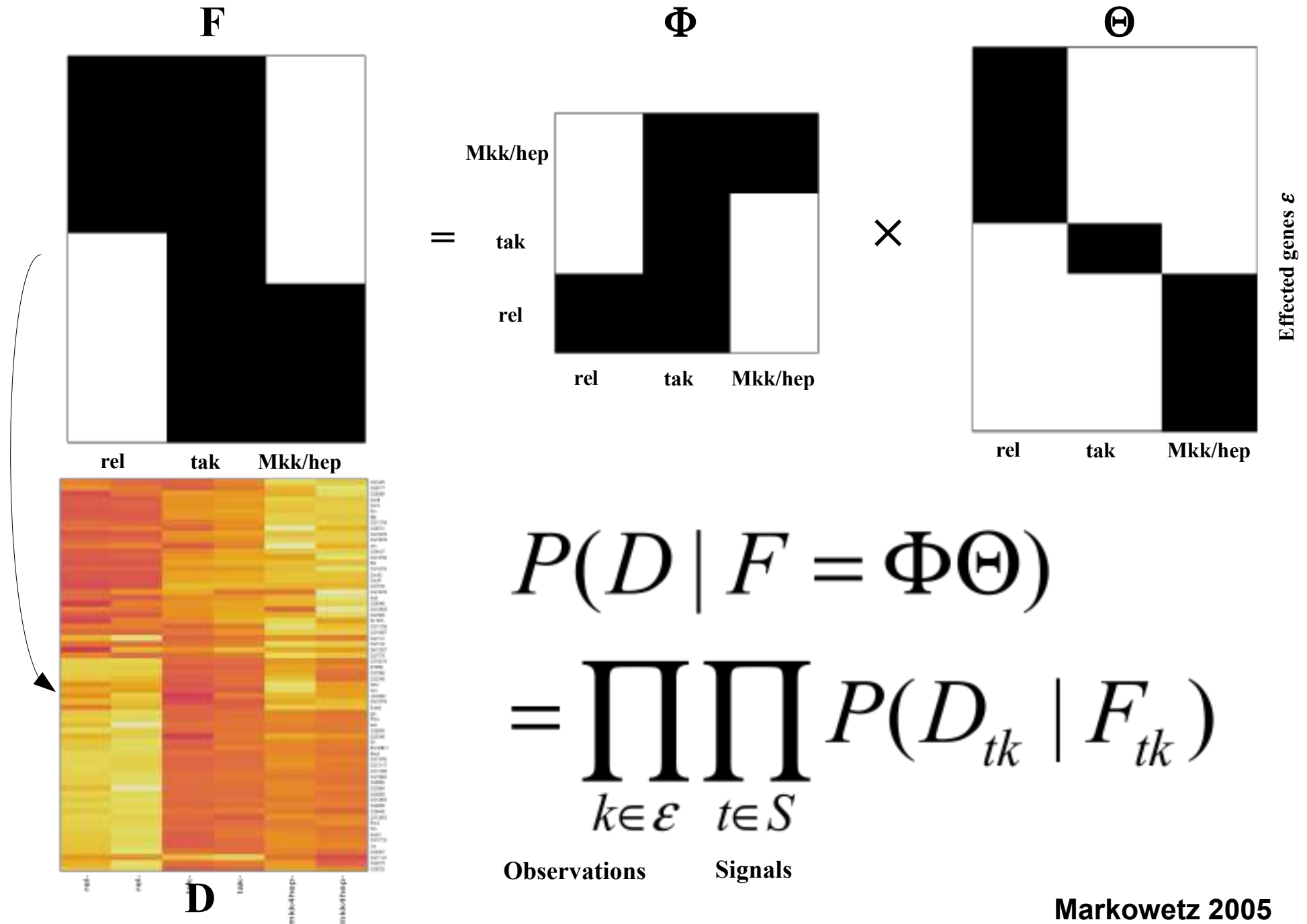
Rel

tak

mkk/hep



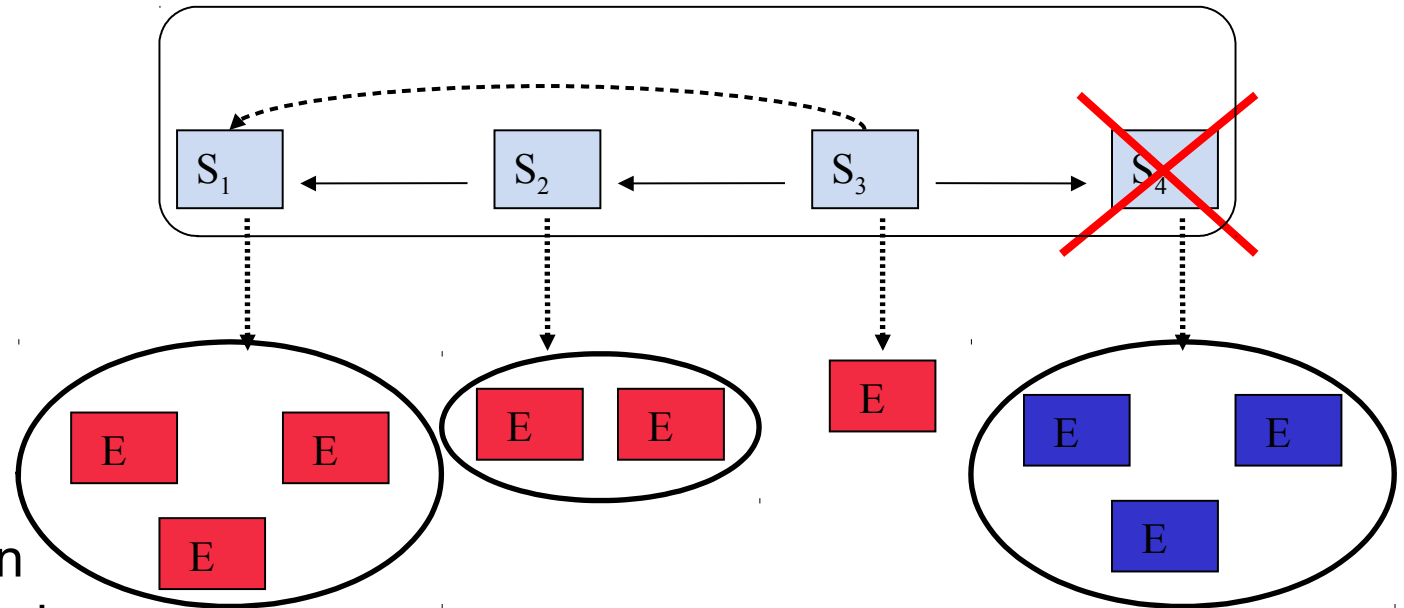
What is a Nested Effects Model



Idea of Nested Effects Models

- Distinguish between:

- S-genes (silenced genes)
- E-genes (effected genes)



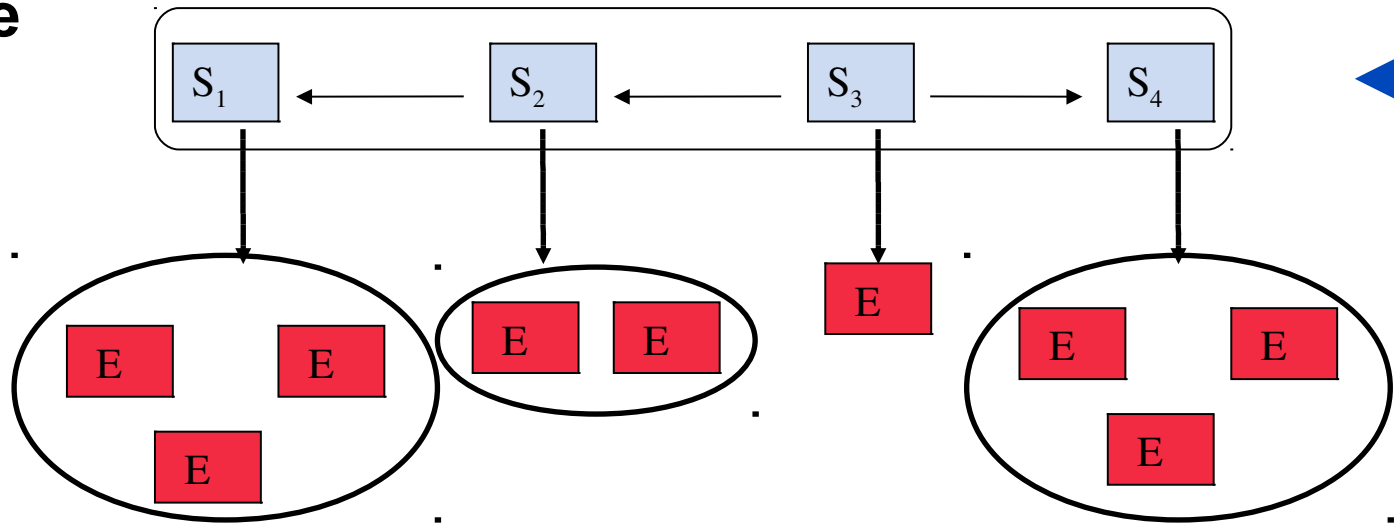
- Perform gene expression study (microarray) for each knock-down experiment.
- Network reconstruction is based on the effects seen at E-genes when specific S-genes are knocked-down

Silencing (S) Experiments

	S1	S2	S3	S4
Effected (E) Genes				

Statistical Network inference

Choose **candidate network topology** of silenced genes (S-genes)



Calculate score using Bayesian statistics (average over E-gene positions)

Likelihood model

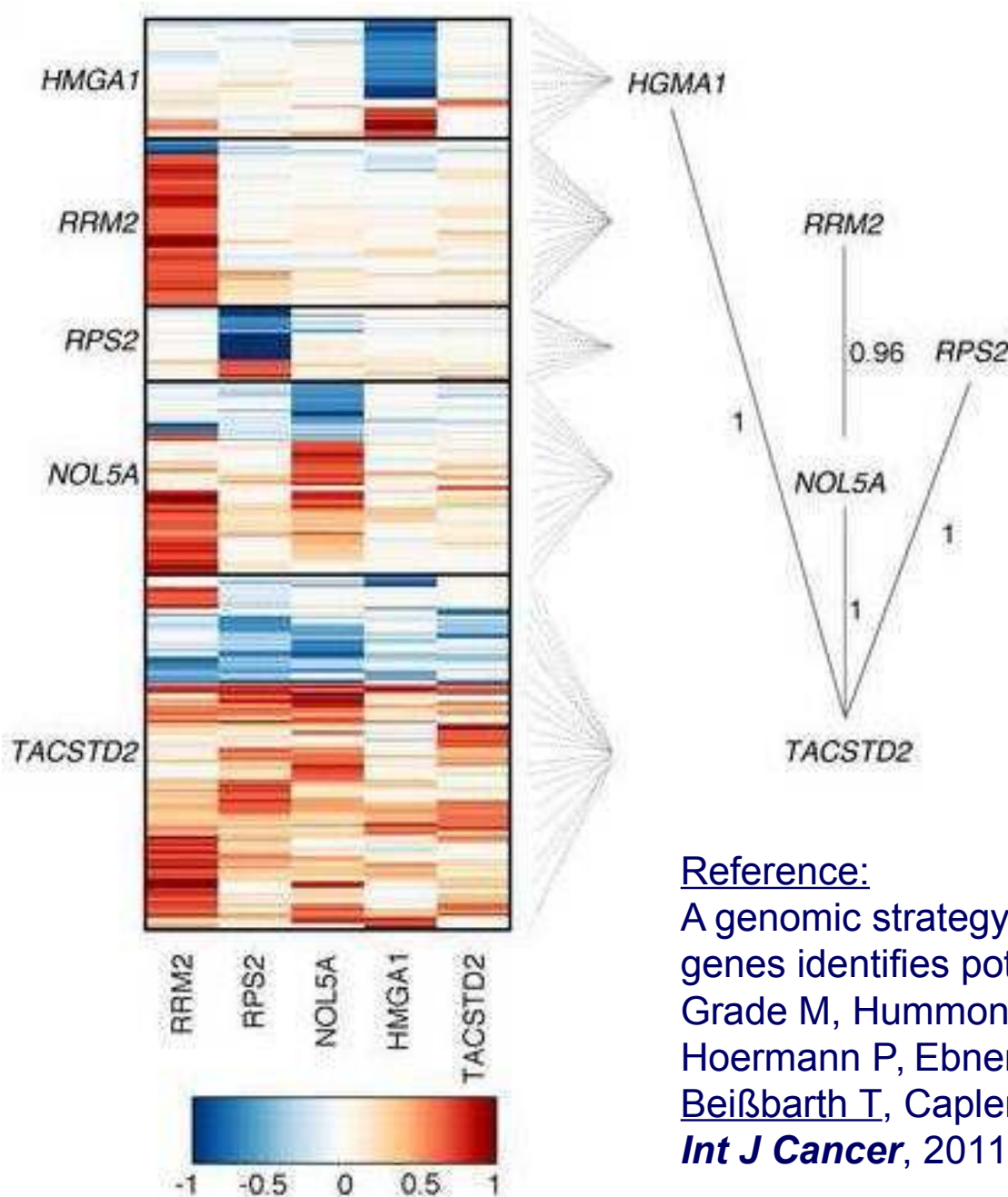
Propose different topology

Review/Method comparison:

Fröhlich H, Tresch A, Beißbarth T. *Biometrical Journal*. 51(2):304-321.

R – Package: *nem*

Example from Colorectal Cancer data-set



Knock-down of 5 genes in colorectal (SW480).
2 siRNAs per gene *
3 microarray replikates *
2 controle-siRNAs

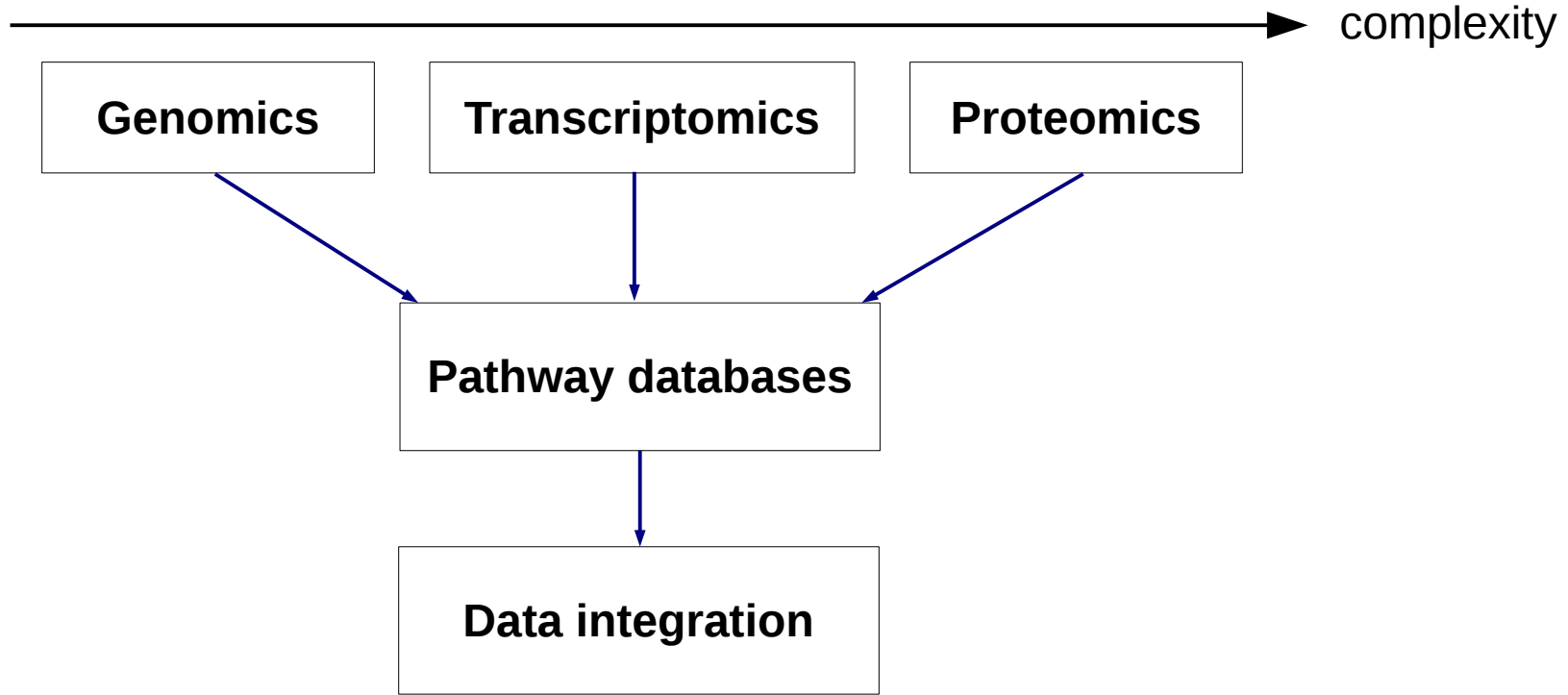
Reference:

A genomic strategy for the functional validation of colorectal cancer genes identifies potential therapeutic targets.

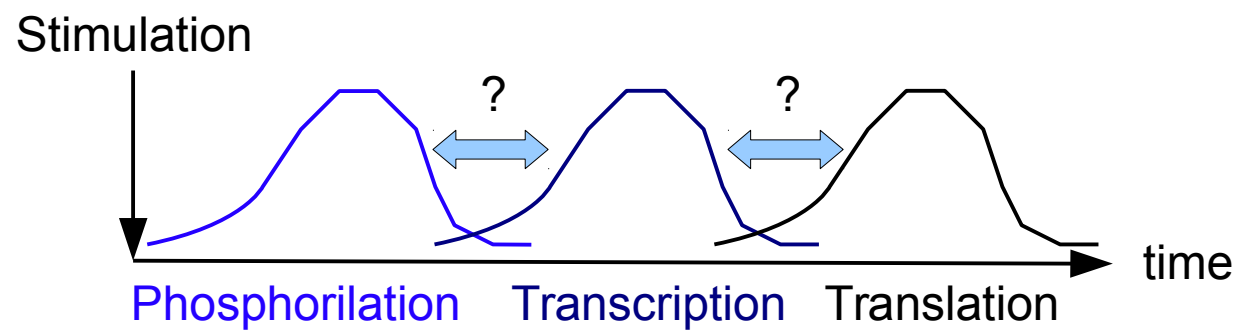
Grade M, Hummon AB, Camps J, Emons G, Spitzner M, Gaedcke J, Hoermann P, Ebner R, Becker H, Difilippantonio MJ, Ghadimi BM, Beißbarth T, Caplen NJ, Ried T.

Int J Cancer, 2011, 128(5):1069-79.

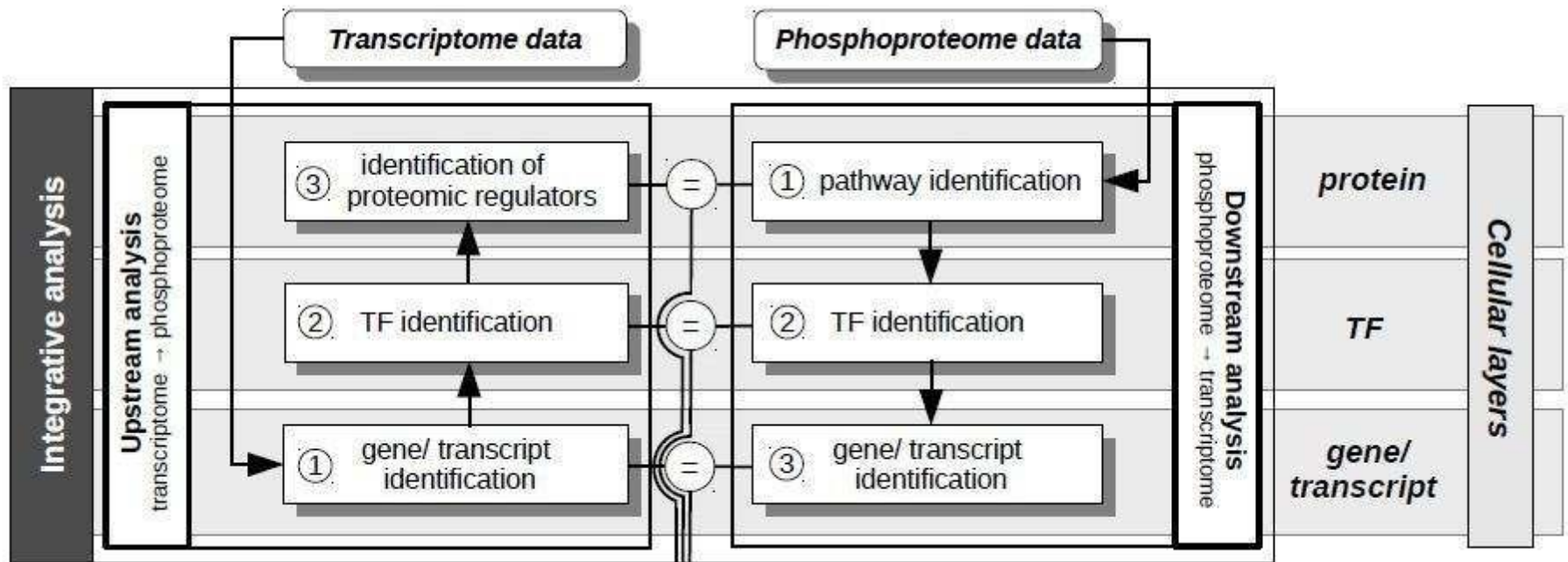
Pathway-based integration using prior pathway knowledge



- Dissolve regulation complexity
- Compare data from different platforms in a layer-specific way



Knowledge based integrative data analysis approach

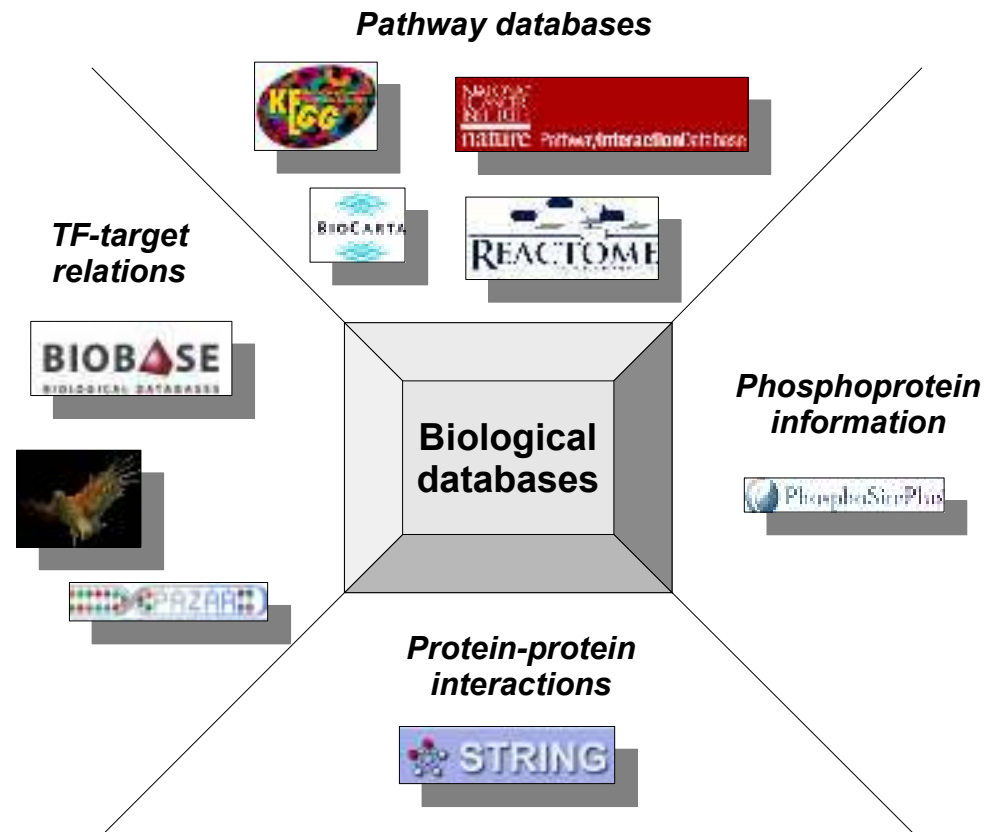


Simplifying assumption:
protein phosphorylation corresponds to downstream pathway activation

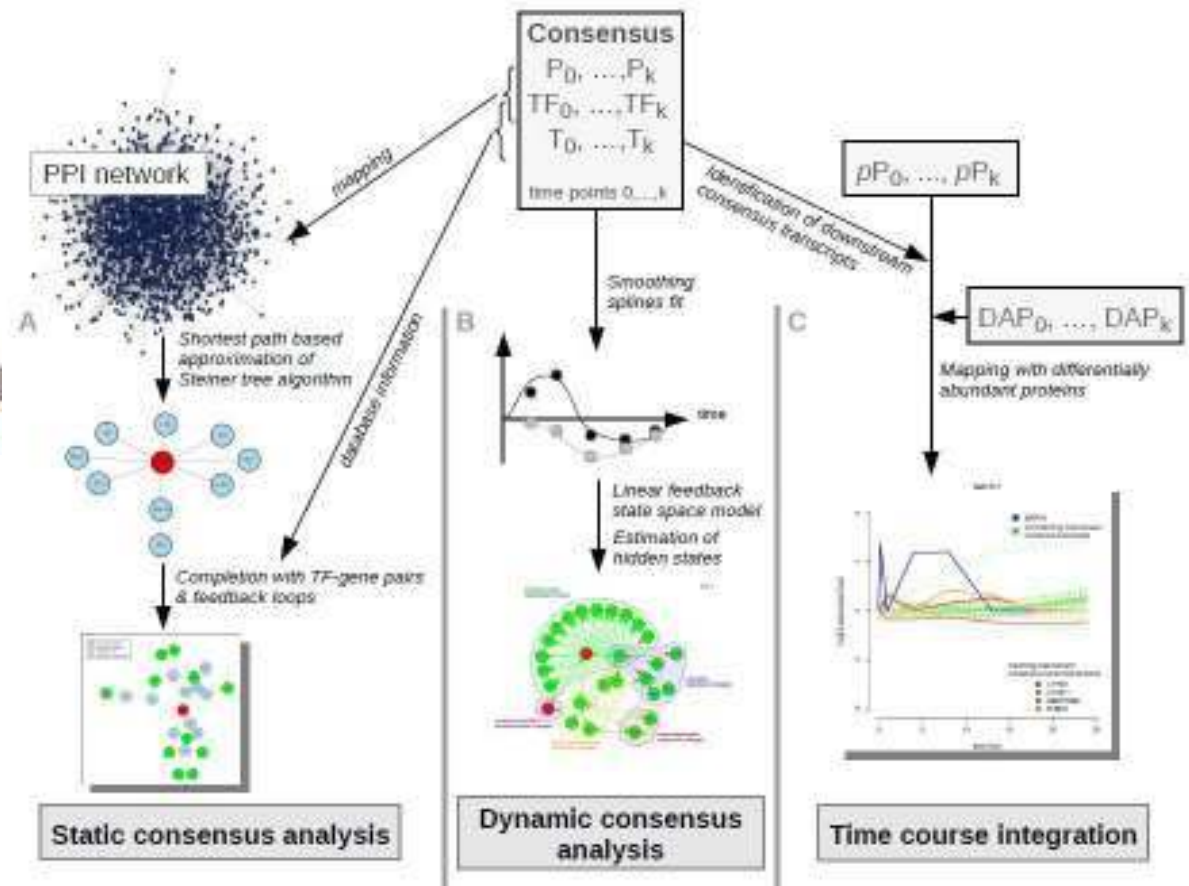
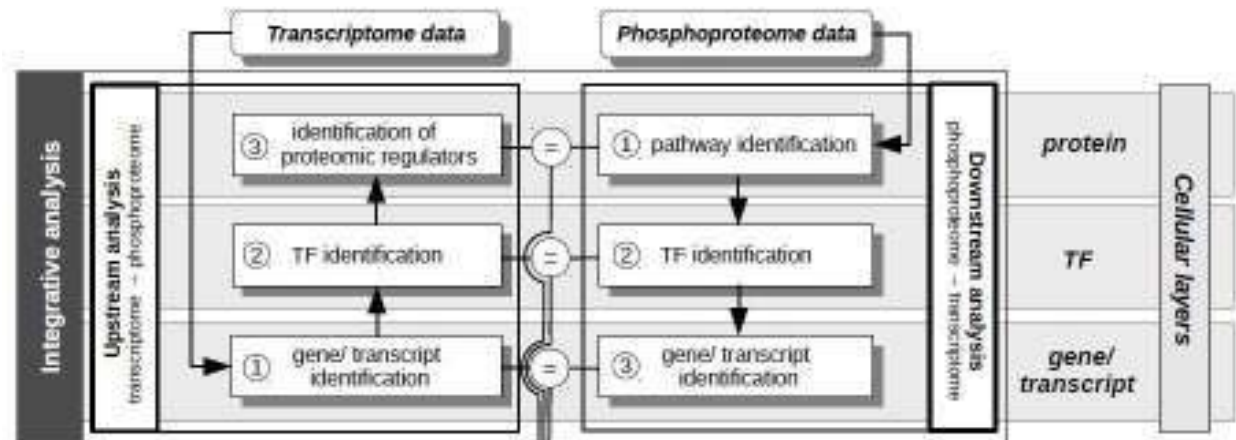
Knowledge based integrative data analysis approach

Based on public databases:

- pathway databases: KEGG, Reactome, PID, Biocarta
- TF-target databases: ChEA, Pazar, user-specified (e.g. Transfak)
- PPI-database: STRING



Knowledge based integrative data analysis approach

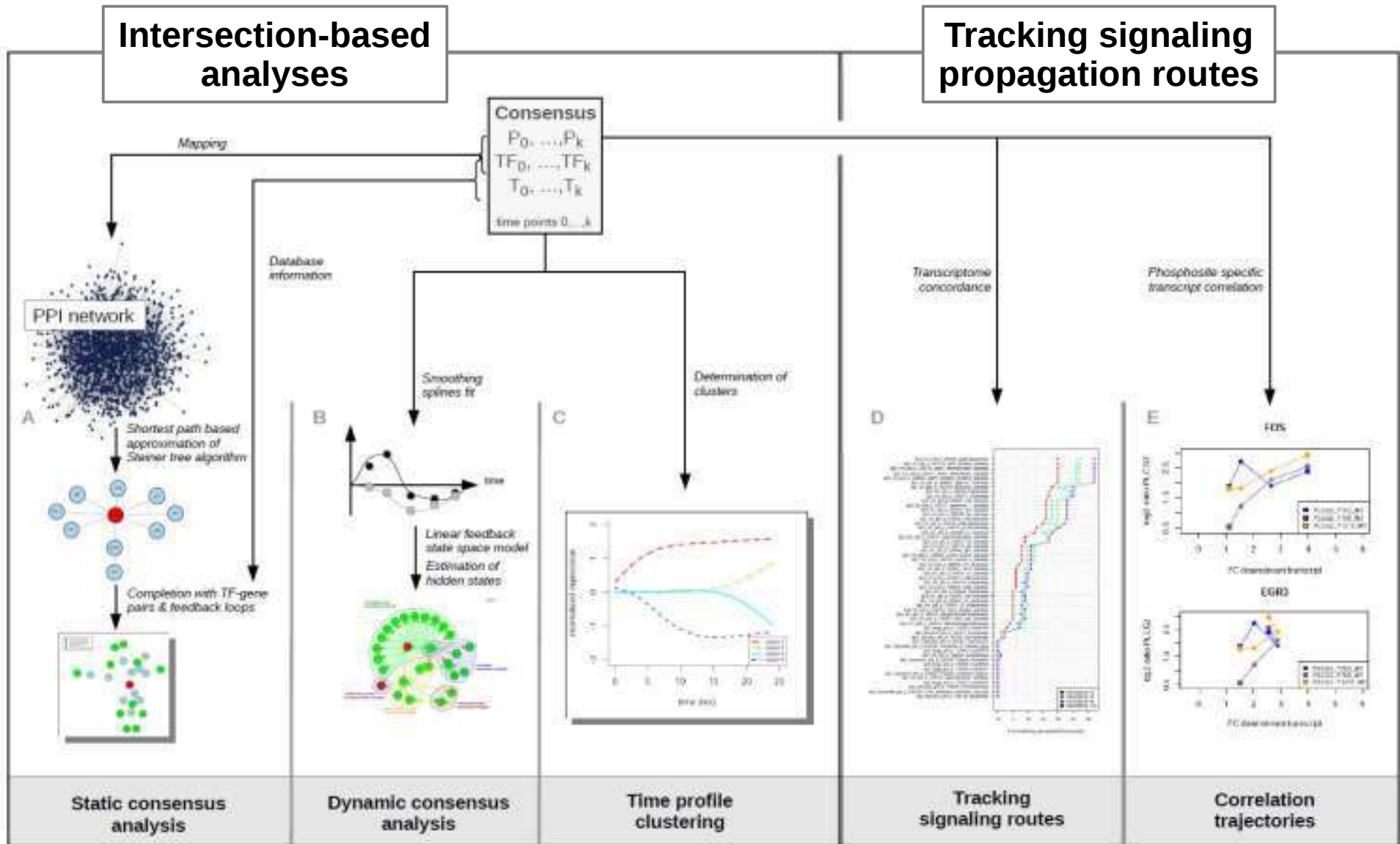


R package 'pwOmics'

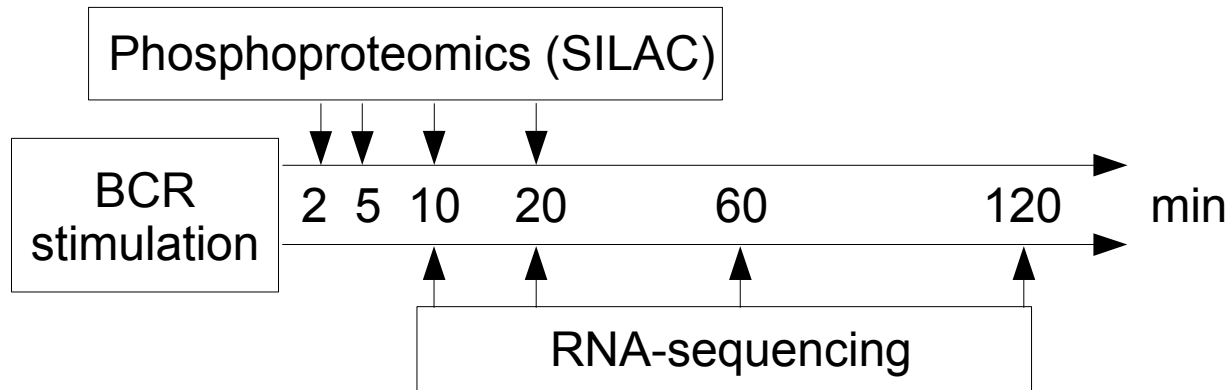


Wachter A, Beissbarth T, Bioinformatics (2015)
 Wachter A, Beissbarth T, Front. Genet. (2015)

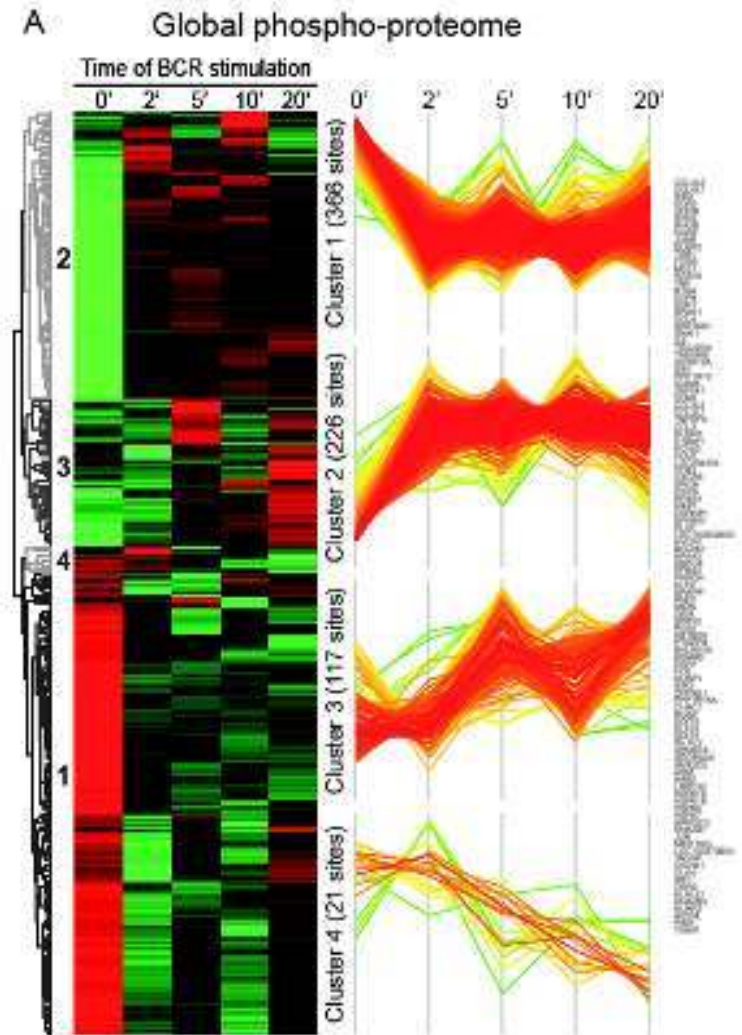
Integrative analysis



Characterization of BCR signaling in Burkitt lymphoma

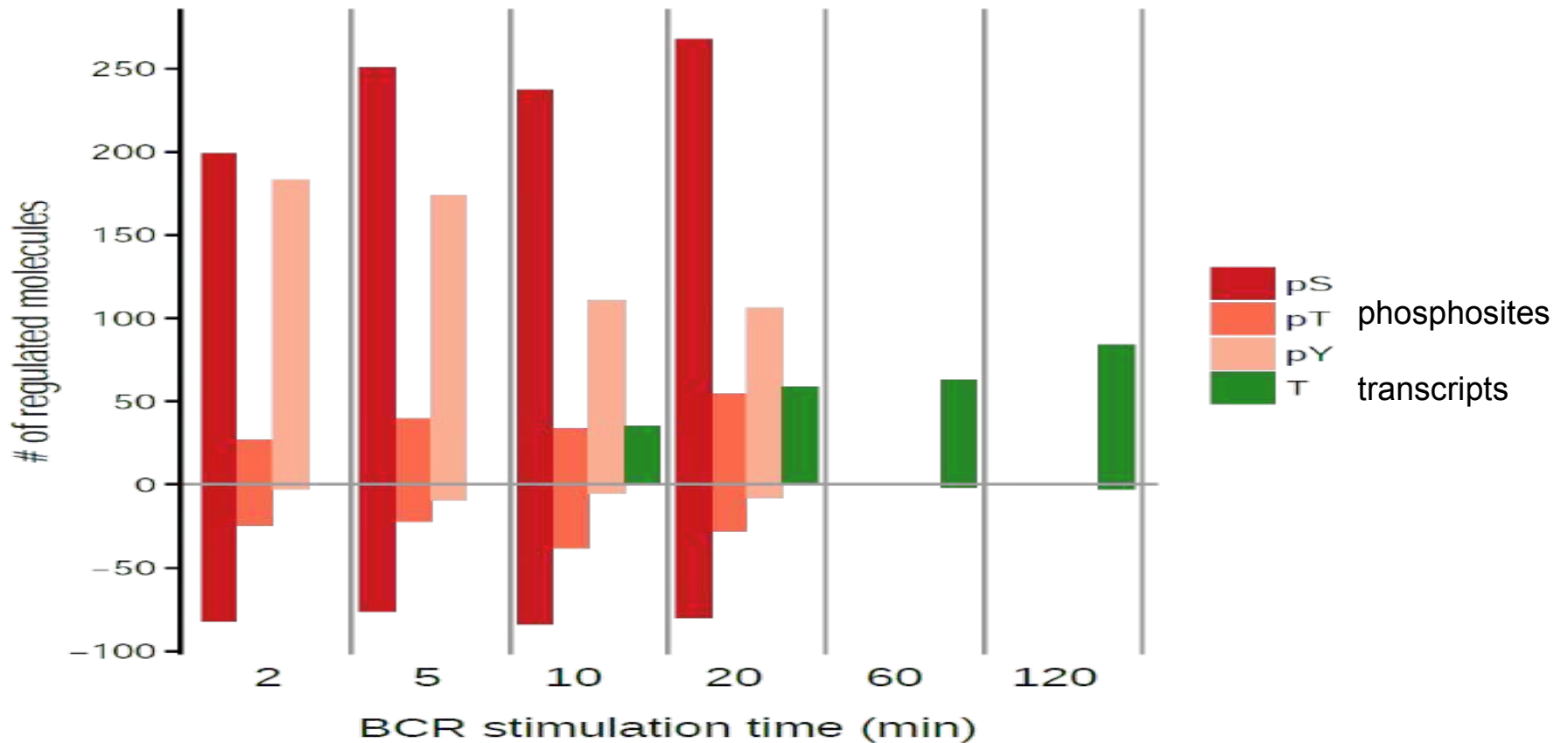
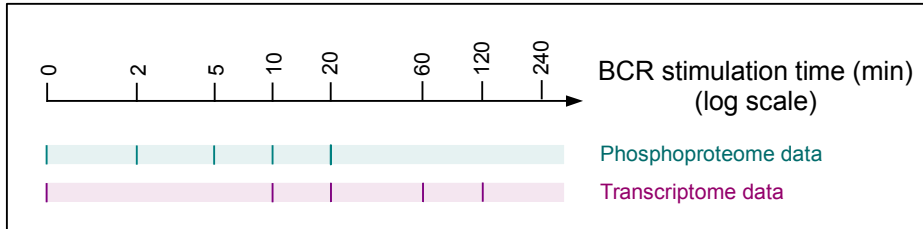


- Human cell line DG75
- Identification of BCR induced processes & druggable signaling pathways
 - Identification of BCR downstream effectors:
 - 12 % transcriptional,
 - 10 % cytoskeleton regulators
 - 9 % kinases



(Collaboration with Thomas Oellerich, Henning Urlaub)

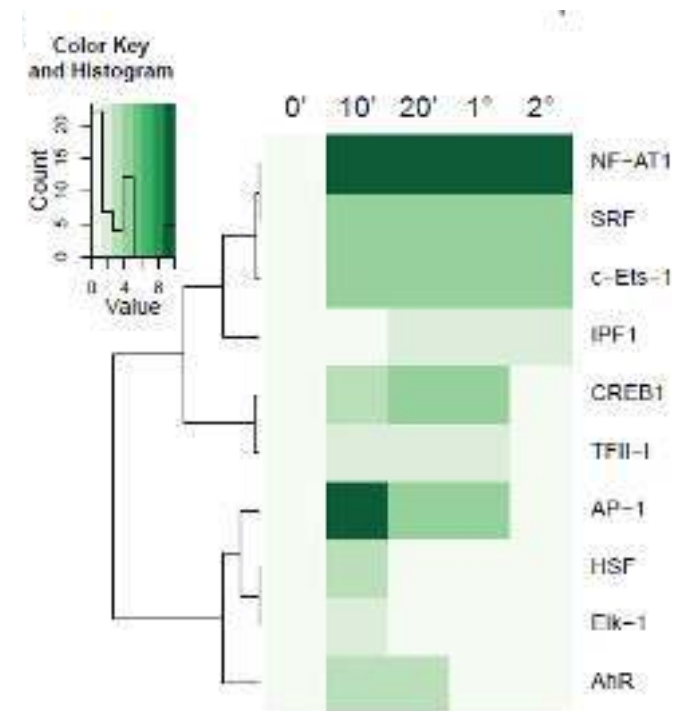
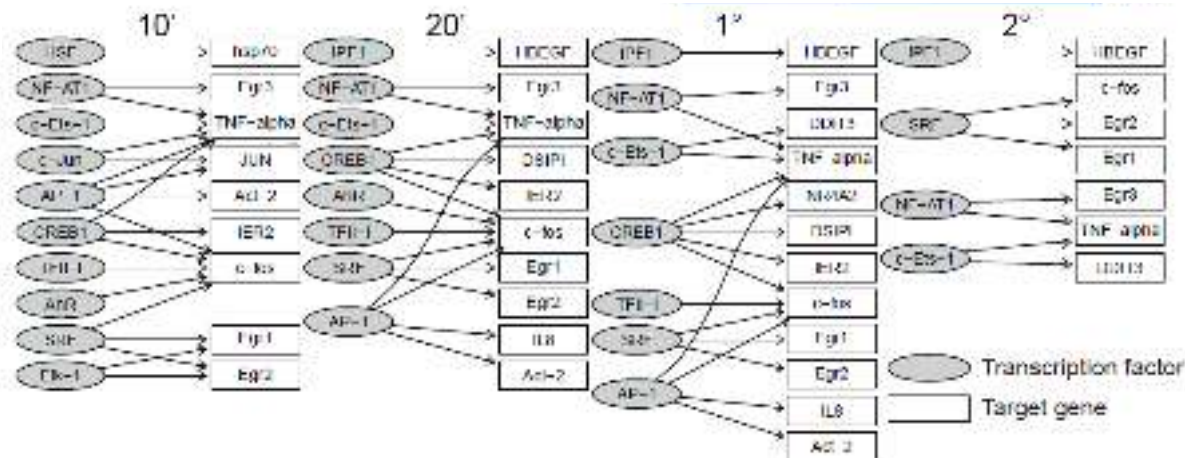
BCR stimulation of DG75 Burkitt's lymphoma cells time course data



Number of significantly regulated sites/transcripts at corresponding BCR stimulation durations. Bars above zero-level indicate upregulation numbers, bars below zero-level downregulation numbers.

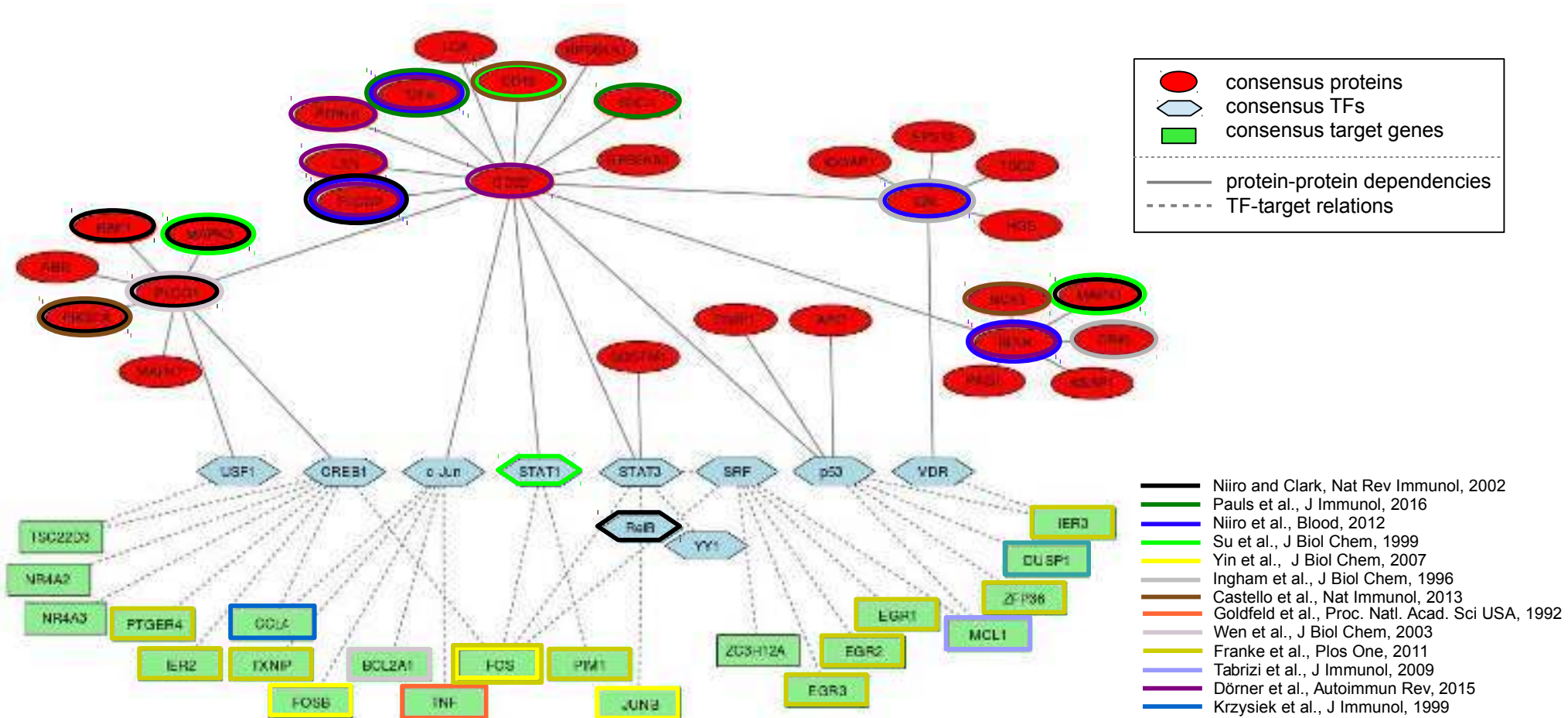
Consensus TF→target gene relations at each time point

- Influence of phosphorylation processes on transcriptome dynamics:



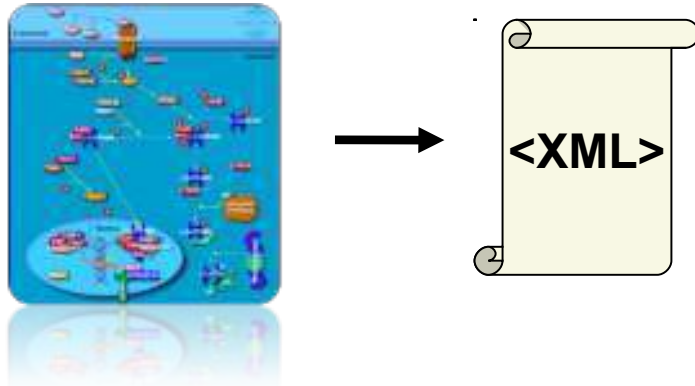
Static consensus graph

→ pooling 2, 5, 10 min phosphoproteome time points & 60, 120 min transcriptome time points



→ High concordance with literature
 → So far mostly level-specific or axis-specific investigation

Encoding Pathway Knowledge



Computationally encoding Pathways has several advantages:

- Separate data and visualization
- Ease Knowledge-Exchange
- Store and curate large amounts of data

BioPAX Ontology:

Pathway = <Interactions>*

Interaction =
<Entity> activates/inhibits <Conversion>

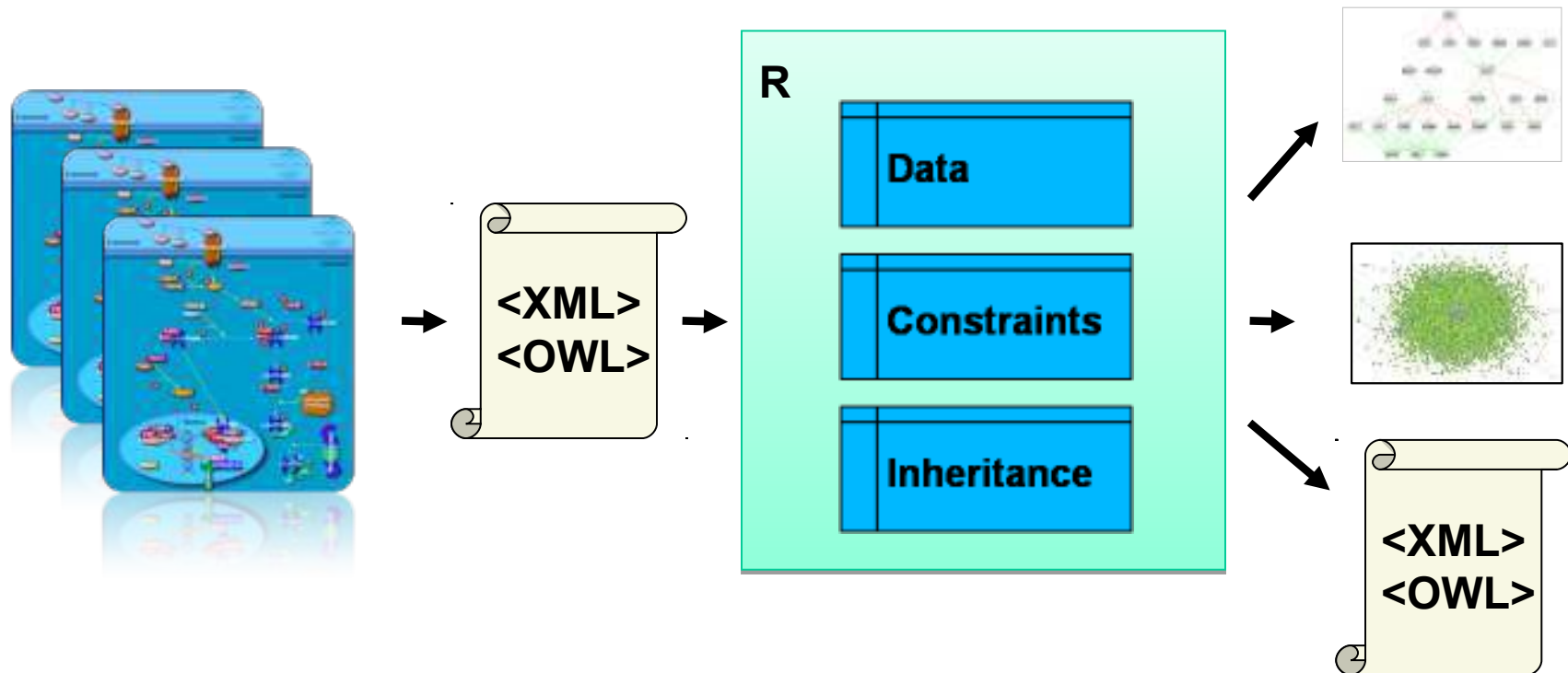
Conversions = <Entity>* → <Entity>*

Main Pathway Encoding-standards:

- Ontologies
BioPAX / SBML
- Graph Representations
KGML / GPML / SBGN

rBiopaxParser

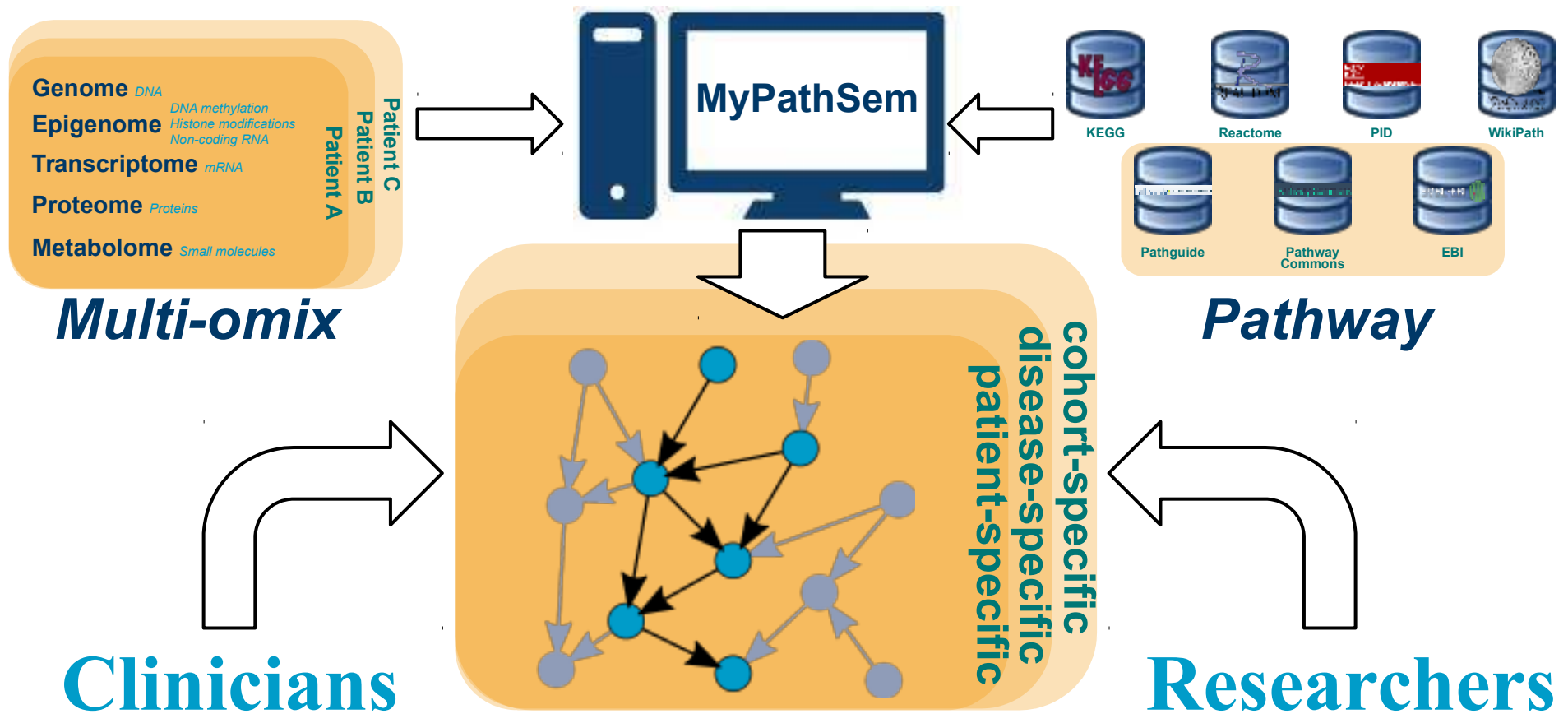
open-source software package for R:
parses BioPAX-encoded pathway databases
internally keeps original data structure
offers programmatic modification & merging
delivers PWs as graphs, matrices, BioPAX



rBiopaxParser on Bioconductor

Kramer F, Bayerlová M, Klemm F, Bleckmann A, Beissbarth T.
rBiopaxParser--an R package to parse, modify and visualize BioPAX data.
Bioinformatics. 2013 Feb 15;29(4):520-2.

MyPathSem: Generating individualised pathways



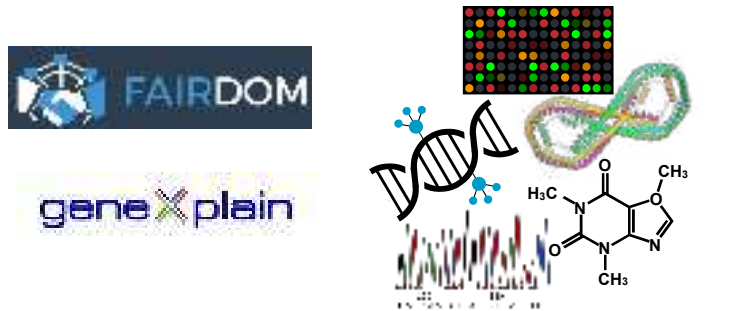
Consortium leader: Tim Beißbath

Partners: UMG (F Kramer, U Sax, E Wingender, A Bleckmann, J Gaedcke), GeneXplain (A Kel)

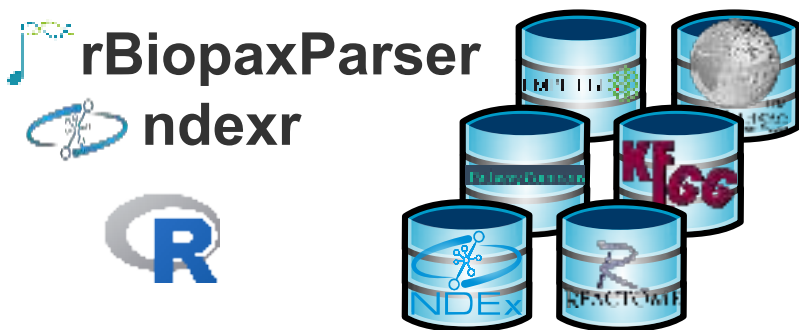
MyPathSem - Daten Integrations Infrastruktur

Integration Platform

Omics Data



Pathway knowledge



Ontology Mapping Service

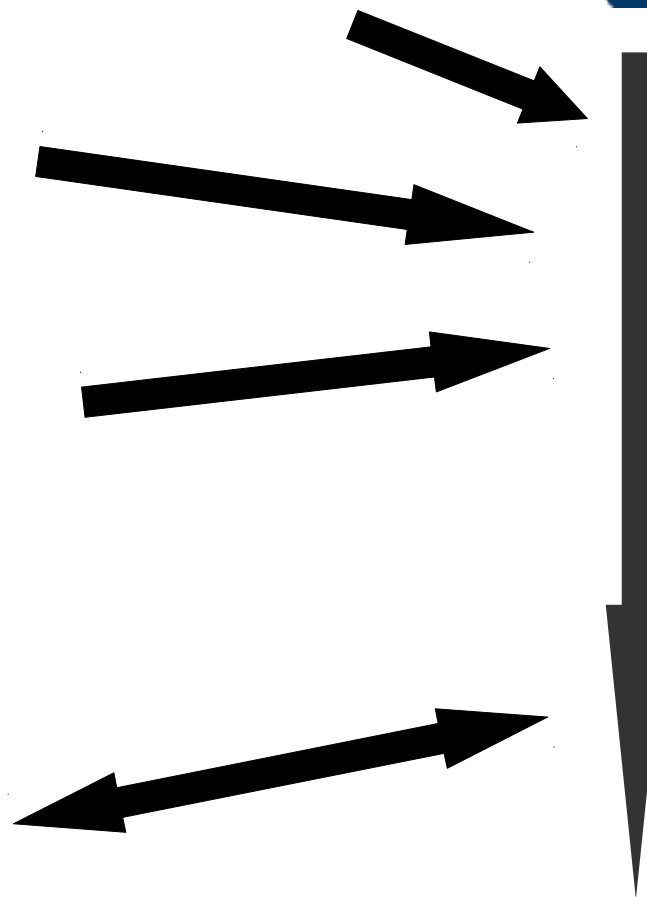
NCI Thesaurus
OBA Service
MeSH - ICD

clinical Data



Angular.js
CerebralWeb
rApache

Docker container



Acknowledgements

Statistical Bioinformatics Group, UMG

Dr. Andreas Leha

Dr. Frank Kramer

Dr. Manuel Nietert

Dr. med. Annalen Bleckmann

Dr. Michaela Biayerlova

Astrid Wachter

Julia Perera Bel

Alexander Wolff

Maren Sitte

Florian Auer

Zaynab Hammoud

Hryhorii Chereda

Felix Reinhardt



Gonsortium Projects

MetastaSys

HER2LOW

MMML-Demonstrators

MyPathSem