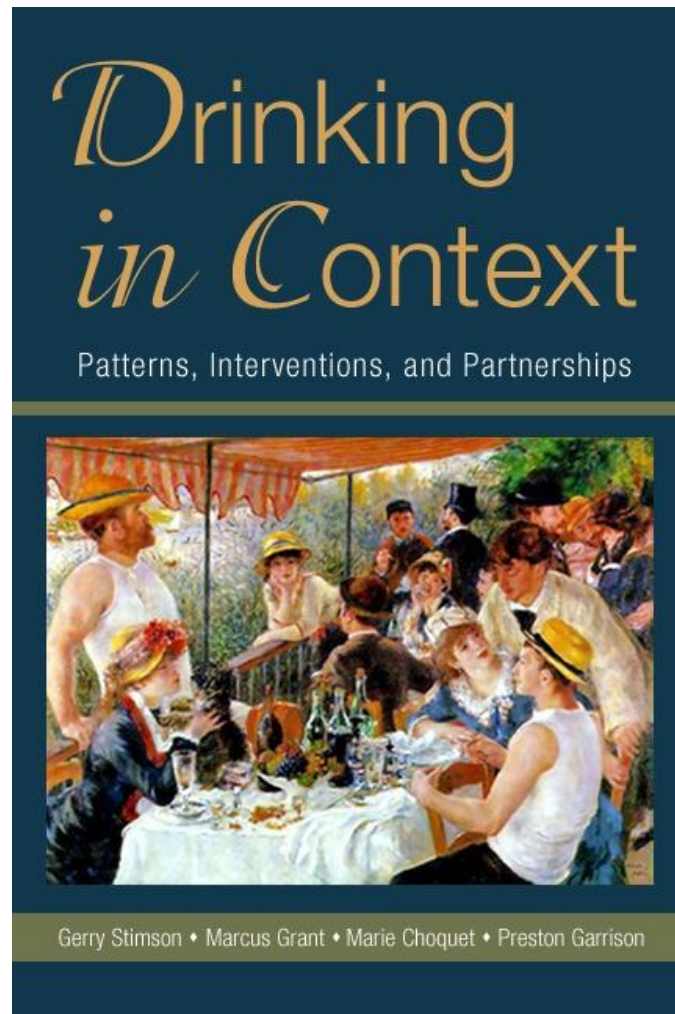




**Interpretation requires context –  
making sense out of gene lists and networks**

Philip Zimmermann, ETH Zurich / NEBION AG

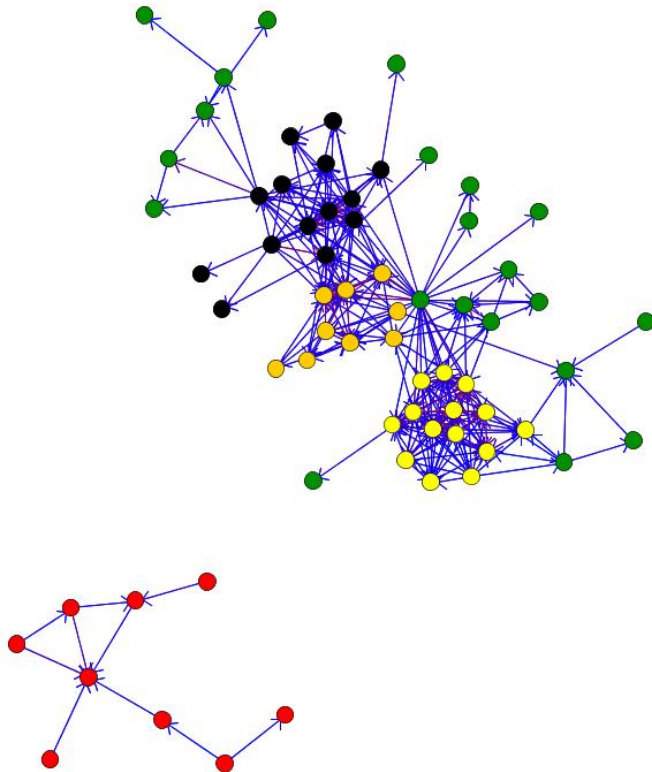
# Importance of context



# Importance of context



# Importance of context

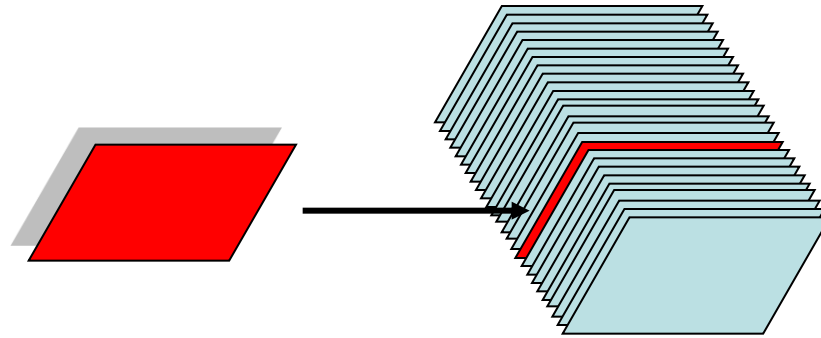


- What does this network represent?
  - a time-course?
  - a snapshot in time?
  - a specific tissue?
  - a stimulus or drug response?
  - a response to a mutation?
  - the output of a specific method?

THE CONTEXT IS OFTEN NEGLECTED!

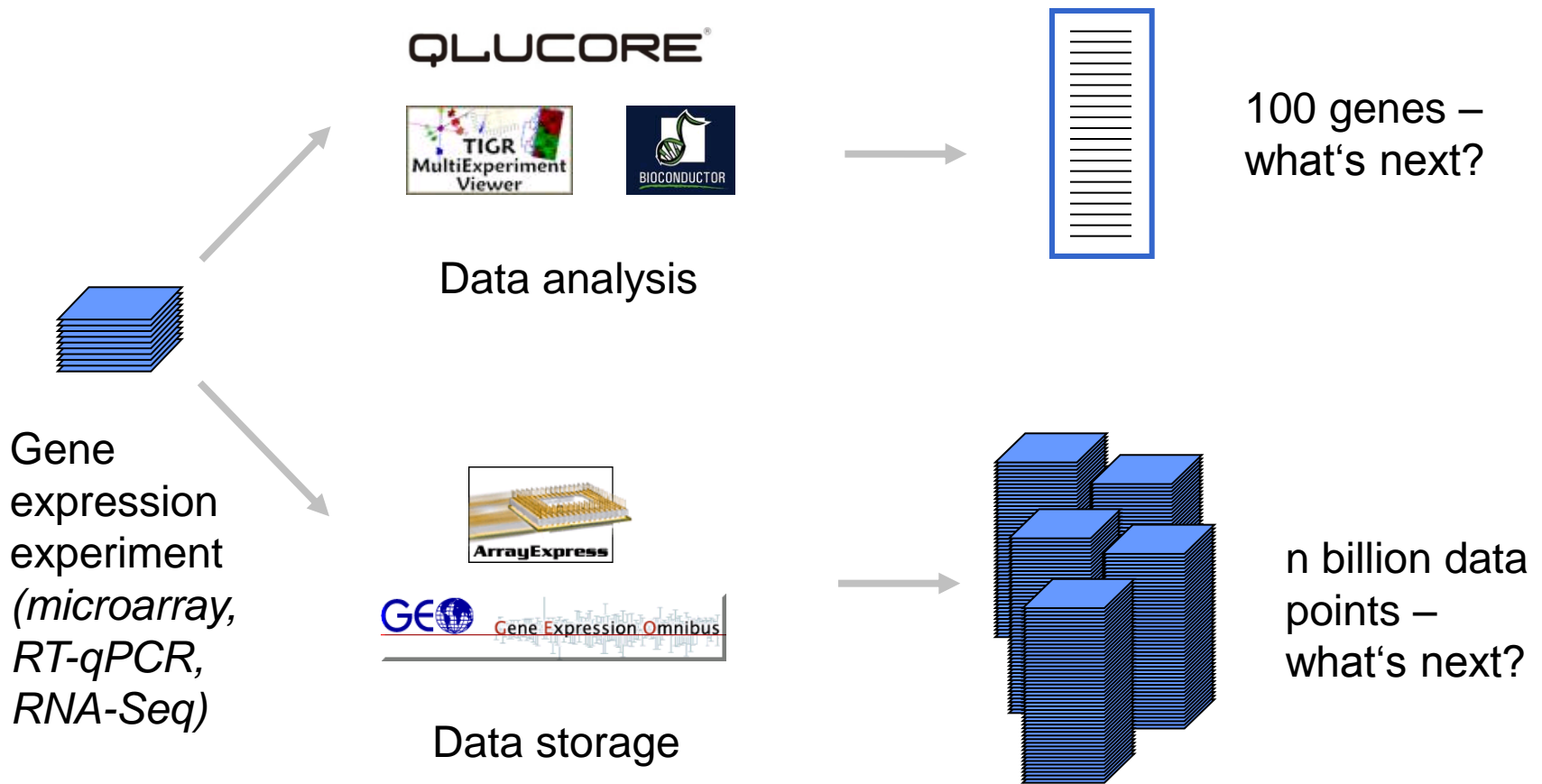
Interpretation requires  
**CONTEXT**

# Contextual meta-analysis - bringing data into context

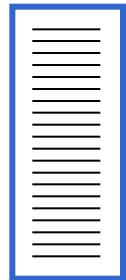


- ▶ What does my experiment tell me in its own context?
- ▶ What does my experiment tell me in the context of other experiments?
- ▶ What do my results tell me in the context of
  - Space?
  - Time?
  - Response?
  - Organisms?

# Flow of a gene expression experiment



# Interpreting gene lists



100 genes –  
what's next?

Pathway Analysis

Gene Set Enrichment Analysis

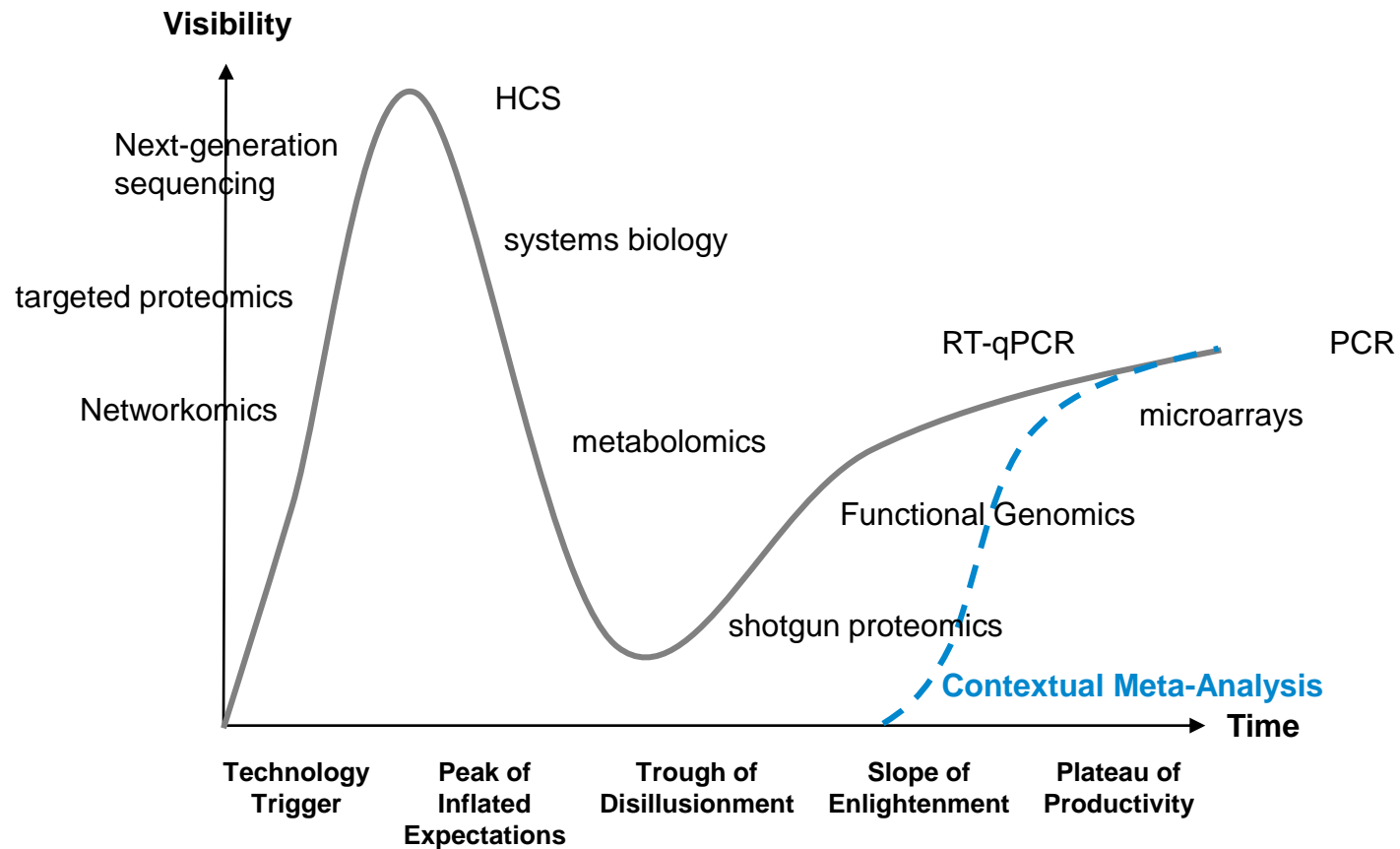
Clustering, PCA, SVM, etc.

**Contextual Meta-Analysis**

[time] [space] [response]

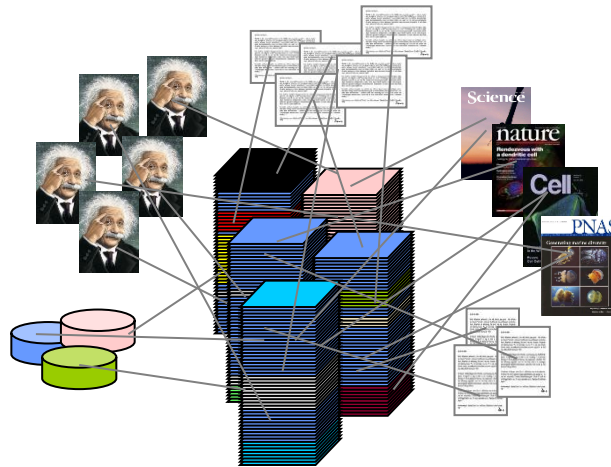


# Contextual Meta-Analysis



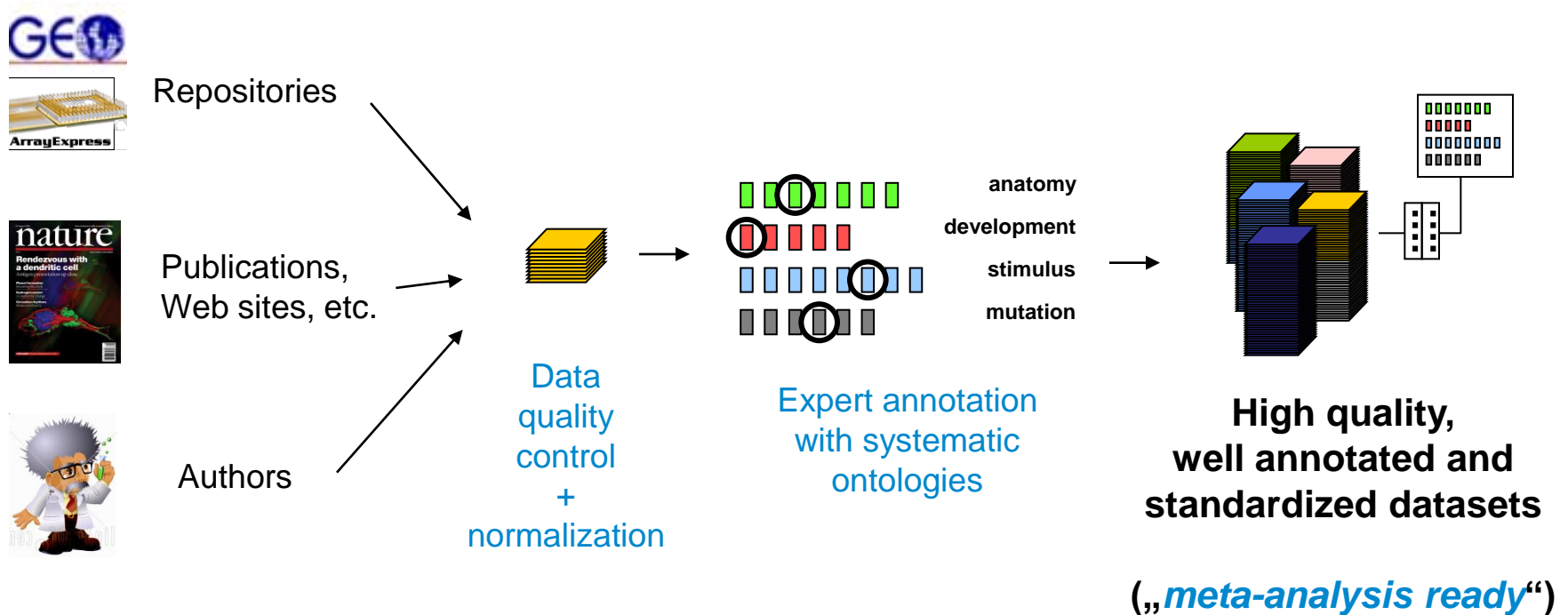
# Meta-analysis across public data is difficult

- ▶ Main bottlenecks for meta-analytical studies:
  - Different **platforms**
  - Different **protocols**, often poorly described
  - Different **normalization** schemes
  - Different approaches for **quality control**
  - Different **analytical models** used during data analysis



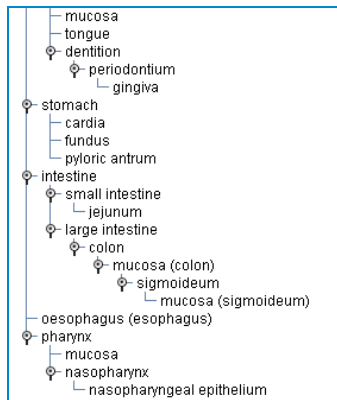
# Building a standardized gene expression compendium

Data sources ➡ QC + Norm ➡ Annotation ➡ Database



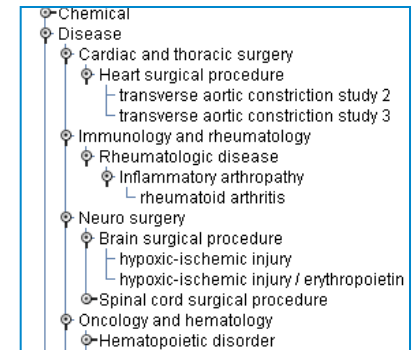
# Annotation of experimental parameters

## Anatomy ontology



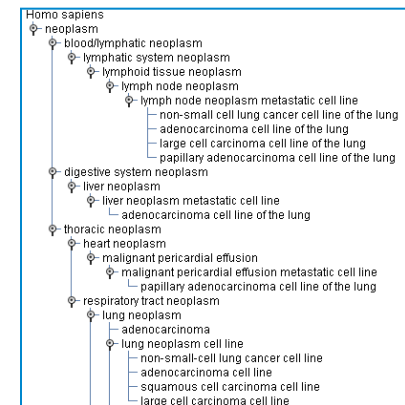
**SPACE**

## Factor ontology



**RESPONSE**

## Neoplasm ontology



**TIME**



Theiler stages

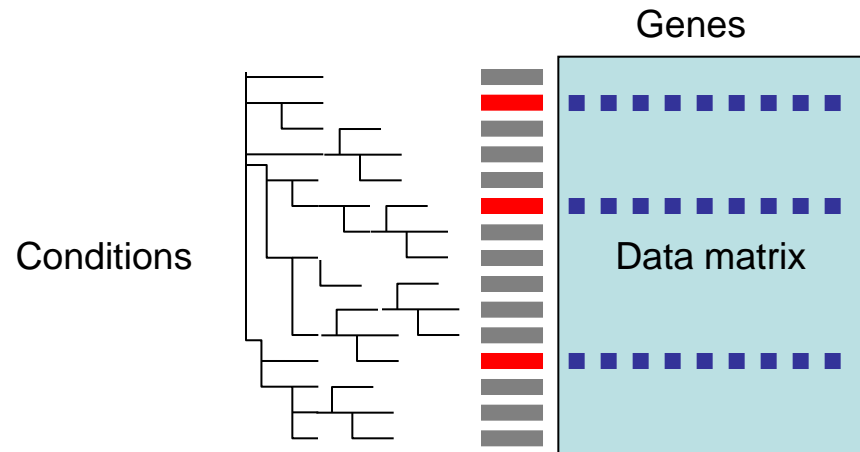
prenatal\_0-1  
 prenatal\_2-4  
 prenatal\_7-8.5  
 prenatal\_9-11  
 prenatal\_11.5-15  
 prenatal\_16-18  
 postnatal\_0  
 postnatal\_1-3  
 postnatal\_4-15  
 adult\_64-255  
 adult\_256-9999

Age (days)

# Why build new ontologies?

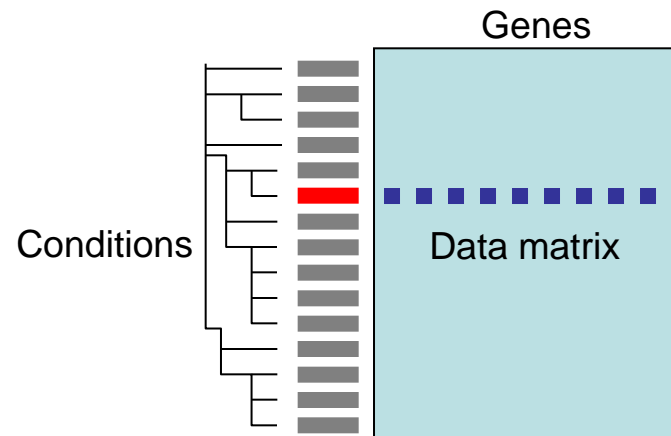
## ► Semantic ontologies

- complex
- community specific
- redundant
- abstract terminology
  - E.g. “immaterial anatomical entity”, “Cardinal body part”
- deep trees



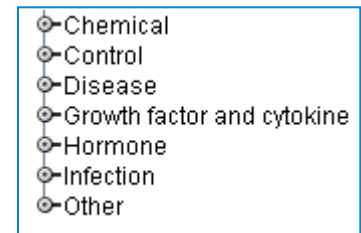
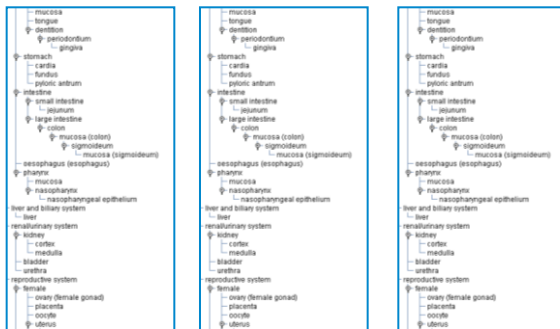
## ► Application ontologies

- minimize redundancies
- minimize abstract terminology
- shallow trees

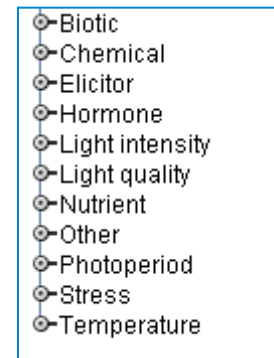
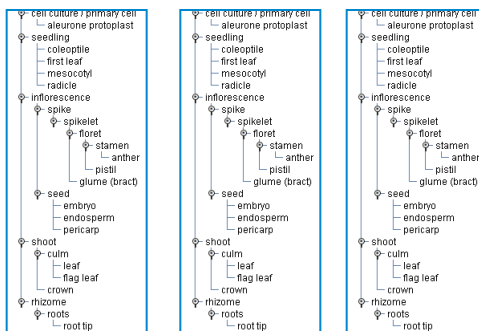


# Cross-species compatibility

## ► Human - Mouse - Rat



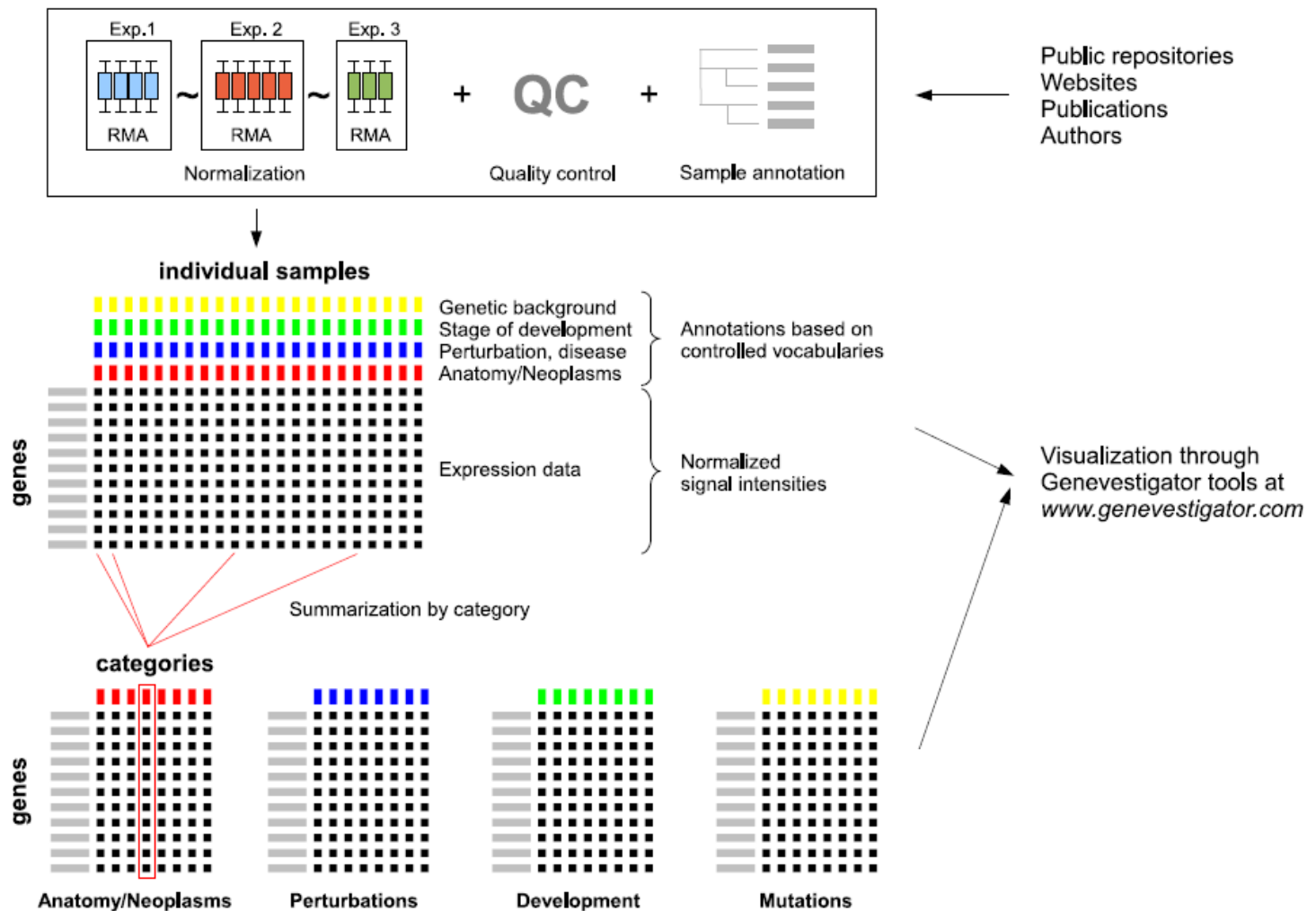
## ► Plant species



# Data Content in Genevestigator

Organism	Studies	Samples	“Conditions”	“Genotypes”	“Anatomy”	“Neoplasm”
Human	496	25'270	1'023	476	234	734
Mouse	397	8'104	515	398	230	
Rat	177	4'787	590	53	123	
Drosophila	123	2'391	209	260	48	
Arabidopsis	374	6'290	775	516	63	
Barley	27	1'098	82	36	21	
Rice	13	166	17	14	14	
Wheat	30	915	557	103	19	
Soybean	12	3'091	55	317	26	
Maize	9	249	20	21	8	
Yeast	63	1'798	176	157	4	
E. coli	21	626	119	90	0	

# Meta-profiles





# RefGenes – now published in BMC Genomics

The screenshot shows a web browser window displaying the abstract of a paper in BMC Genomics. The browser's address bar shows the URL: <http://www.biomedcentral.com/1471-2164/12/156/abstract>. The page header includes the BMC Genomics logo, an Impact Factor of 3.76, and navigation links. The article title is "RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization". The authors listed are Tomas Hruz, Markus Wyss, Mylene Docquier, Michael W Pfaffl, Sabine Masanetz, Lorenzo Borghi, Phebe Verbrugghe, Luba Kalaydjieva, Stefan Bleuler, Oliver Laule, Patrick Descombes, Wilhelm Gruissem, and Philip Zimmermann. The article is published in BMC Genomics 2011, 12:156, with a DOI of 10.1186/1471-2164-12-156. The abstract is provisional. The background section states that RT-qPCR is a sensitive method for gene expression quantification, and the results show that a subset of stable genes exists that has smaller variance than commonly used reference genes. The results section describes the use of the Genevestigator database to analyze transcriptomes of several organisms.

Abstract | RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization - Mozilla Firefox

File Edit View History Bookmarks Tools Help

[http://www.biomedcentral.com/1471-2164/12/156/abstract](#)

BMC Genomics

IMPACT FACTOR 3.76

Log on/register  
Feedback | Support | My details

home | journals A-Z | subject areas | advanced search | authors | reviewers | libraries | about | my BioMed Central

Methodology article **Open Access**

## RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization

Tomas Hruz ✉, Markus Wyss ✉, Mylene Docquier ✉, Michael W Pfaffl ✉, Sabine Masanetz ✉, Lorenzo Borghi ✉, Phebe Verbrugghe ✉, Luba Kalaydjieva ✉, Stefan Bleuler ✉, Oliver Laule ✉, Patrick Descombes ✉, Wilhelm Gruissem ✉ and Philip Zimmermann ✉

BMC Genomics 2011, **12**:156 doi:10.1186/1471-2164-12-156

Published: 21 March 2011

### Abstract (provisional)

#### Background

RT-qPCR is a sensitive and increasingly used method for gene expression quantification. To normalize RT-qPCR measurements between samples, most laboratories use endogenous reference genes as internal controls. There is increasing evidence, however, that the expression of commonly used reference genes can vary significantly in certain contexts.

#### Results

Using the Genevestigator database of normalized and well-annotated microarray experiments, we describe the expression stability characteristics of the transcriptomes of several organisms. The results show that a) no genes are universally stable, b) most commonly used reference genes yield very high transcript abundances as compared to the entire transcriptome, and c) for each biological context a subset of stable genes exists that has smaller variance than commonly used reference genes or genes that were selected for their stability across all conditions.

#### Viewing options:

- Abstract
- PDF (2.4MB)

#### Associated material:

- Readers' comments
- PubMed record

#### Related literature:

- Articles citing this article on PubMed Central
- Other articles by authors on Google Scholar on PubMed
- Related articles/pages on Google on Google Scholar on PubMed

#### Tools:

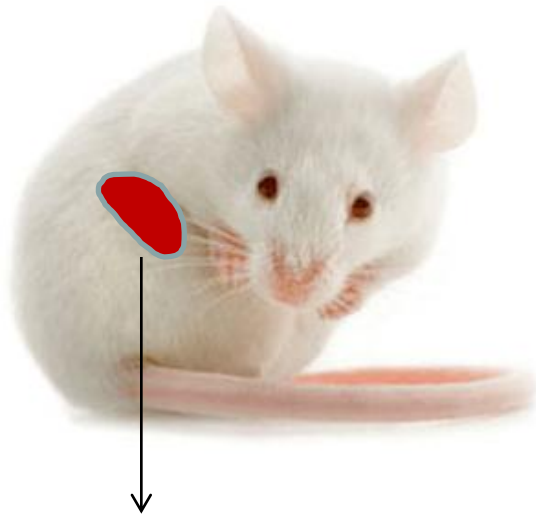
- Download citation(s)
- Email to a friend
- Order reprints
- Post a comment

#### Post to:

- Citeulike
- Connotea
- Del.icio.us
- Facebook
- Mendeley

# RT-qPCR normalization – current concepts

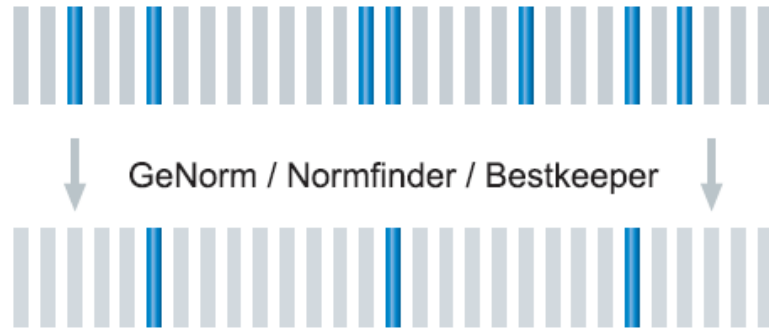
Which genes are most stably expressed in mouse liver?



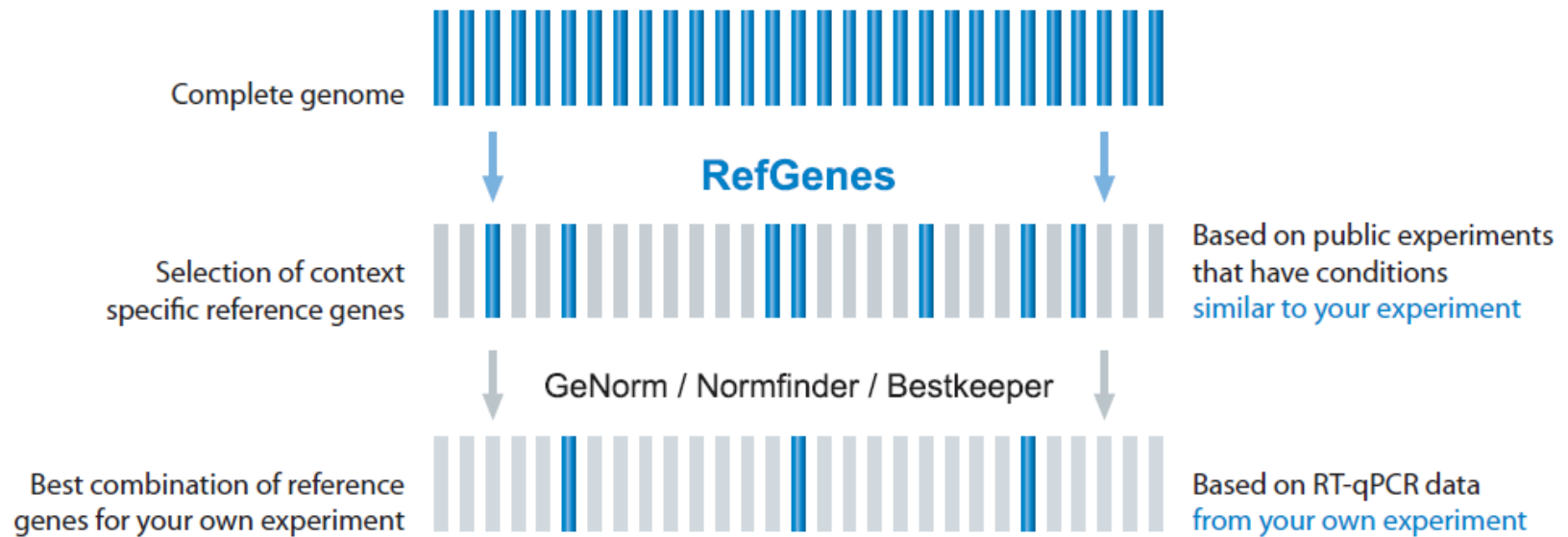
Mouse liver samples

OpenArray™ Human Housekeeping Panel Kit

Human Endogenous Control Gene Panel



# RT-qPCR normalization – concept of RefGenes



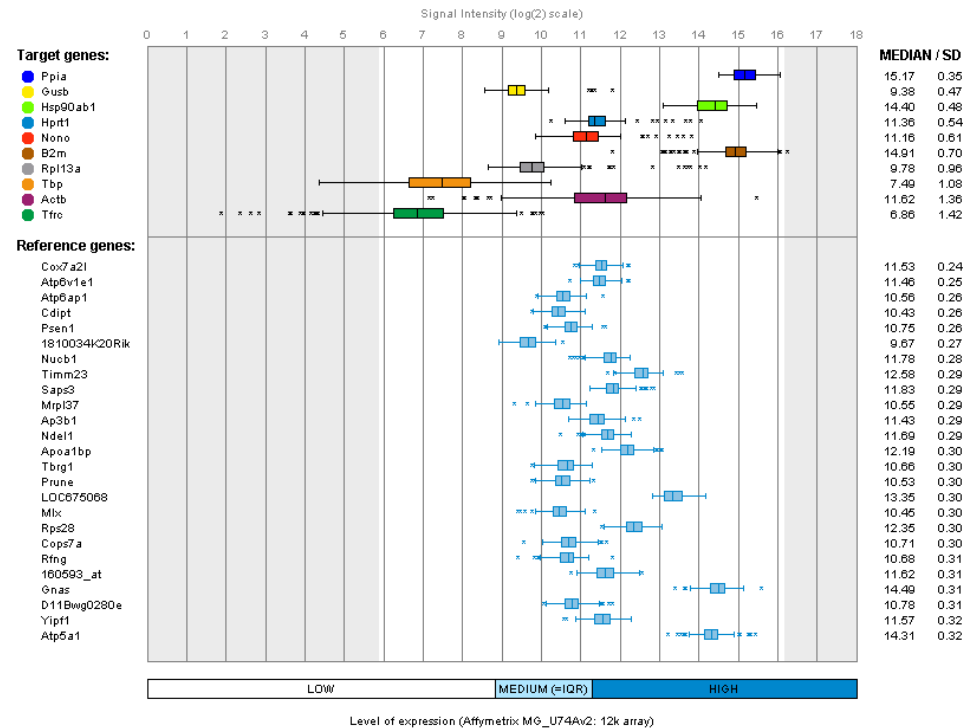
# RT-qPCR normalization – beyond gene panels



Selected microarray dataset:

197 mouse liver samples

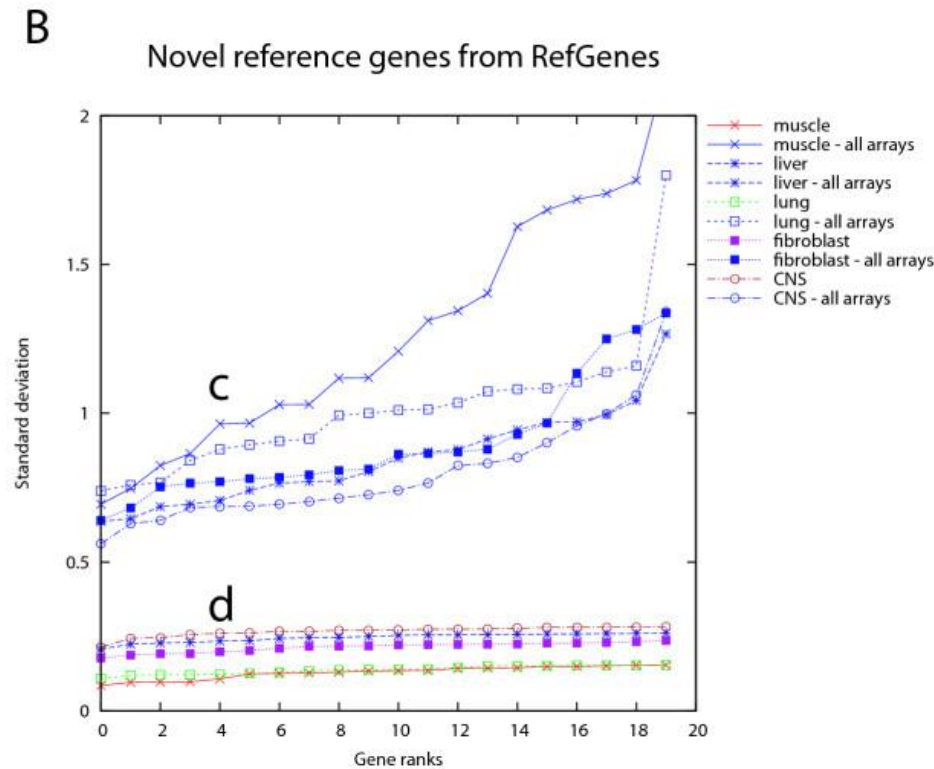
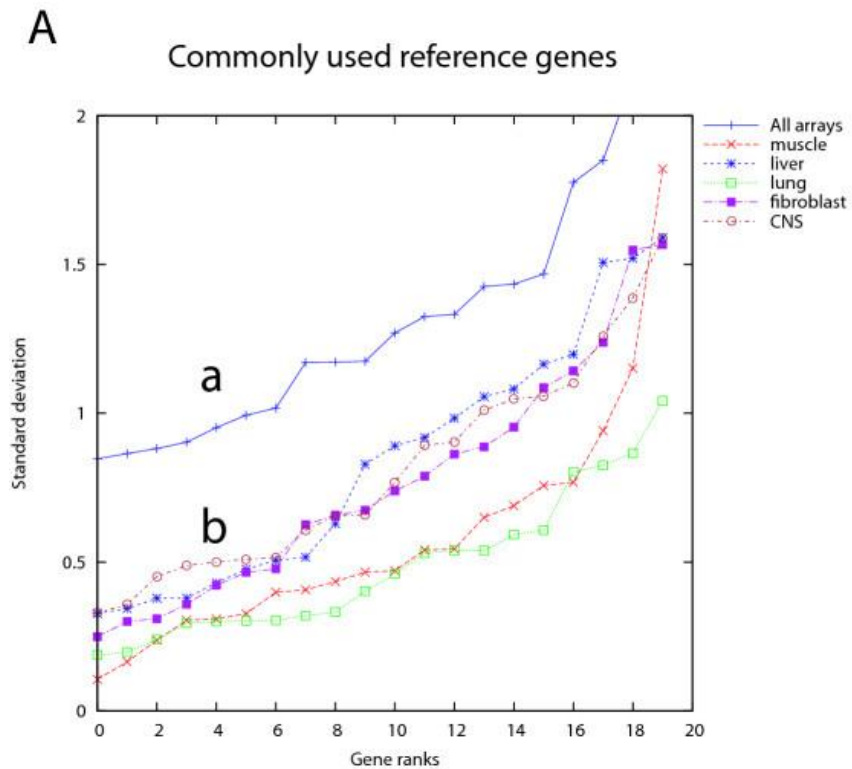
## Example of mouse liver



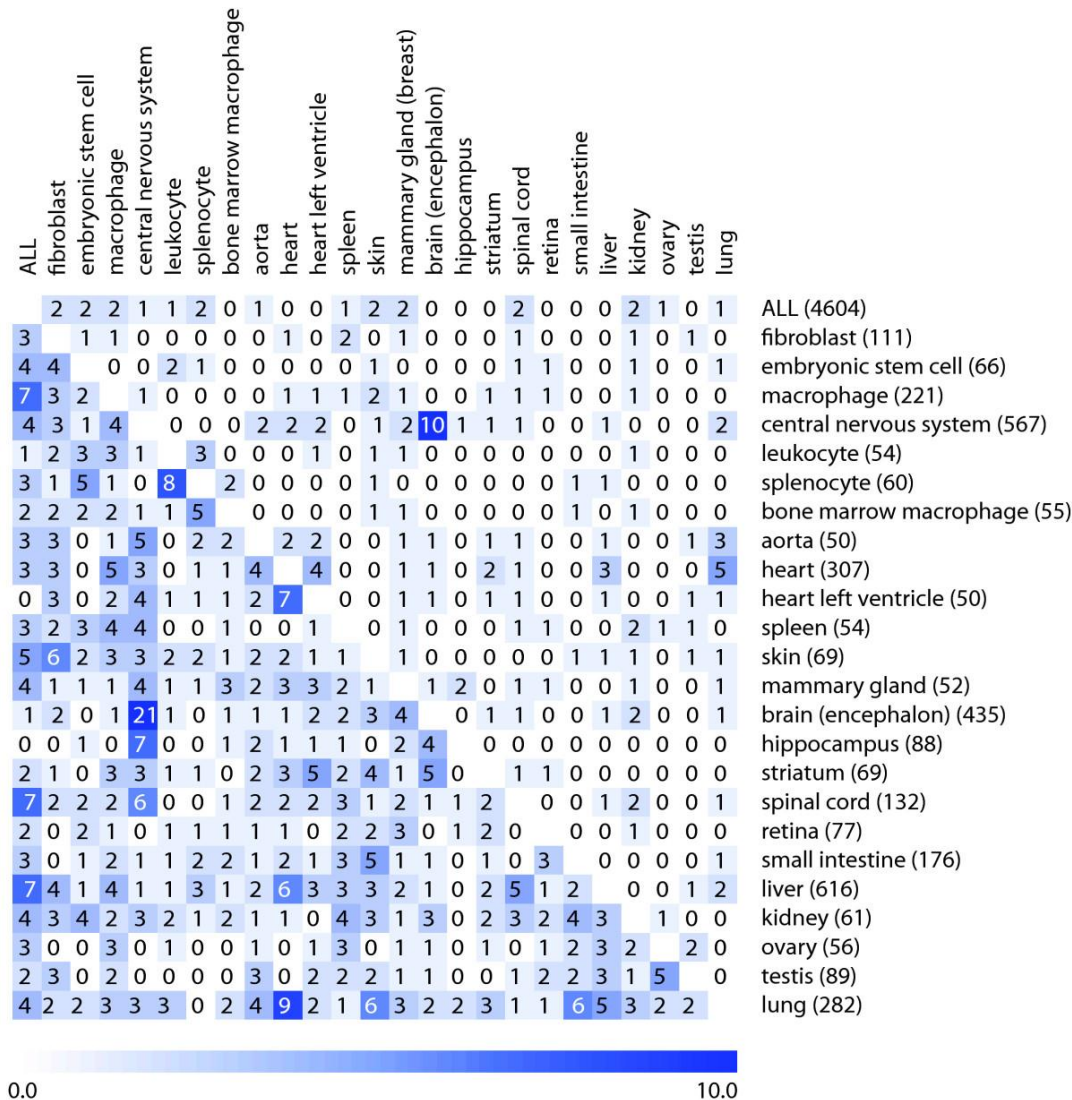
# Validation of RefGenes approach

		Rank of the average expression stability values of remaining reference genes								Mean values for		
	Samples	1	2	3	4	5	6	7	8	Top 3 genes	RefGenes candidates	Common ref. genes
<b>SPECIFIC TISSUES</b>												
<b>Mouse liver</b>	16	GAK	SRP72	mRpL16	VPS4A	ACTB	HPRT	GAPDH	TUBB			
GeNorm (Avg M)		0.15	0.15	0.17	0.19	0.21	0.24	0.27	0.30	0.16	0.17	0.26
Mean Ct		25.02	24.68	26.56	26.91	20.47	25.09	19.50	24.41	25.42	25.79	22.37
<b>Arabidopsis seedling</b>	16	At3g24160	At1g13320	At3g27820	GADPH	ACTB	UBQ10					
GeNorm (Avg M)		0.19	0.19	0.22	0.25	0.28	0.32			0.20	0.20	0.28
Mean Ct		20.23	21.04	21.47	17.74	17.51	17.73			20.91	20.91	17.66
<b>Arabidopsis leaf</b>	16	At3g01150	GAPDH	At3g61710	ACTB	At1g32050	UBQ10					
GeNorm (Avg M)		0.16	0.16	0.31	0.42	0.50	0.63			0.21	0.32	0.40
Mean Ct		26.65	21.07	25.80	21.66	20.03	23.63			24.51	24.16	22.12
<b>Arabidopsis apex</b>	10	At2g17390	AT3G17920	At5g51880	ACTB	GADPH	UBQ10					
GeNorm (Avg M)		0.11	0.11	0.15	0.20	0.22	0.49			0.12	0.12	0.30
Mean Ct		18.89	23.14	22.22	17.91	17.92	21.86			21.42	21.42	19.23

# RefGenes – essential figures

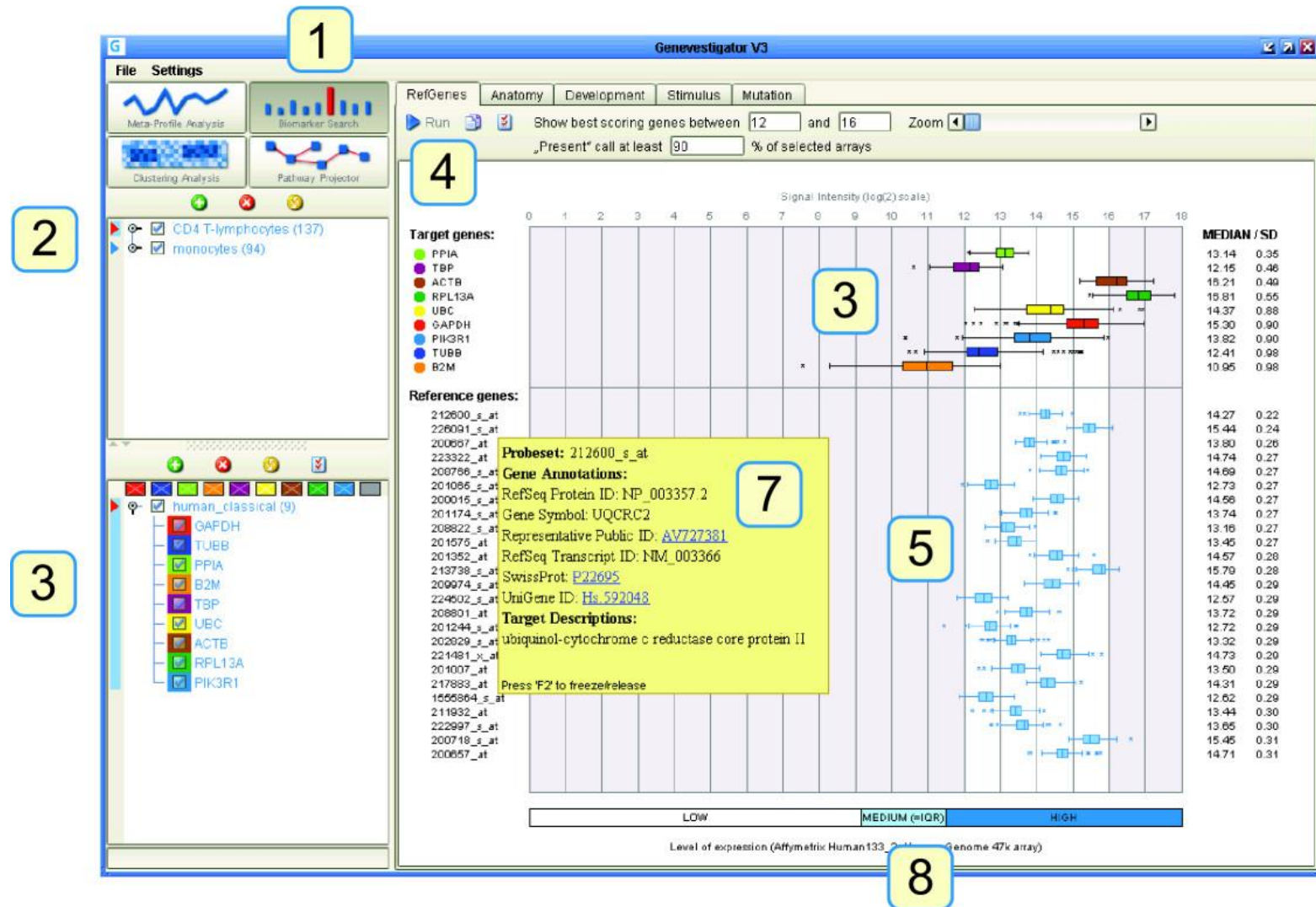


# RefGenes – essential figures





# RefGenes – essential figures

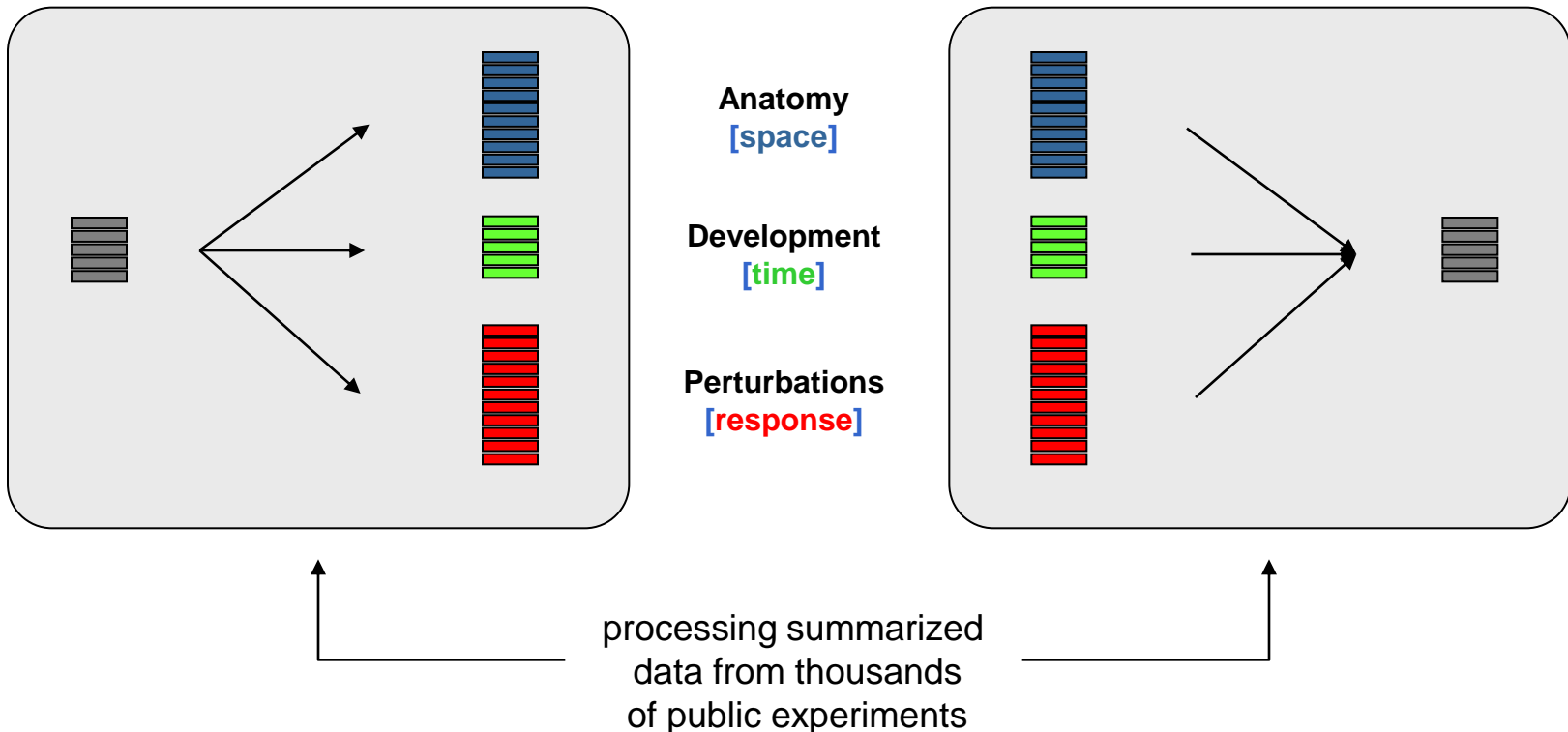




# Analytical approaches in Genevestigator

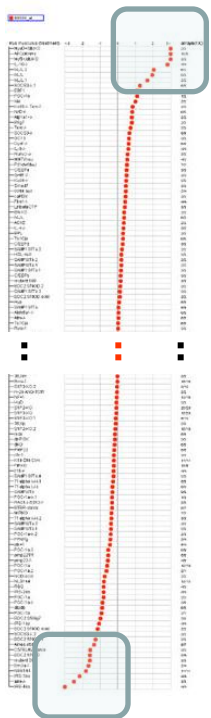
genes → which conditions?

conditions → which genes?



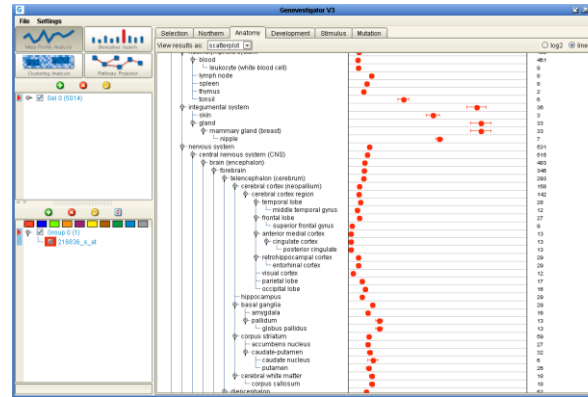
# Example output

Most upregulating conditions

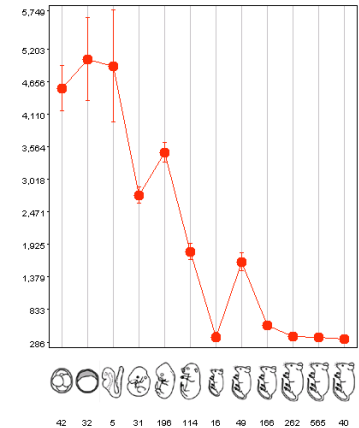


experimental conditions summarized from many experiments

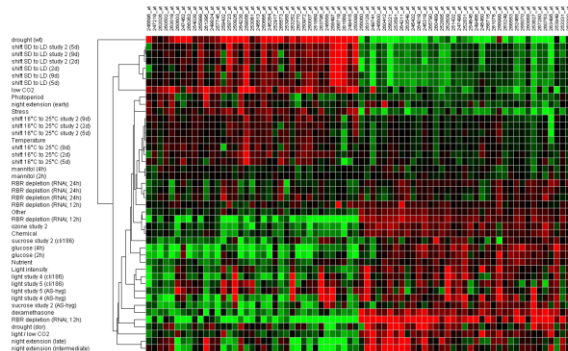
Most downregulating conditions



Anatomical expression profile of human ERBB2



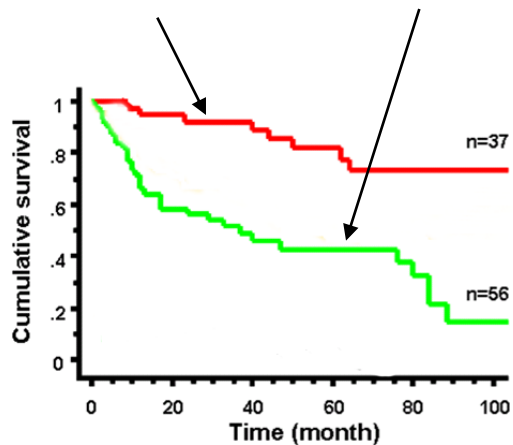
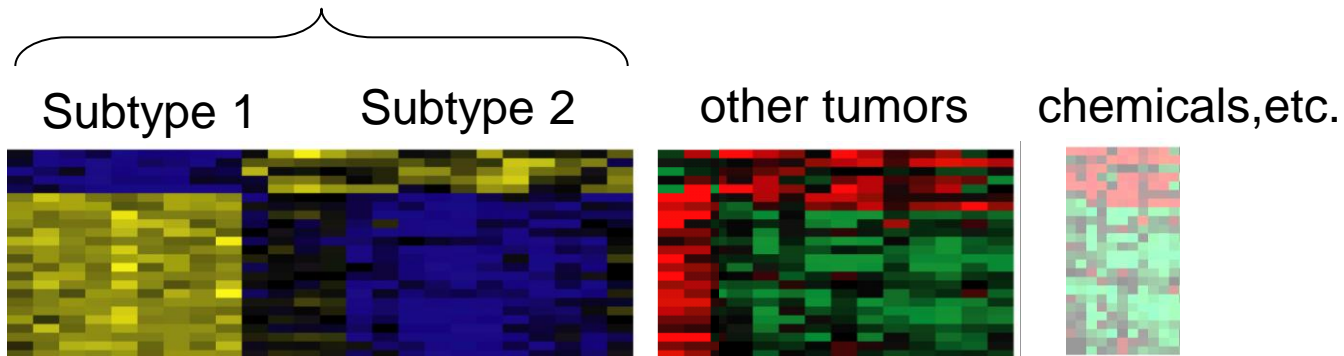
Development expression profile of mouse BRCA1



Response of growth-related genes in Arabidopsis

# Example 1: conserved gene signatures

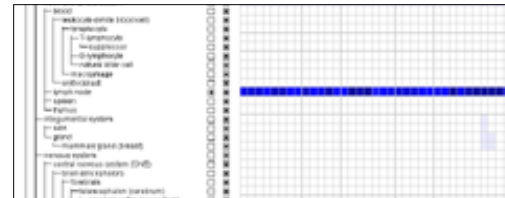
Tumors with identical  
clinico-morphological  
characteristics!



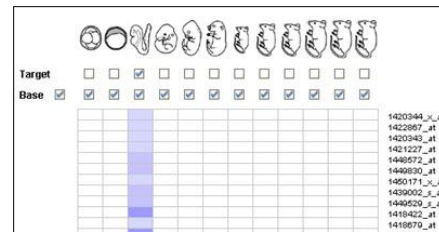
# Biomarker Search

- Identify genes that exhibit specific expression characteristics

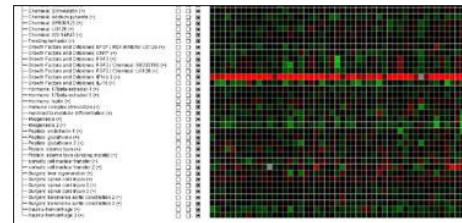
## Anatomy



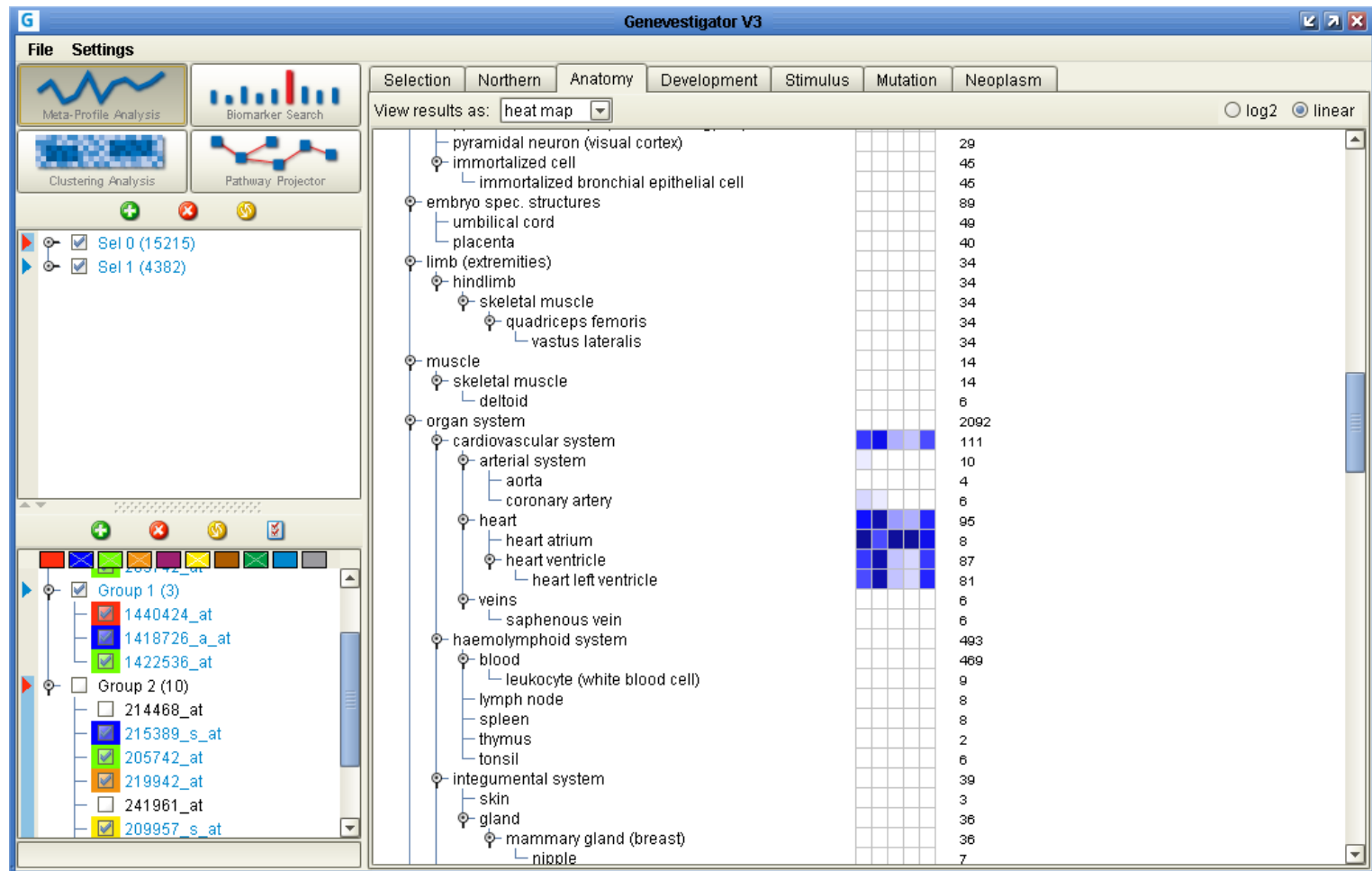
## Development



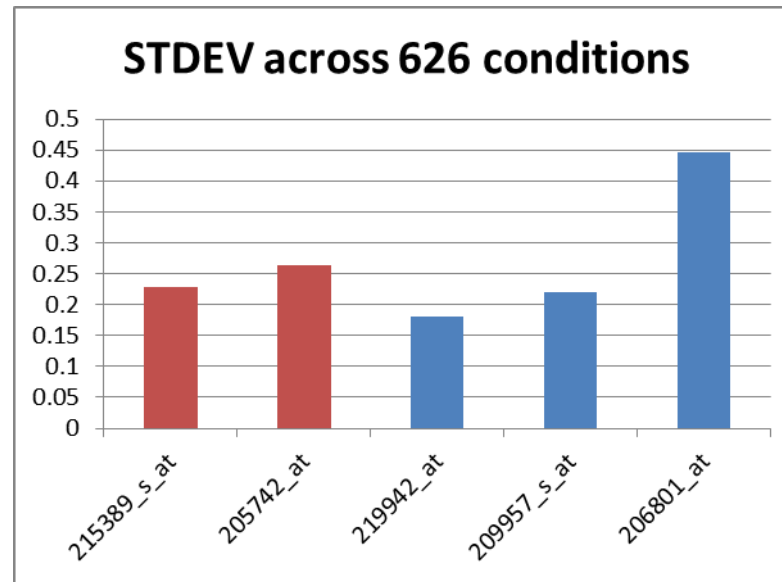
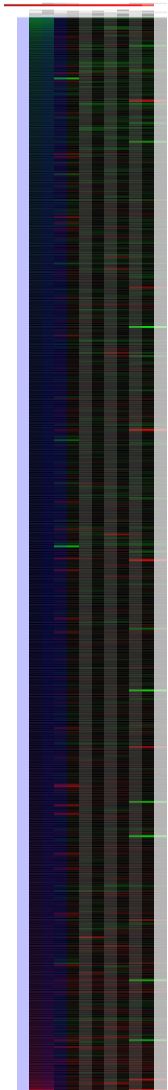
## Conditions/ Genotypes



## Example 2: heart specific genes



## Example 2: Troponins T and I



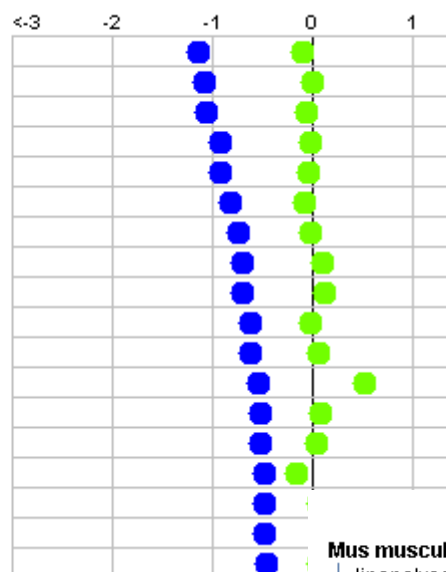
Troponin T

Troponin I

## Example 2: Troponin T and I

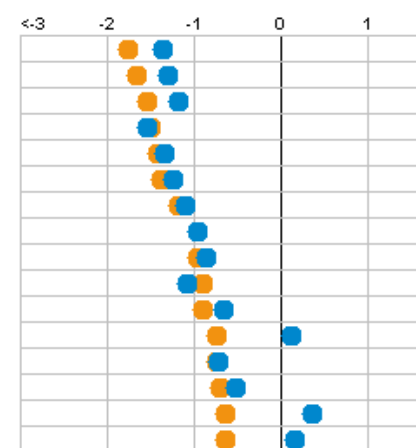
### Homo sapiens (experimental)

- melanoma
- kidney transplantation study 8 (early)
- gefitinib
- kidney transplantation study 8 (late)
- kidney transplantation study 8 (intermediate)
- renal cell carcinoma (metastasis)
- renal cell carcinoma (prim. tumor)
- baculovirus
- MAGP2
- differentiation medium: dex / AA (late)
- serum Amyloid A stimulation (intermediate)
- adenocarcinoma study 3 (negative)
- smoking study 33
- differentiation medium: dex / AA / BGP (late)
- kidney transplantation
- smoking study 32 (1 m)
- bFGF (nero)
- mucociliary differentiation (intermediate)



### Mus musculus (experimental)

- lipopolysaccharide
- resiquimod / protein (ovalbumin) study 2
- simvastatin / lipopolysaccharide
- sodium arsenite study 14
- sodium arsenite study 16
- 17beta-estradiol study 8
- sodium arsenite study 12
- mammary gland development study 5 (lactating)
- 17beta-estradiol study 9
- N. brasiliensis
- resiquimod / protein (ovalbumin)
- spinal cord injury study 3 (intermediate)
- mechanical ventilation study 4
- ovalbumin / particulate matter
- heat shock study 2 (late)
- FGF2

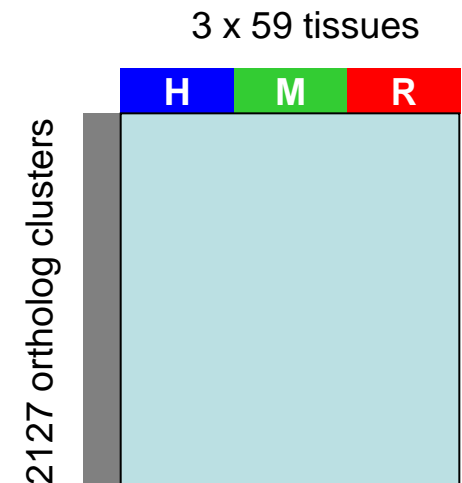


## Example 3: gene expression in the context of tissues

- ▶ Expression data matrix with orthologous genes
  - Human133A (20k), MouseU74Av2 (12k), RatU34 (8k)
  - Ortholog clusters from OMA
  - 1 combination per gene was selected (probe specificity + good hybridization)
  - Total: **2127 unique gene ortholog mappings** were created

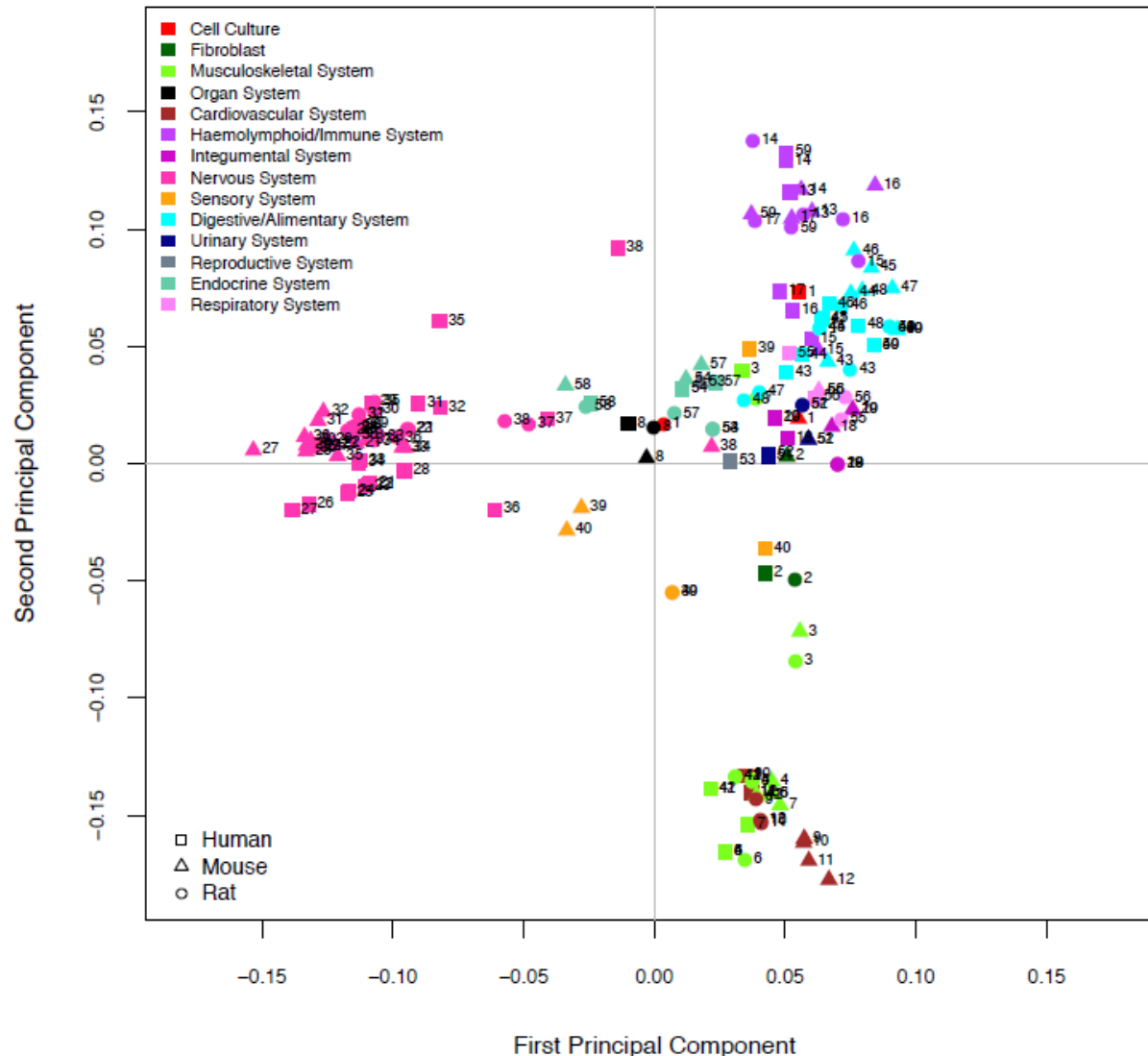
- ▶ **59 common tissue types** for human, mouse and rat

- ▶ Number of hybridizations: **9300 Affymetrix arrays**



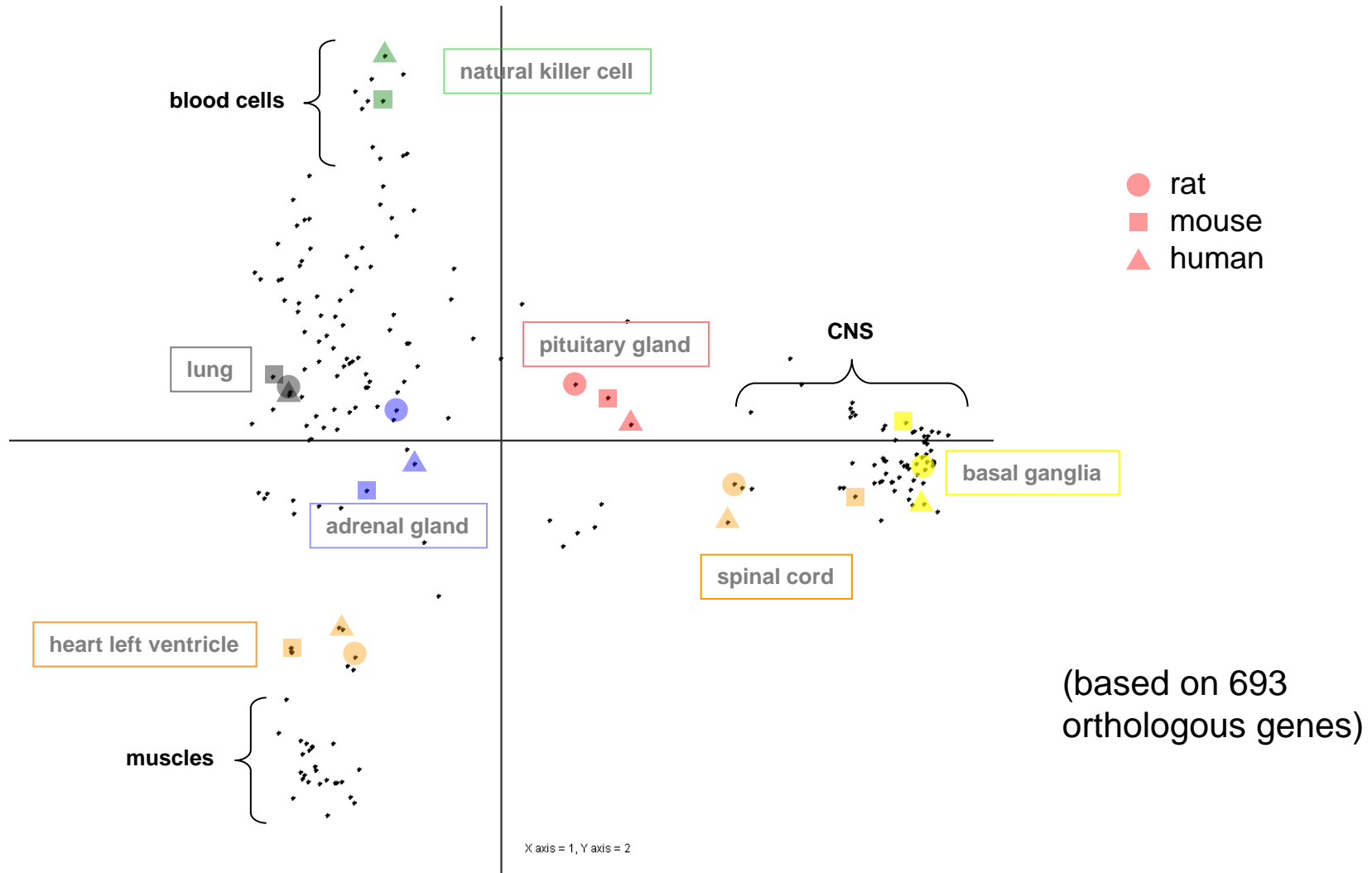


# Tissue expression across human, mouse and rat

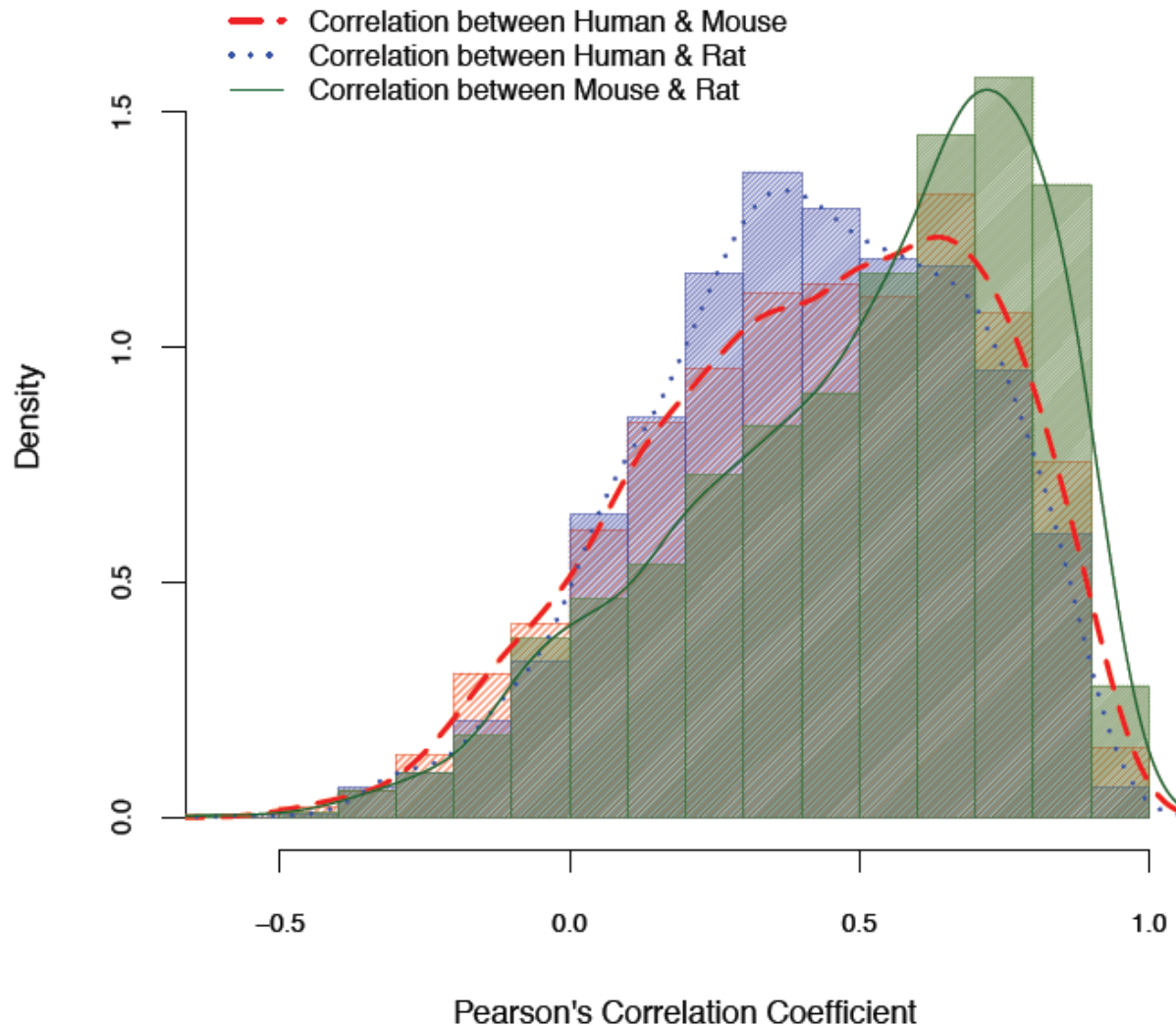


(based on 1897  
orthologous genes)

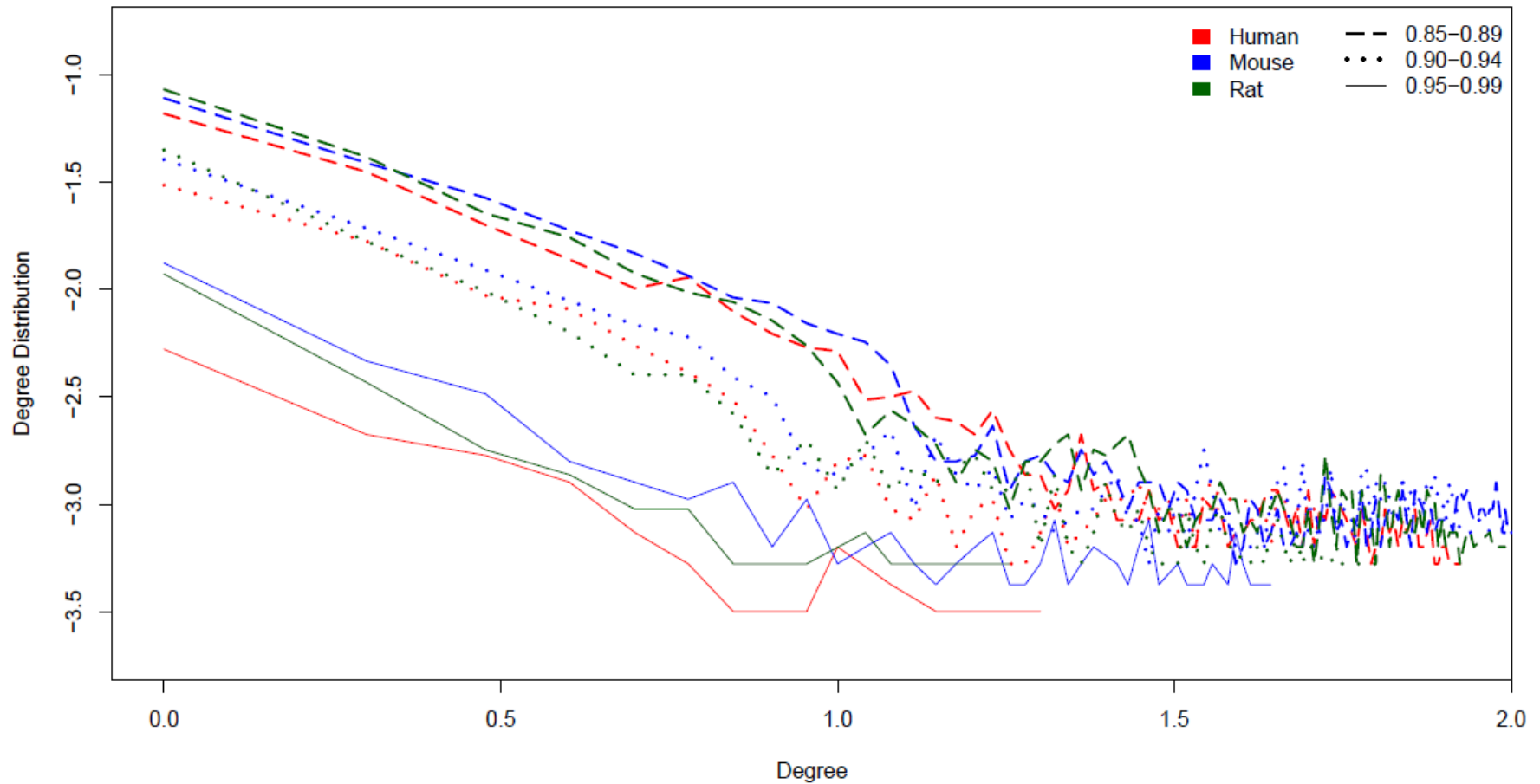
# Tissue expression across human, mouse and rat



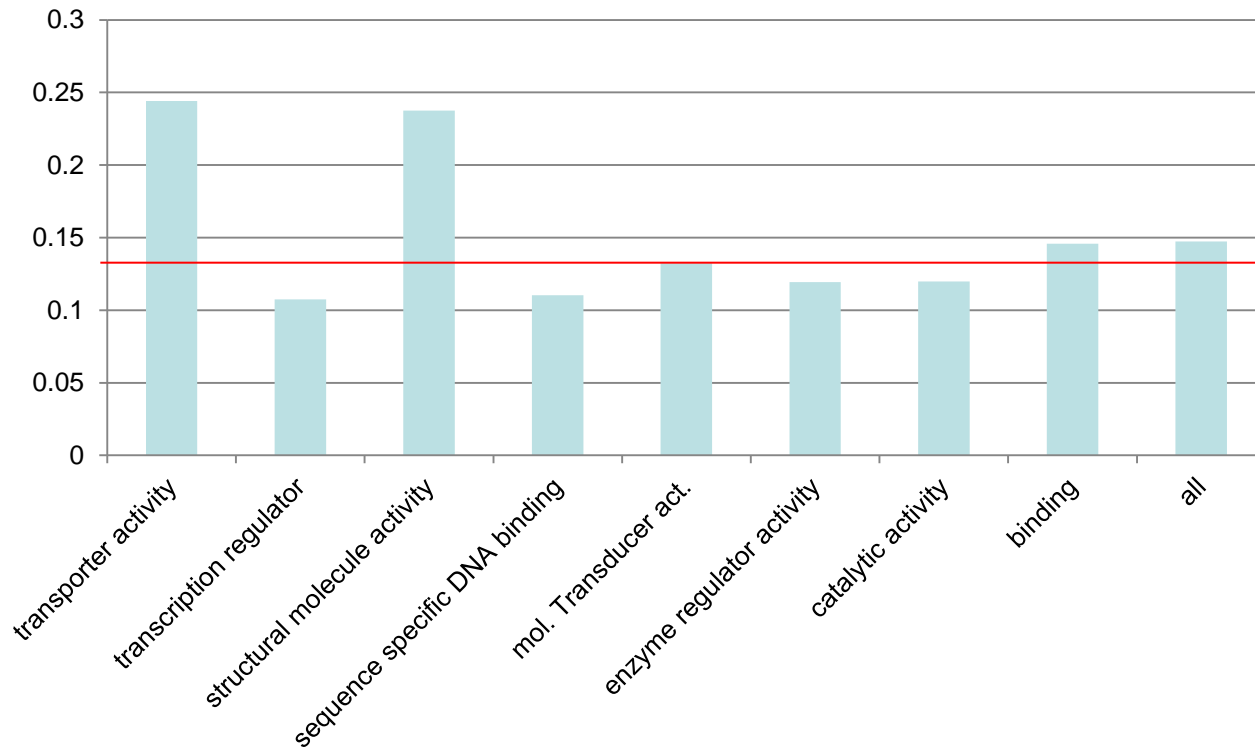
# Expression correlation between H, M, R



# Scale free structure of tissue correlation networks



# Most conserved biological processes



## Tissue profile is predominantly conserved between Human, Mouse and Rat

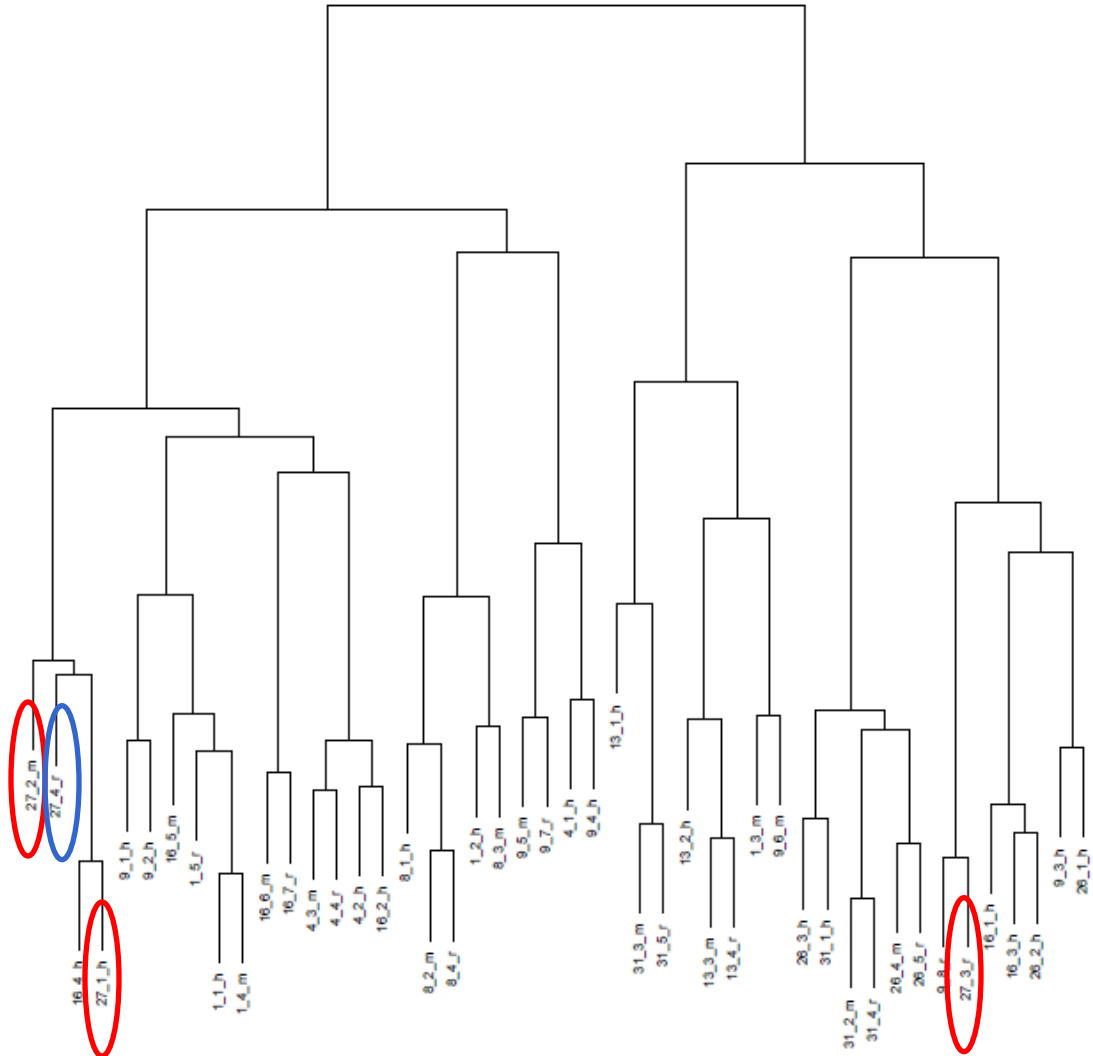
	Similar Anatomical Profile	Similar Level of Expression
25%	Yes	yes
65%	Yes	No
10%	No	Yes/No

## Protein sequence

- ▶ Human:
  - 27\_1\_h
- ▶ Mouse:
  - 27\_2\_m
- ▶ Rat:
  - 27\_3\_r
  - 27\_4\_r

Best sequence match

## Gene expression



## Take home message

- ▶ Contextual Meta-Analysis allows to
  - Better interpret your data
  - Get better confidence about your results
  - Identify highly specific biomarkers

# CONTEXT

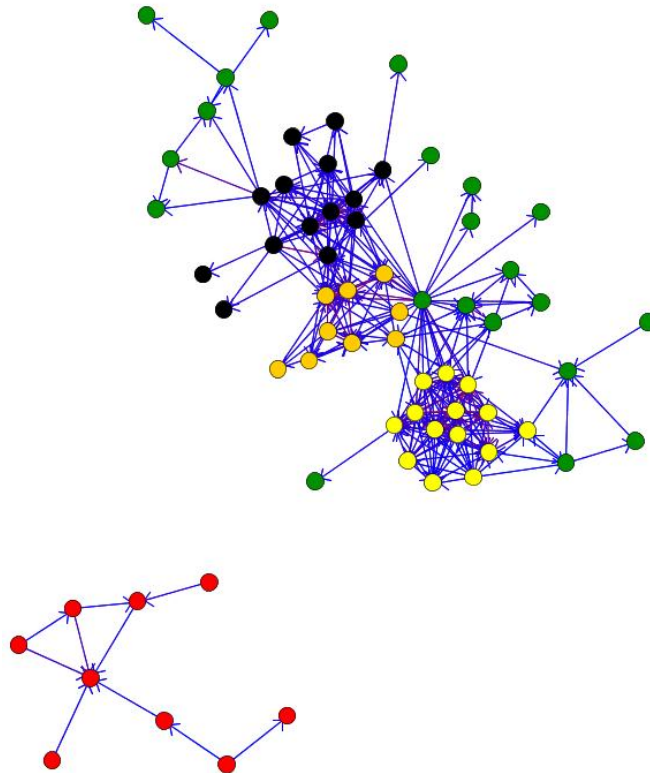


Experimental  
context?

Data types used?

Organism?

Cell type?



Network dynamics?

Reproducibility?

# Conclusions

- ▶ Contextual meta-analysis gives very valuable insight into the regulation of genes across a wide variety of conditions
- ▶ The integration of human, mouse and rat databases increases the number of biological contexts
- ▶ At the level of tissue expression, there is a high conservation between human, mouse and rat.

[www.genevestigator.com](http://www.genevestigator.com)

OPEN

CLASSIC

ADVANCED

# Acknowledgements

- ▶ ETH Zurich
  - Wilhelm Gruissem, Philip Zimmermann, Ajay Prasad, Suchitra Suresh
  - Peter Widmayer, Tomas Hruz
- ▶ Genevestigator Team / NEBION AG
  - Stefan Bleuler
  - Oliver Laule
  - Markus Wyss
  - Peter von Rohr
  - Kirsten Laule
  - Bernd Stahl
  - Karsten Keldermann
  - Rasa Meskauskiene
  - Igor Asner
- ▶ KTI/CTI, EU-Agron-Omics Project