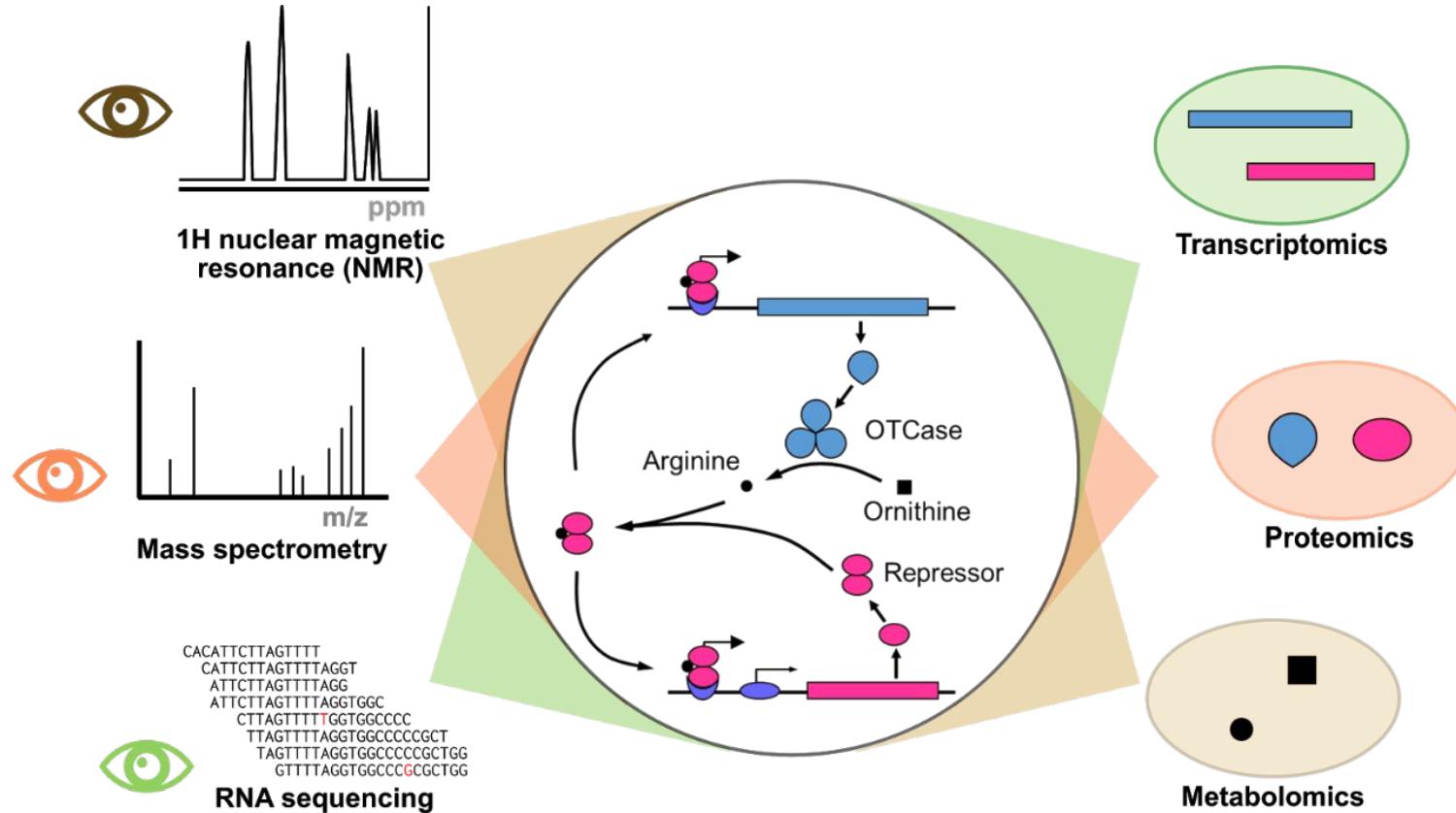


AMIDD Lecture 6: From network to cellular modelling



Omics data are projections of high-dimensional biological space. It is an *inverse problem* to infer a high-dimensional space from its projections.

Multiscale Modelling of Drug Mechanism and Safety by Zhang, Sach-Peltason, Kramer, Wang and Ebeling, Drug Discovery Today, 2020

Dr. Jitao David Zhang, Computational Biologist

¹ Pharmaceutical Sciences, Pharma Research and Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche

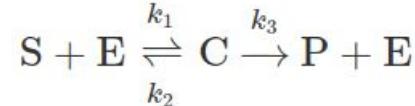
² Department of Mathematics and Informatics, University of Basel

Topics

- **Gene expression profiling: a case study of omics and cellular modelling**
- **Applications for drug safety: TG-GATEs**
- **Applications for drug mechanism of action: molecular phenotyping**

Simulation of biological networks with ordinary differential expression: the simplest case

Given the reaction



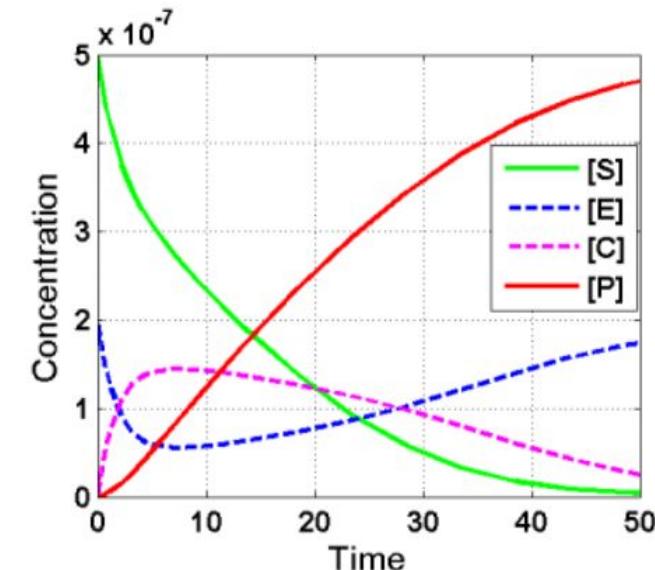
According to the law of mass action

$$\begin{aligned}\frac{d[S]}{dt} &= -k_1[E][S] + k_2[C], \\ \frac{d[E]}{dt} &= -k_1[E][S] + (k_2 + k_3)[C], \\ \frac{d[C]}{dt} &= k_1[E][S] - (k_2 + k_3)[C], \\ \frac{d[P]}{dt} &= k_3[C],\end{aligned}$$

Given the initial values and rate constants

- $S(0) = 5e^{-7}$
- $E(0) = 2e^{-7}$
- $C(0) = P(0) = 0$
- $k_1 = 1e^6$
- $k_2 = 1e^{-4}$
- $k_3 = 0.1$

It is possible to simulate the concentration changes by time *deterministically*.



See [Systems Engineering Wiki \(tue.nl\)](#) for MATLAB/COPASI codes and *Stochastic Modelling for Systems Biology* by Darren J. Wilkinson

Chemical Master Equations (CME): a particle model of chemical reaction

Given the reaction $A + B \xrightleftharpoons[k_2]{k_1} C + D$ and the initial condition $X(0) = \begin{bmatrix} K \\ K \\ 0 \\ 0 \end{bmatrix}$ (K molecules of species A and of species B respectively)

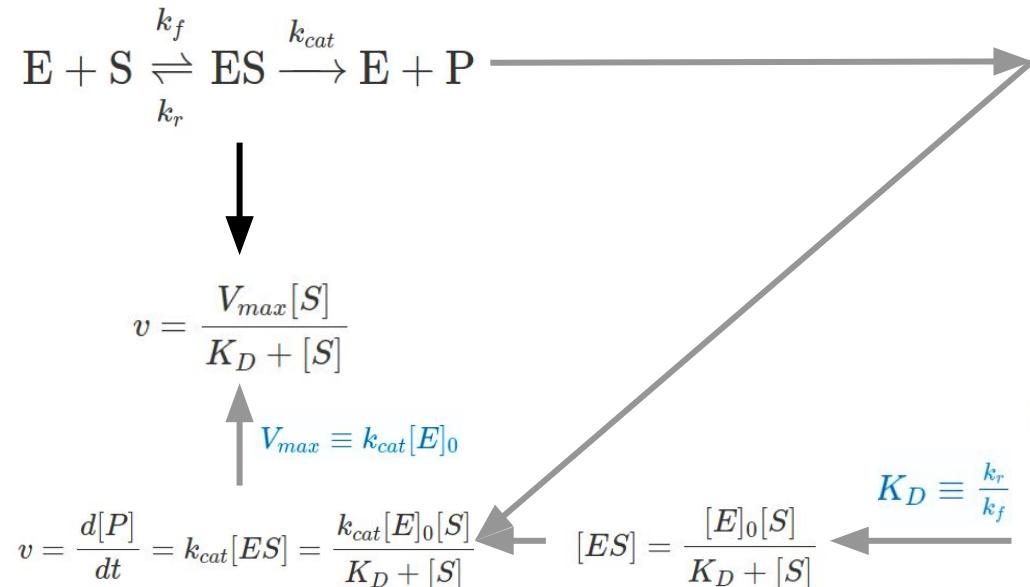
The state vector $X(t)$ can take at any time point *one* of the values

$$\begin{bmatrix} K \\ K \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K-1 \\ K-1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} K-2 \\ K-2 \\ 2 \\ 2 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ K \\ K \end{bmatrix},$$

Theoretically we can build an ODE system with $K+1$ equations to model *every state of the reaction*, down to every particle. In reality, the dimension is so high so that a simulation is not feasible.

CME is a set of ODEs, with each ODE representing one possible state of the system. Solution of the k th equation at time t is a real number giving the probability of system being in that particular state at that time.

Reaction Rate Equations (RRE): a compartment model



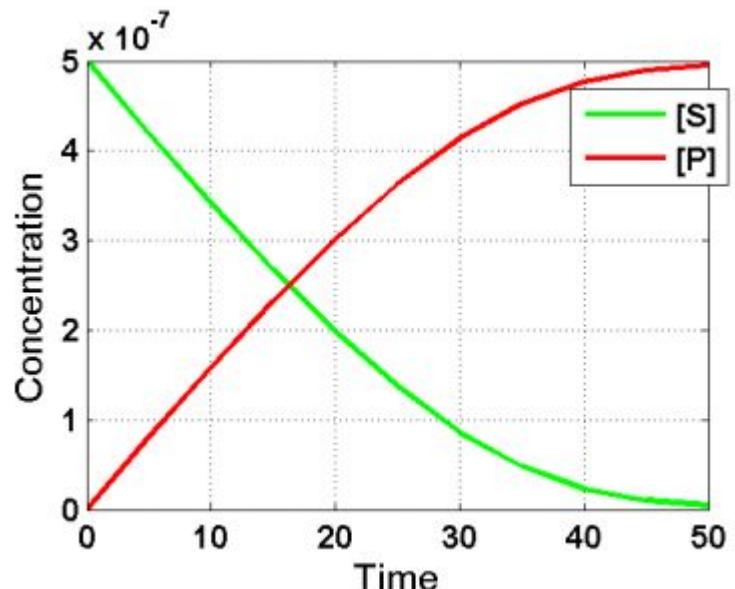
$$\begin{aligned} \frac{d[E]}{dt} &= -k_f[E][S] + k_r[ES] + k_{cat}[ES], \\ \frac{d[S]}{dt} &= -k_f[E][S] + k_r[ES], \\ \frac{d[ES]}{dt} &= k_f[E][S] - k_r[ES] - k_{cat}[ES], \\ \frac{d[P]}{dt} &= k_{cat}[ES], \end{aligned}$$

\downarrow

$$k_f[E][S] = k_r[ES]$$

$$\begin{aligned} k_f([E]_0 - [ES])[S] &= k_r[ES] \\ k_f[E]_0[S] - k_f[ES][S] &= k_r[ES] \\ k_f[E]_0[S] &= k_r[ES] + k_f[ES][S] \\ k_f[E]_0[S] &= [ES](k_r + k_f[S]) \\ [ES] &= \frac{k_f[E]_0[S]}{k_r + k_f[S]} \\ [ES] &= \frac{k_f[E]_0[S]}{k_f(\frac{k_r}{k_f} + [S])} \end{aligned}$$

RRE simulation of the Michaelis-Menten model



Source: [Systems Engineering Wiki \(tue.nl\)](#)

RRE is a set of ODEs, with each ODE representing one chemical species. Solution of the j th equation at time t is a real number representing the concentration of species j at time t .

The Gillespie's algorithm and the chemical Langevin equation allow stochastic simulation of biological networks

- The *stochastic simulation algorithm* (exact SSA), also called *Gillespie's algorithm*, allows stochastic simulation of a reaction.
- It is performed in four steps
 - **Initialize** the system with initial conditions
 - Given a state at time t , we can define a probability p that reaction j takes place in the time interval $[t+\tau, t+\tau+d\tau]$. It is the product of two density functions of two random variables: the probability of reaction j happens (proportional to the number of substrate molecules), multiplied by the time until next reaction, which is exponentially distributed. This is known as the **Monte Carlo** step.
 - Let the randomly selected reaction happen and **update** the time.
 - **Iterate** until substrates are exhausted or simulation time is over.
- Further computation tricks such as ‘tau-leaping’ by lumping together reactions are possible. The chemical Langevin equation (CLE) replaces further accelerates stochastic simulation by approximating the Poisson distribution with the normal distribution.

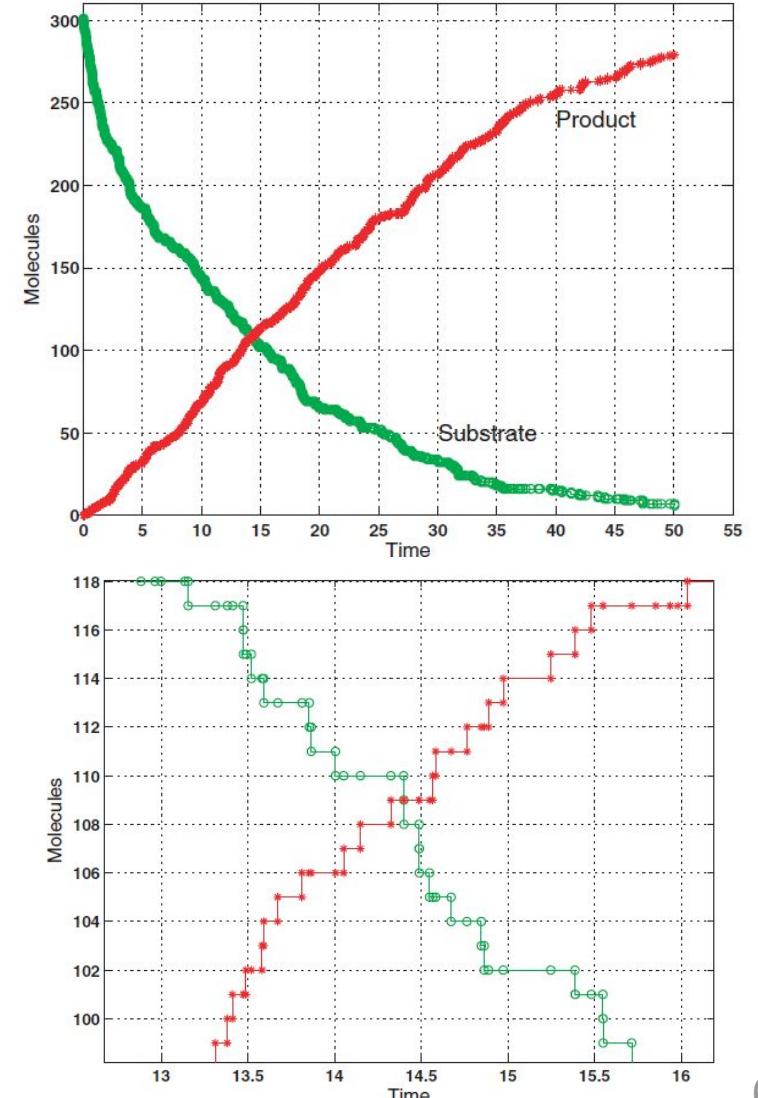
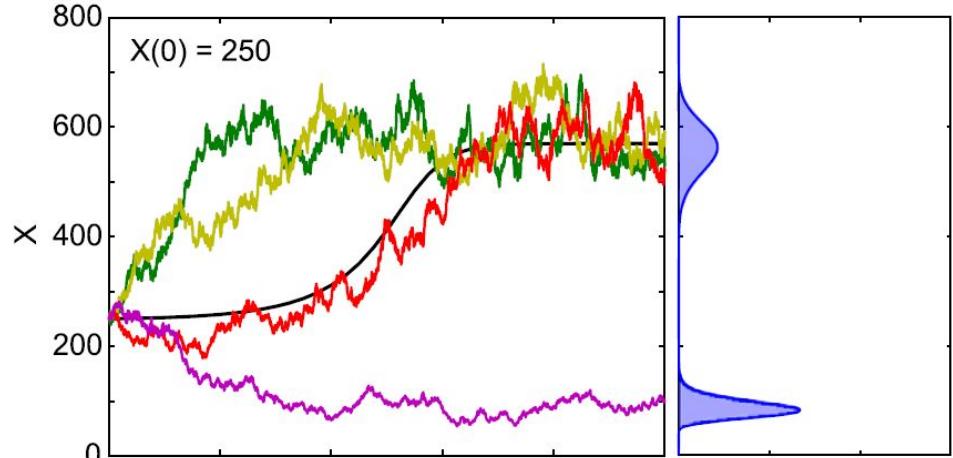
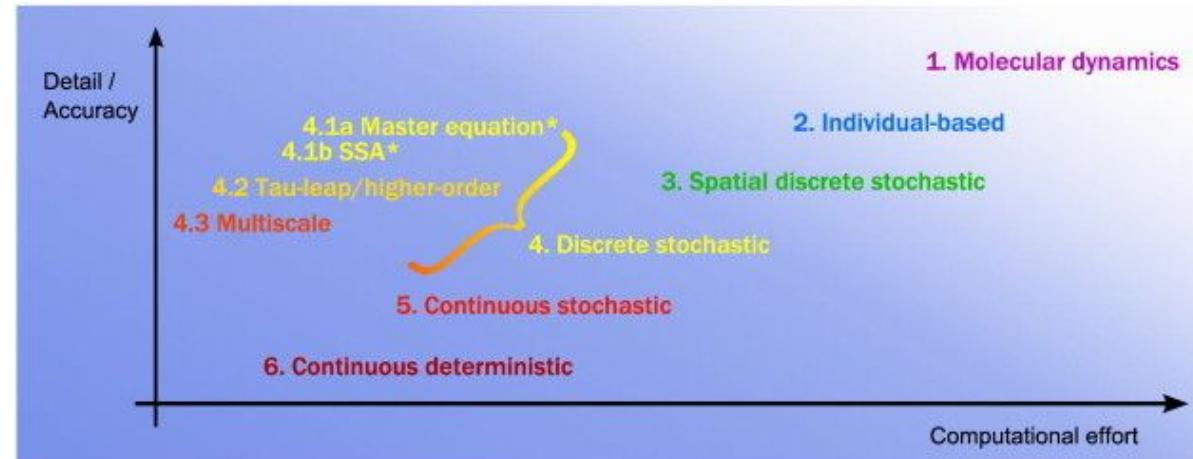


Figure source and further reading: Higham, Desmond J. 2008. “Modeling and Simulating Chemical Reactions.” *SIAM Review* 50 (2): 347–68. <https://doi.org/10.1137/060666457>.

Why stochastic modelling?



- Stochastic modelling can reveal individual trajectories that are otherwise ‘averaged’ by ODE models.
- Small systems and single-molecule studies show stochastic behaviour.
- It is possible to consider both extrinsic and intrinsic factors and take them into the model.



Advantages and disadvantages of several modelling/simulation methods.

Simulation method	Cat.	Advantages	Disadvantages	References	Software
Master equation	4	Exact	Very computationally intensive	[85,143]	
SSA	4	Statistically exact	Very computationally intensive	[82,109]	COPASI [144] StochKit [145] STOCKS [146] BioNetS [147]
Tau-leap	4	Relatively fast	Approximate; too slow for large systems or frequent/multiscale reactions	[83,113,118]	StochKit [145]
Higher-order	4	Relatively fast; accurate	Approximate; too slow for large systems or frequent/multiscale reactions	[83,121,122,124,125]	
Multiscale/hybrid	4	Fast; good for systems with disparate reaction scales	Approximate; problems with coupling different scales	[131,132,137,139,148]	COPASI [144] BioNetS [147]
Brownian dynamics	2	Tracks individual molecules	Slow; molecule size must be artificially added	[149,150]	Smoldyn [149,151] MCell [152]
Compartment-based	3	Accounts for diffusion between homogeneous compartments	Slow; compartment size must be set manually; each compartment is homogeneous	[150,153,154]	MesoRD [153] URDME [155]
SDE	5	Fast	Continuous; Gaussian noise	[76]	BioNetS [147]
PDE (R-D)	6	Very fast; spatial	Continuous; no noise	[156]	
ODE	6	Very fast	Continuous; no noise	[157]	

Székely and Burrage. 2014. “[Stochastic Simulation in Systems Biology](#).”

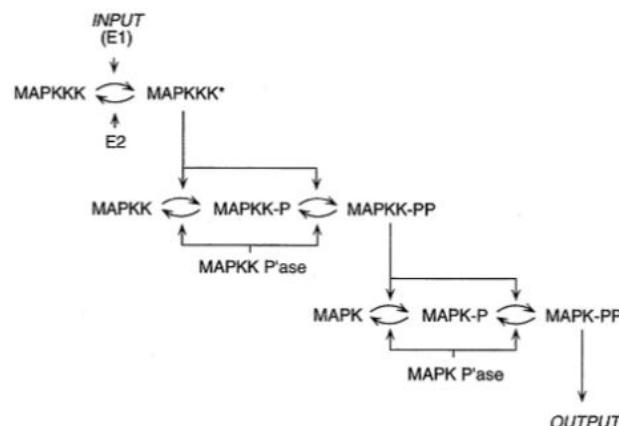
Computational and Structural Biotechnology Journal 12 (20–21): 14–25.

Also see *Stochastic Modelling for Systems Biology* by Darren J. Wilkinson.

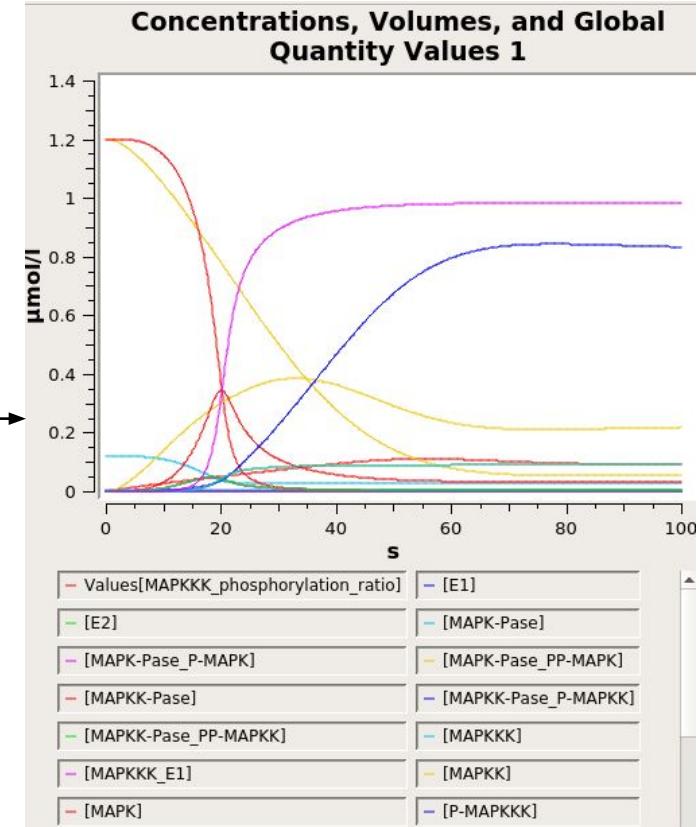
Cat. represents Category from Fig. 2. Abbreviations: SSA, stochastic simulation algorithm; SDE, stochastic differential equation; PDE (R-D), partial differential equation (classical reaction-diffusion equations); ODE, ordinary differential equation.

Biochemical system simulator COPASI

- Freely available at <http://COPASI.org/>
- COPASI supports two types of simulation
 - Ordinary differential equation (ODE) based simulation**
 - Stochastic kinetic simulation**, among others using the [stochastic Runge–Kutta method](#) (RI5) and [Gillespie's algorithm](#)
 - Resources to learn more about stochastic modelling: [MIT OpenCourseWare](#) by Jeff Gore, and [Stochastic Processes: An Introduction, Third Edition](#) by Jones and Smith
- Tutorials also available on [the website of European Bioinformatics Institute \(EBI\)](#)
- The mathematical concept and software tools are important for detailed analysis of enzymatic reactions, especially in the presence of drugs and/or disease-relevant mutation

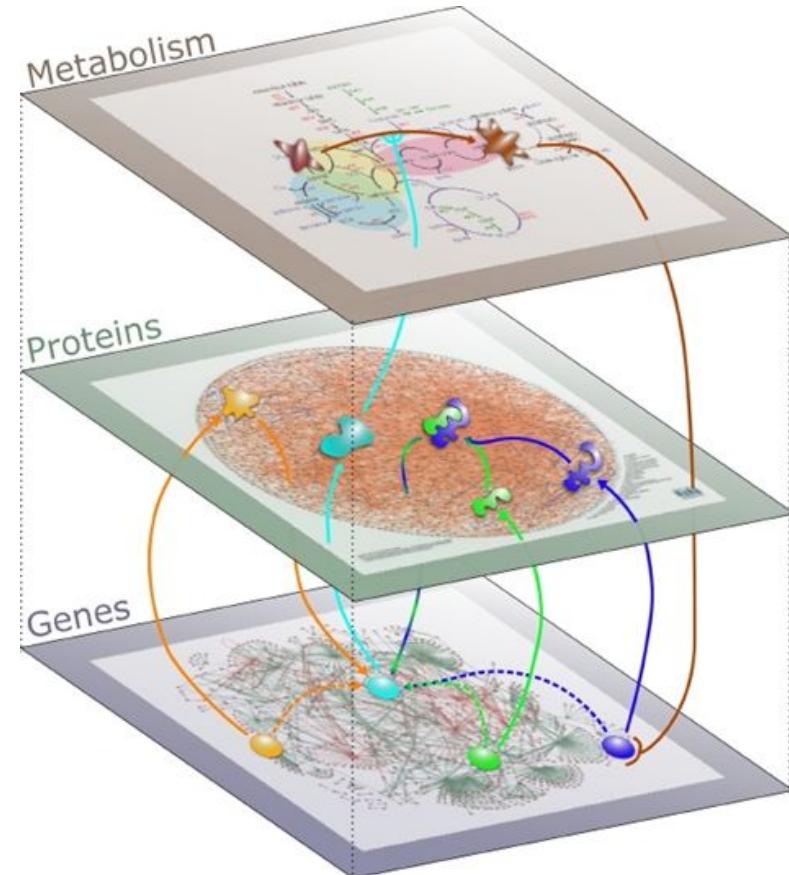


Huang and Ferrell, PNAS, 2006

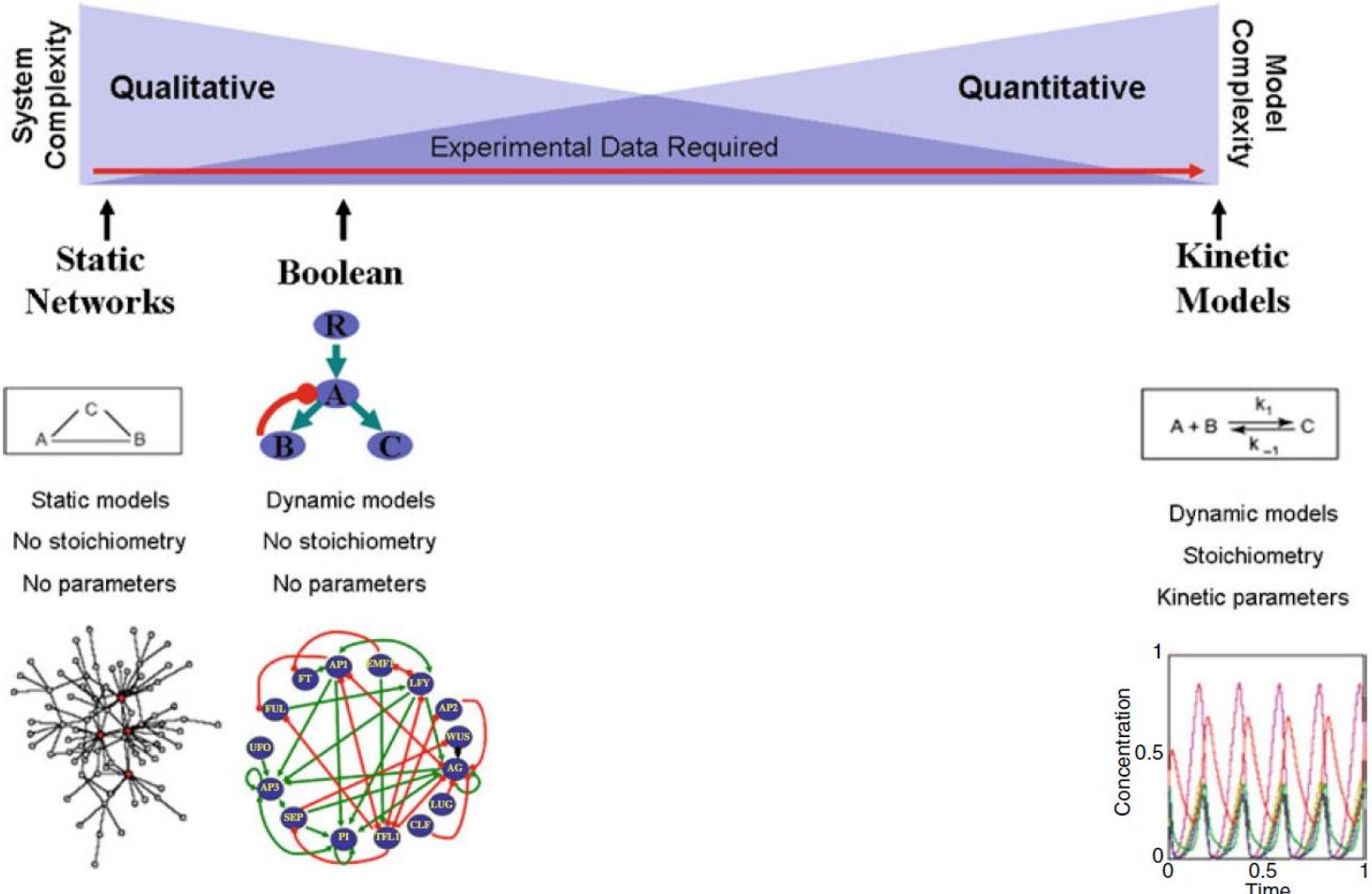


ODE-based simulation of dynamics

Modelling biological networks

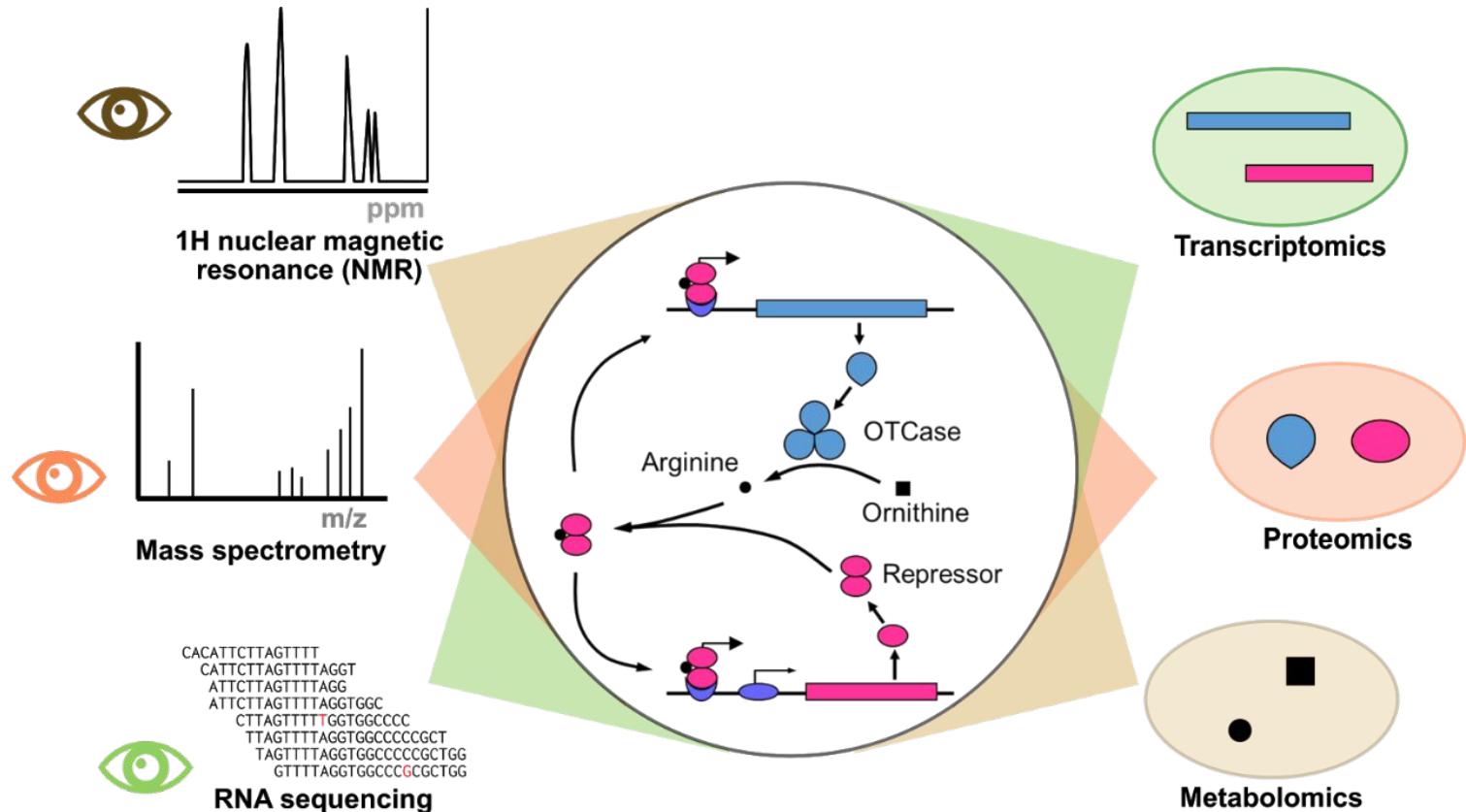


Stéphane CHÉDIN & Jean LABARRE, www-dsv.cea.fr

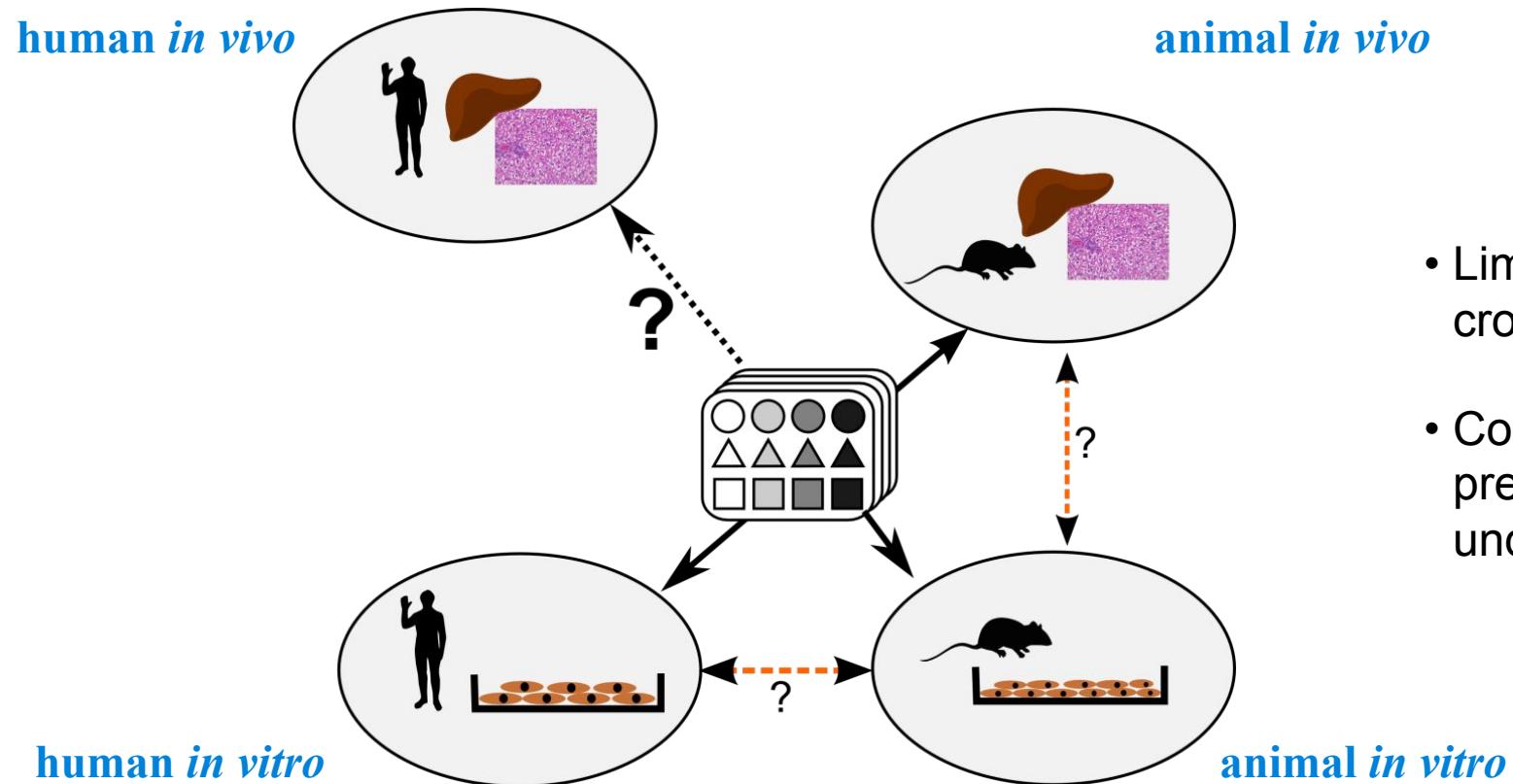


Garg, Abhishek, Kartik Mohanram, Giovanni De Micheli, and Ioannis Xenarios. 2012. "[Implicit Methods for Qualitative Modeling of Gene Regulatory Networks](#)." In *Gene Regulatory Networks: Methods and Protocols*, edited by Bart Deplancke and Nele Gheldof, 397–443. Methods in Molecular Biology. Totowa, NJ: Humana Press.

Omics data are projections of high-dimensional biological space



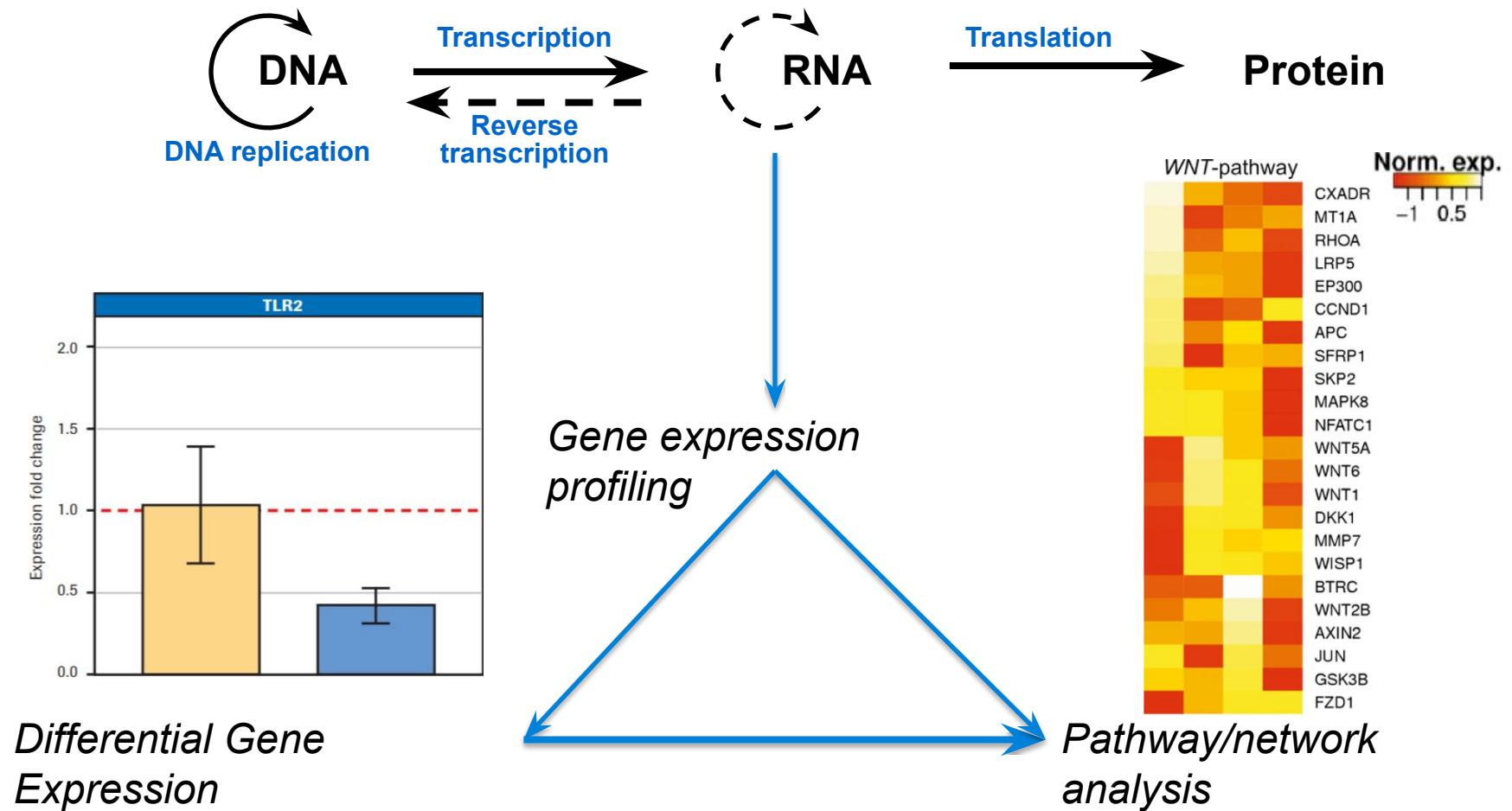
One challenge in drug discovery: non-clinical safety assessment



- Limited *in vitro-in vivo* and cross-species translatability
- Conflict between black-box prediction methods and the need to understand the mode of action

We need better (and interpretable) tools to predict safety profiles of drug candidates

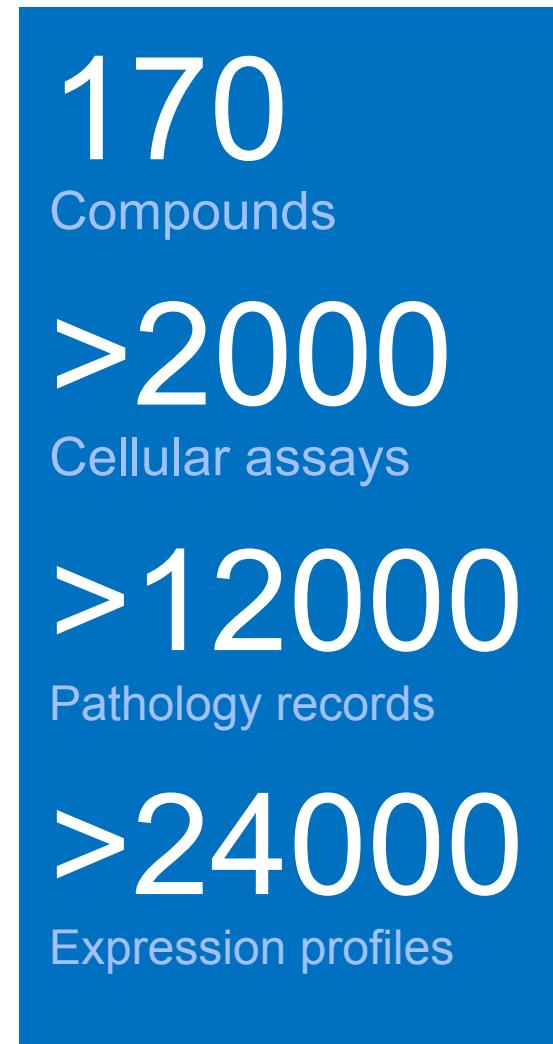
Principles of gene expression profiling



Figures: Wikimedia Commons/Thomas Shafee, CC/Adapted

TG-GATEs: Toxicogenomics Project-Genomics Assisted Toxicity Evaluation system

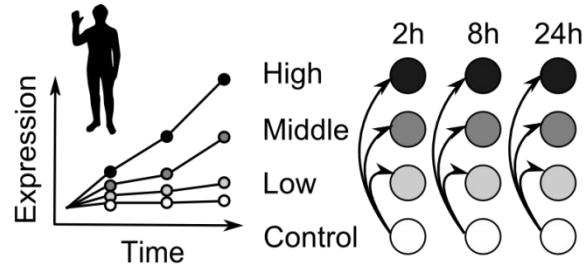
- **Japanese Consortium 2002-2011**
 - National Institute of Biomedical Innovation, National Institute of Health Sciences, and 15 pharmaceutical companies, including Roche Chugai.
- **Data fully released in 2012 to the public:** Time-series and dose-dependent experiments using 170 bioactive compounds
 - *In vitro & in vivo* gene expression profiling, each containing gene expression data of about 20,000 genes
 - *In vitro* PicoGreen DNA quantification assay
 - *In vivo* histopathology in liver and kidney
 - *In vivo* clinical chemistry
- **Total raw data size >2 TB**



TG-GATEs is a valuable data source to study drug-induced toxicity *in vitro* and *in vivo*

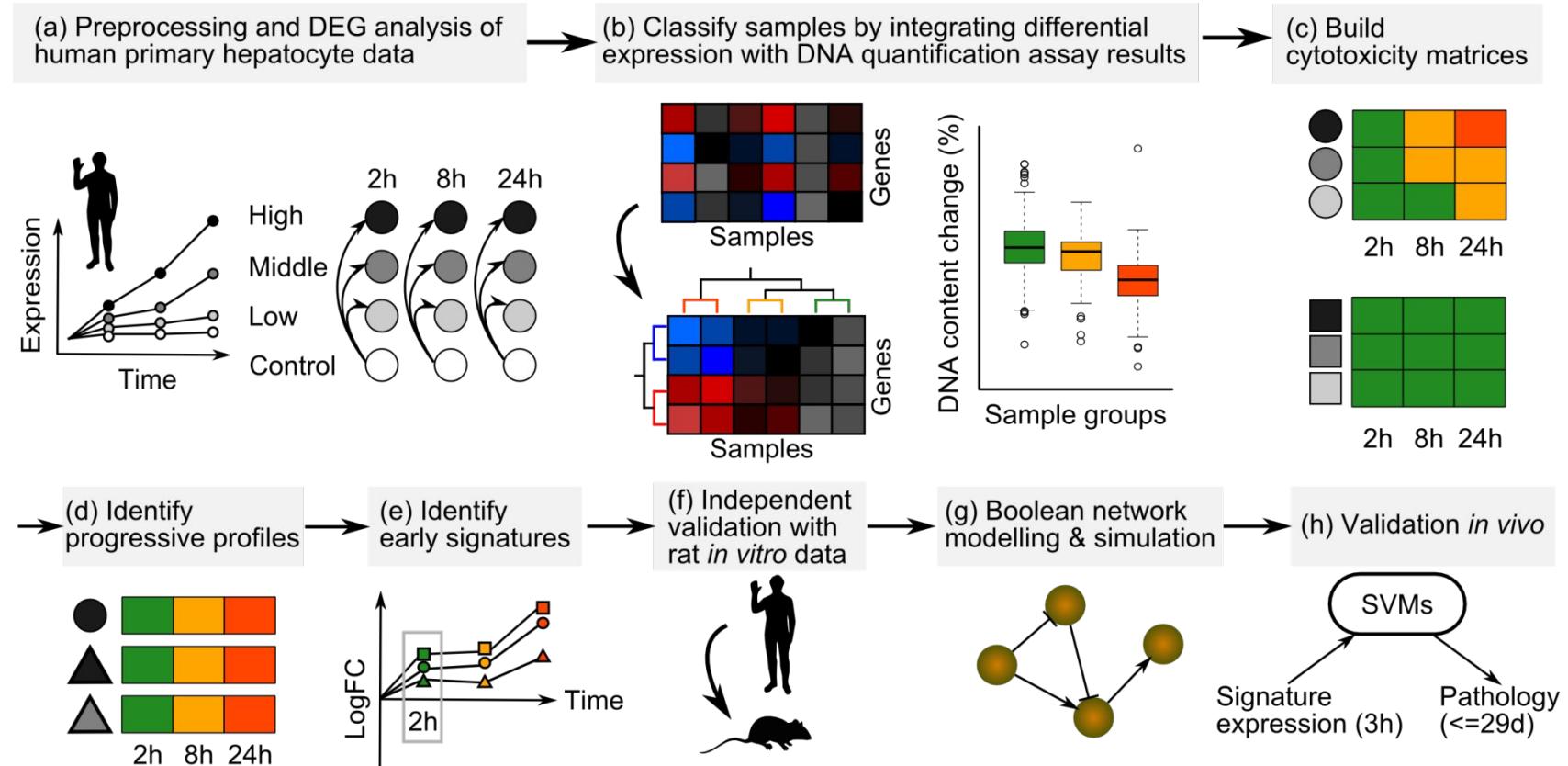
We built a computational pipeline to identify early signatures of toxicity

(a) Preprocessing and DEG analysis of human primary hepatocyte data



We integrate unsupervised learning, regression analysis, and network modelling to reach the goal

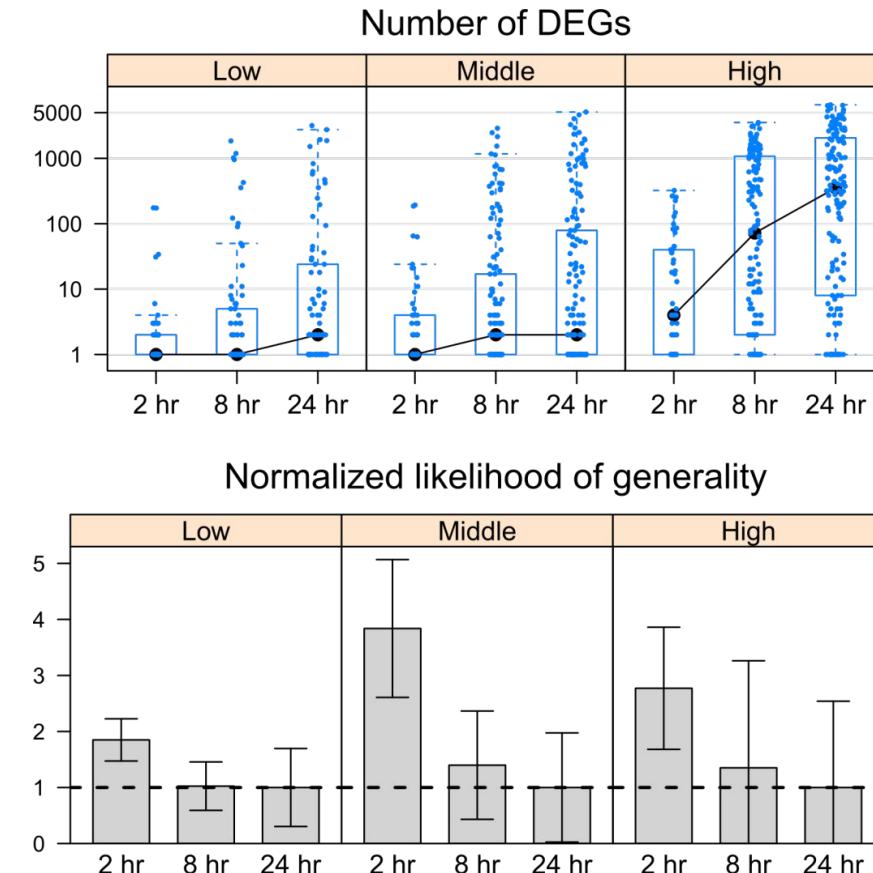
We built a computational pipeline to identify early signatures of toxicity (without animation)



We integrate unsupervised learning, regression analysis, and network modelling to reach the goal

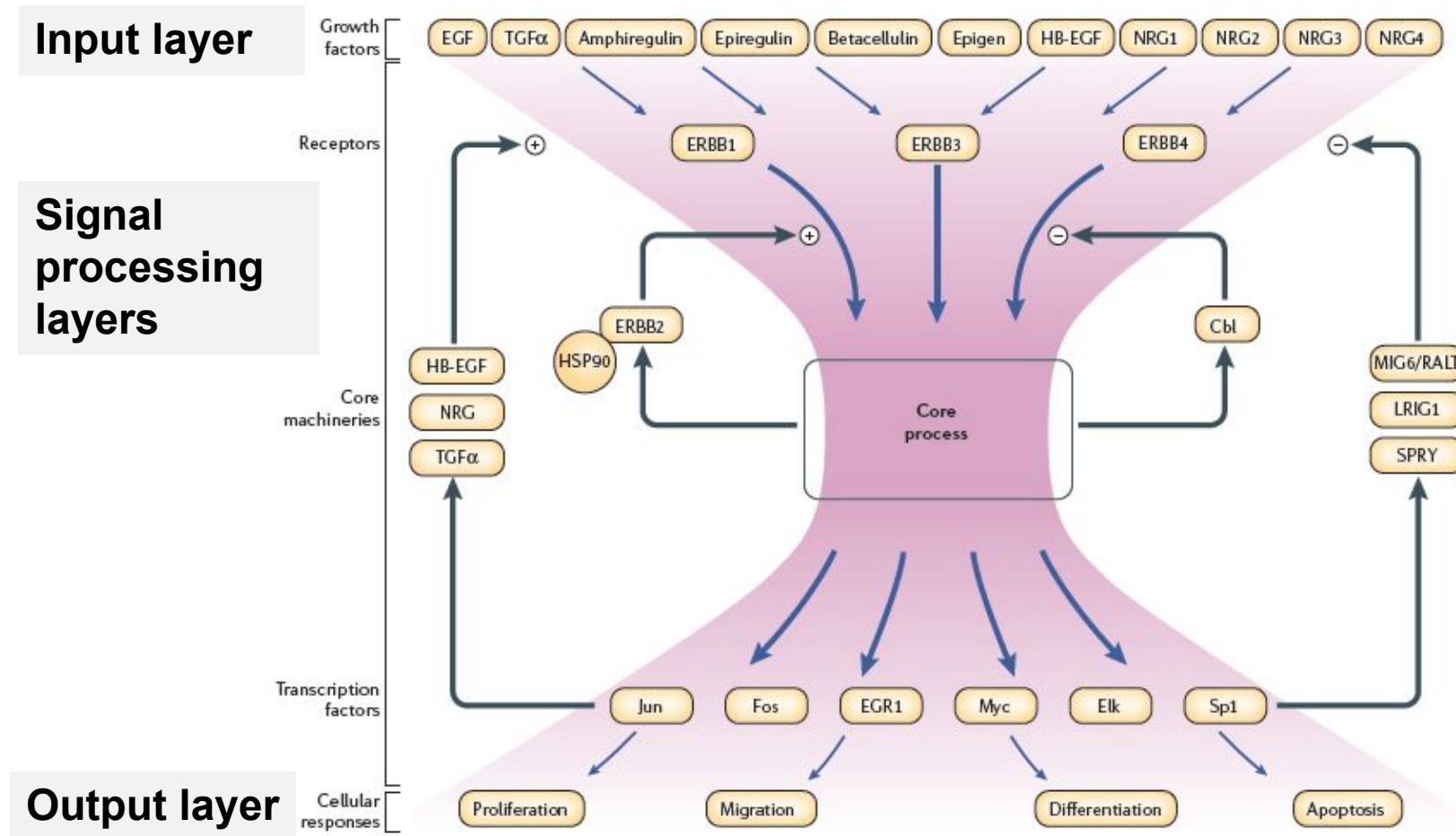
It is worth observing early

- We found that early-response genes induced 2h after compound administration are more generic (less specific) than late-induced genes: they are more likely to be induced by multiple compounds.
- □ We hypothesize that diverse signalling pathways «back-converge» to a few early-response genes, which can be toxicity signatures.



Contrary to common wisdom (at the time), we argue that toxicogenomics should focus on early time points

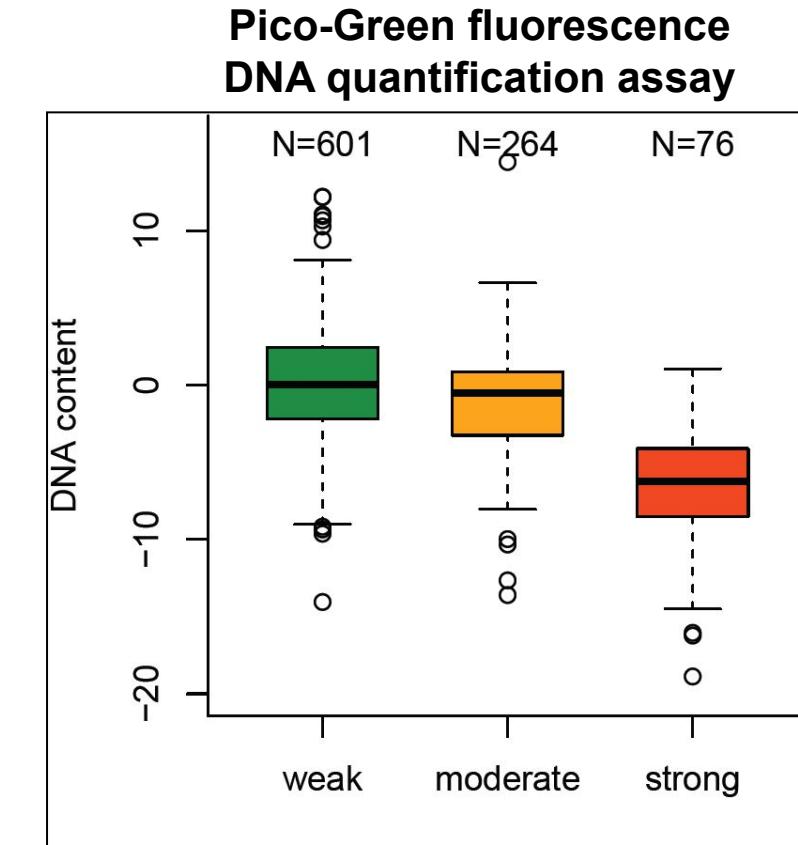
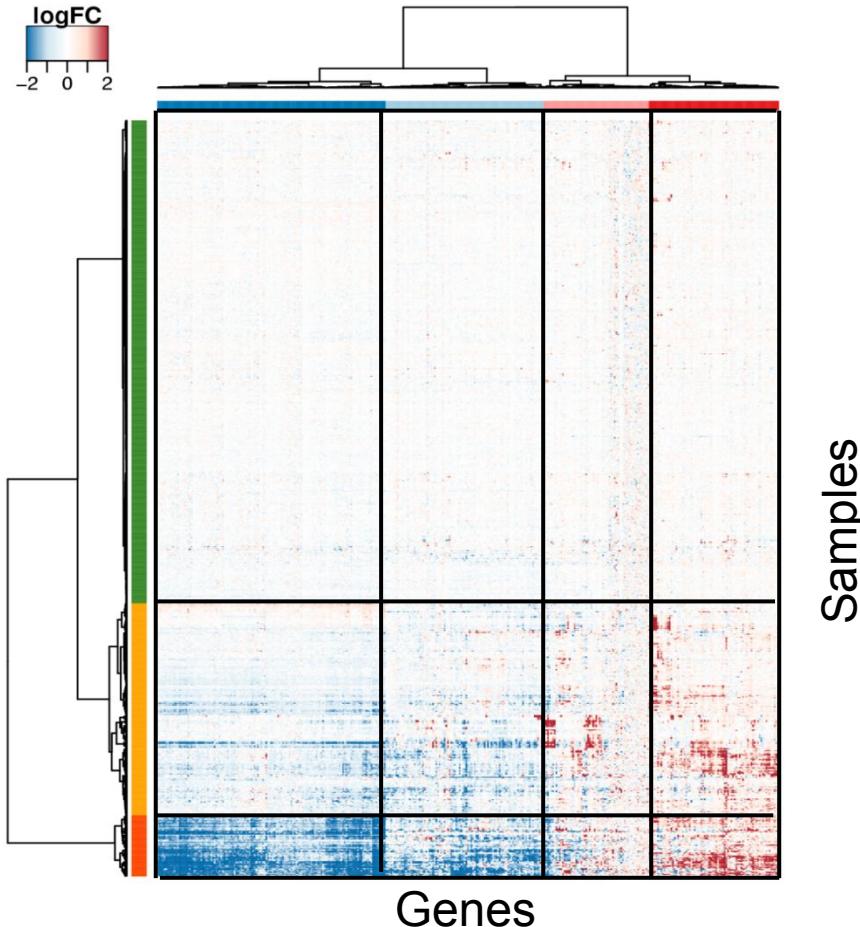
The bow-tie structure of signalling networks as a model that explains the power of early time point



Adapted from Ami Citri and Yosef Yarden,
Nature Reviews Molecular Cell Biology (2005)

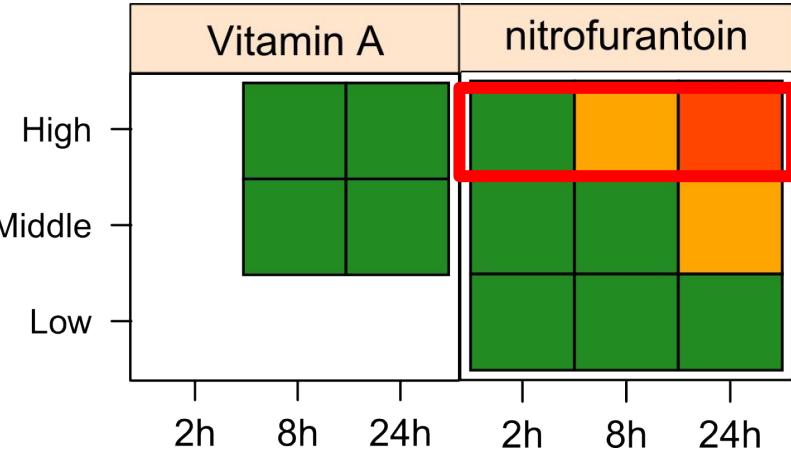
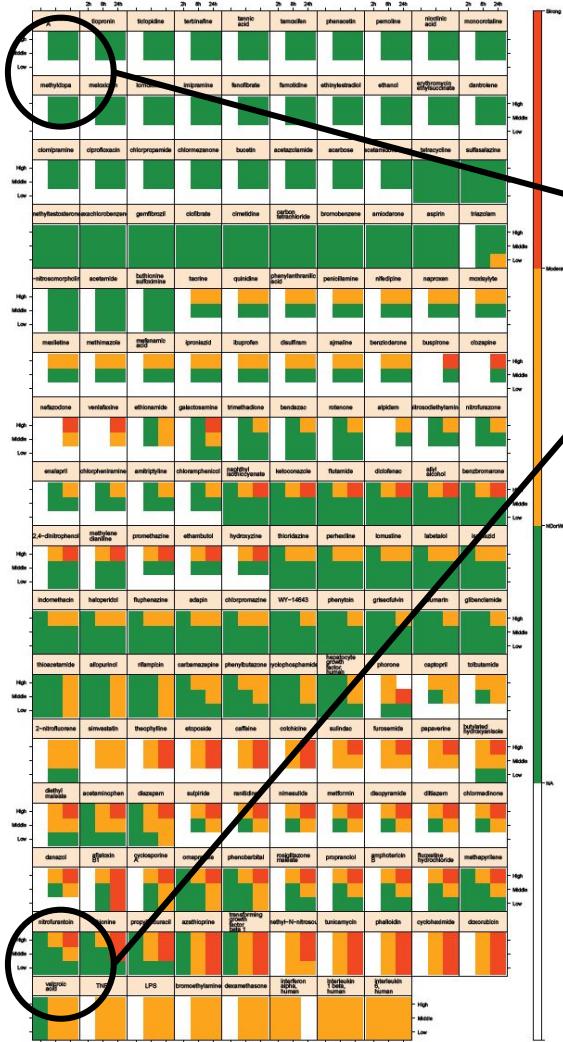
We hypothesize that signalling pathways that mediate toxicity «back-converge» to a few early-response genes

Compound-induced cytotoxicity can be classified into three levels by molecular phenotypes



Unsupervised clustering identified groups of compounds associated with cytotoxicity

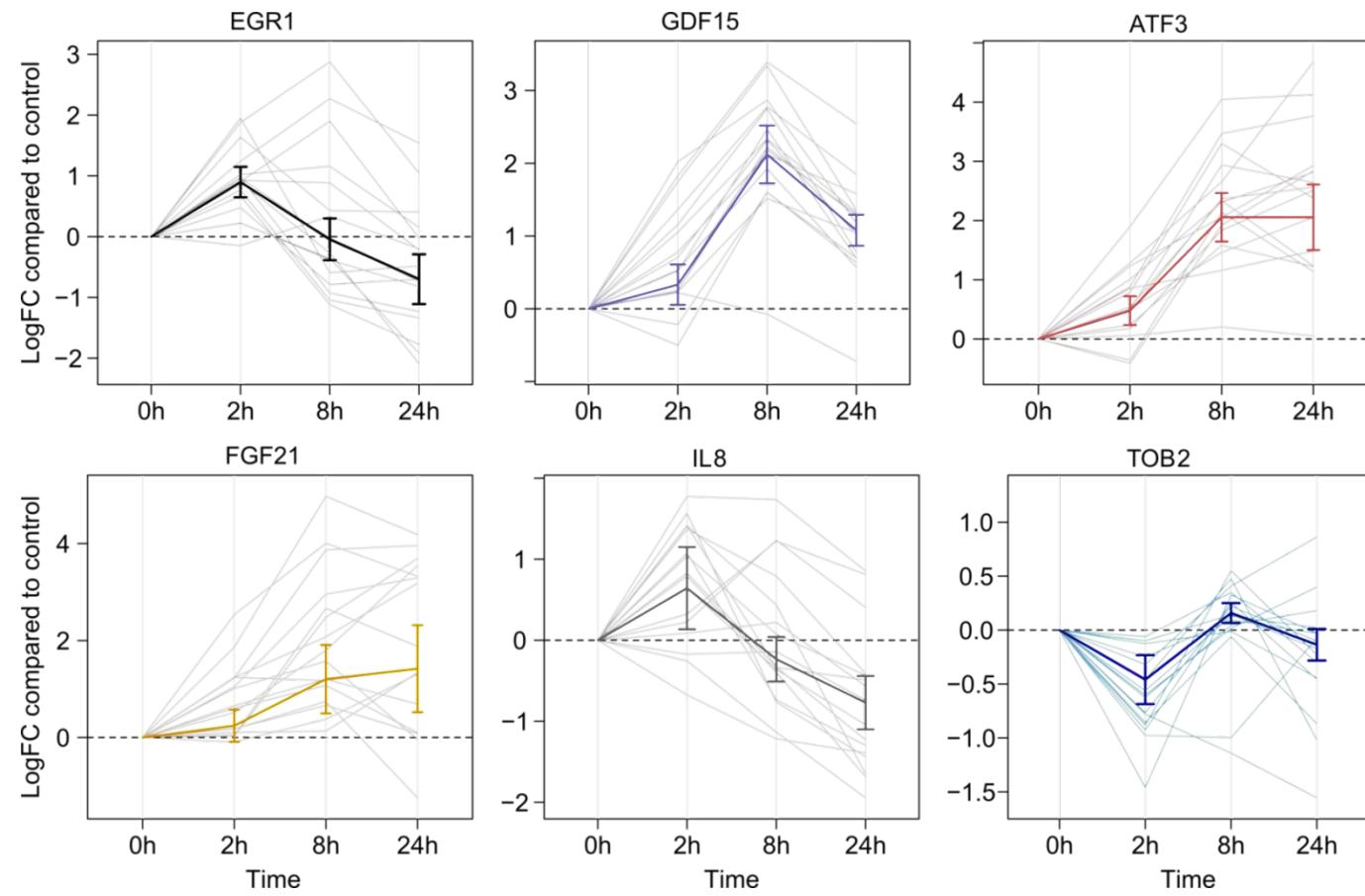
Cytotoxicity matrix and early signatures identified from progressive profiles *in vitro*



For **predictive signatures**, we focused on *progressive cytotoxicity profiles* (such as the one in the red box) and identified genes commonly regulated at 2h in such profiles.

Unsupervised clustering allowed us to identify progressive cytotoxicity profiles

Expression patterns of early signatures in human



Genes which were consistently and significantly up- or down-regulated at 2h in progressive profiles were chosen as signatures ($|logFC|>0.25$ & $p<0.05$). Purely data-driven: no biological knowledge was used for prioritization.

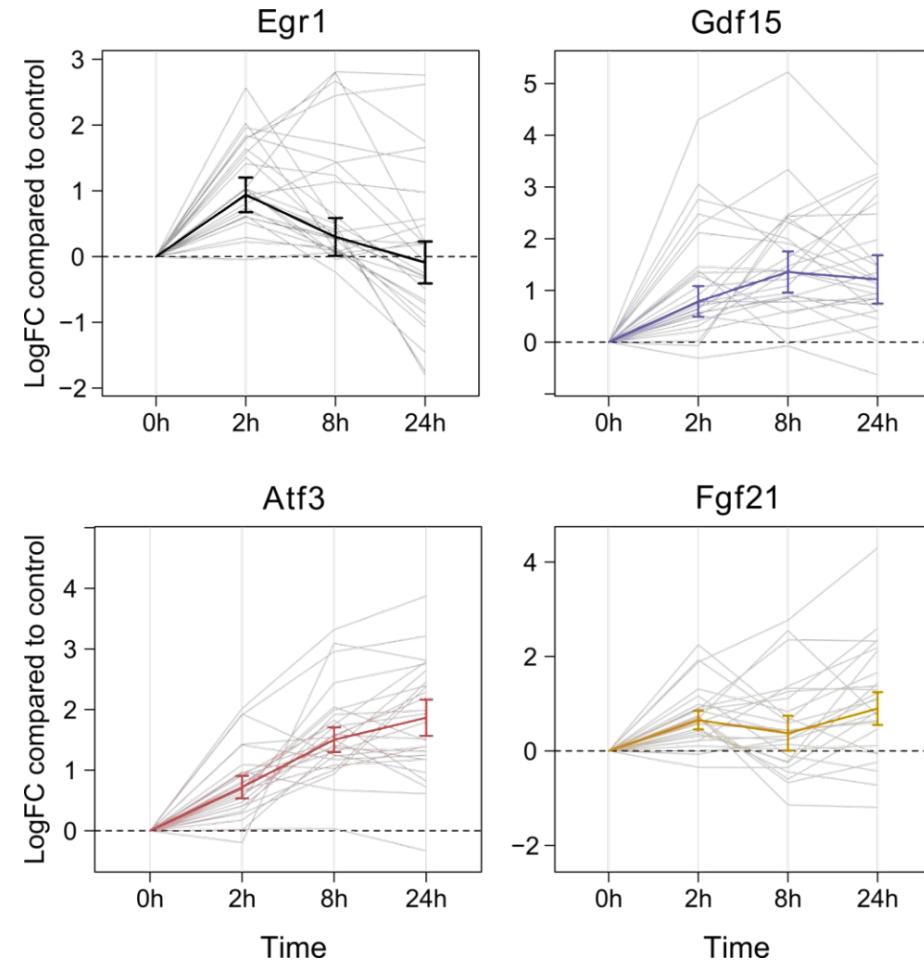
Each thin line represents one treatment, and the thick line represents the average.

Six early gene signatures of cytotoxicity were identified from human *in vitro* data

A consensus signature set of cytotoxicity emerges

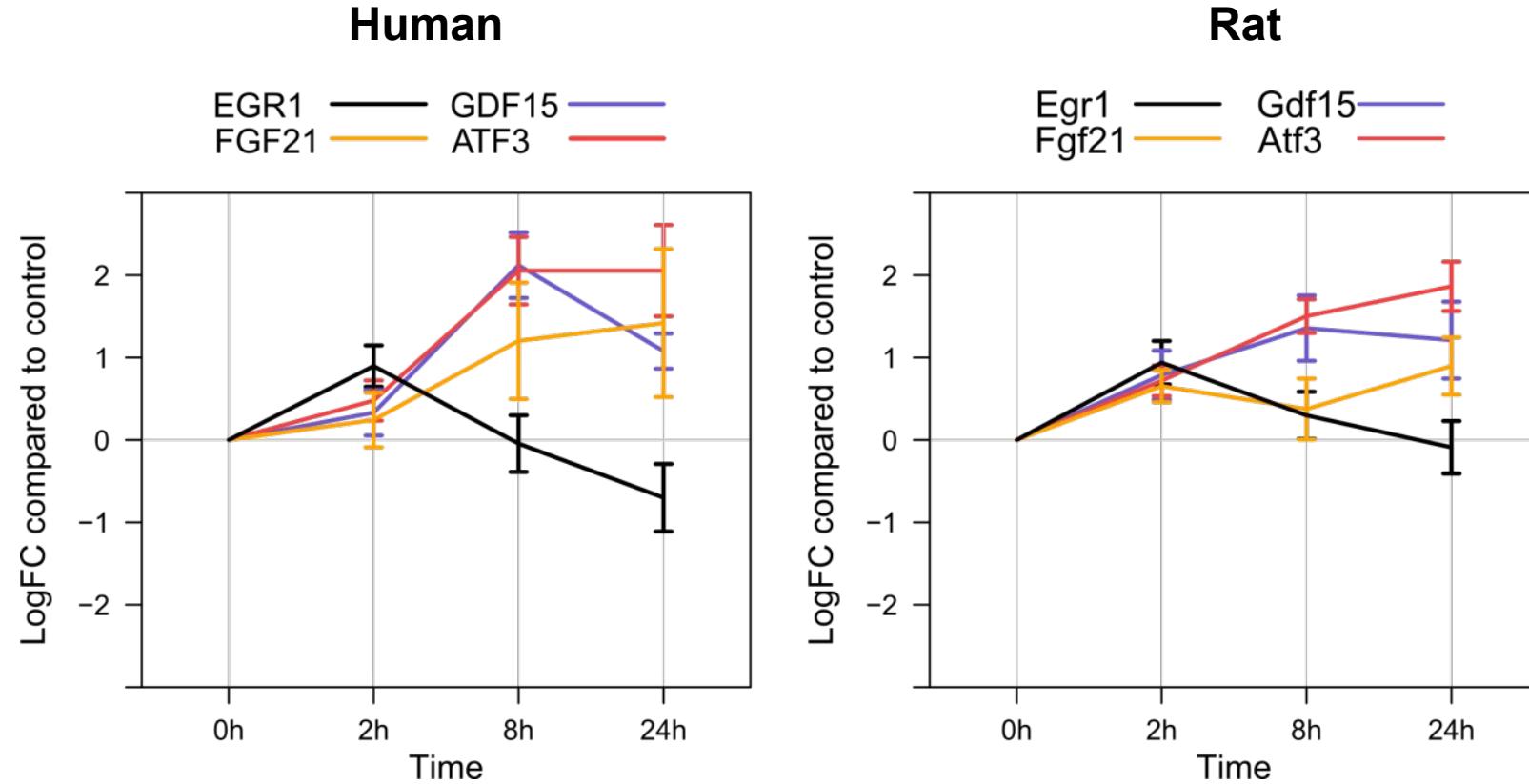
Out of six early signatures in human, four are early signatures of progressive profiles in rat: Egr1, Atf3, Gdf15, and Fgf21.

IL-8 does not have rat orthologue; Tob2 shows a similar pattern, but statistically was not significant.



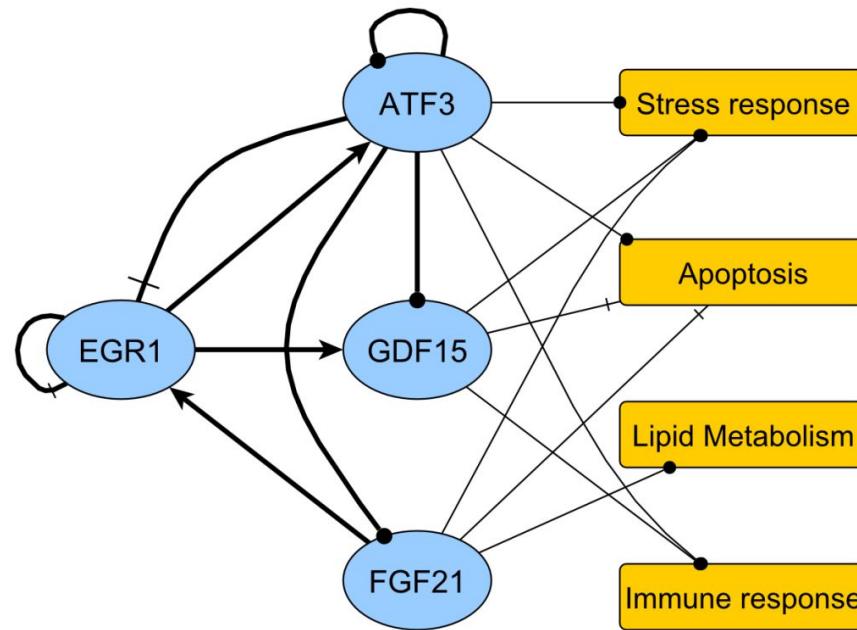
Early-response signatures in rat. Each thin line represents one treatment, and the thick line represents the average. The identification was driven by rat data only.

Conserved dynamics of the early signatures in human and rat primary hepatocytes



Lines represent average inductions, and error bars indicate 95% confidence interval of the average induction.

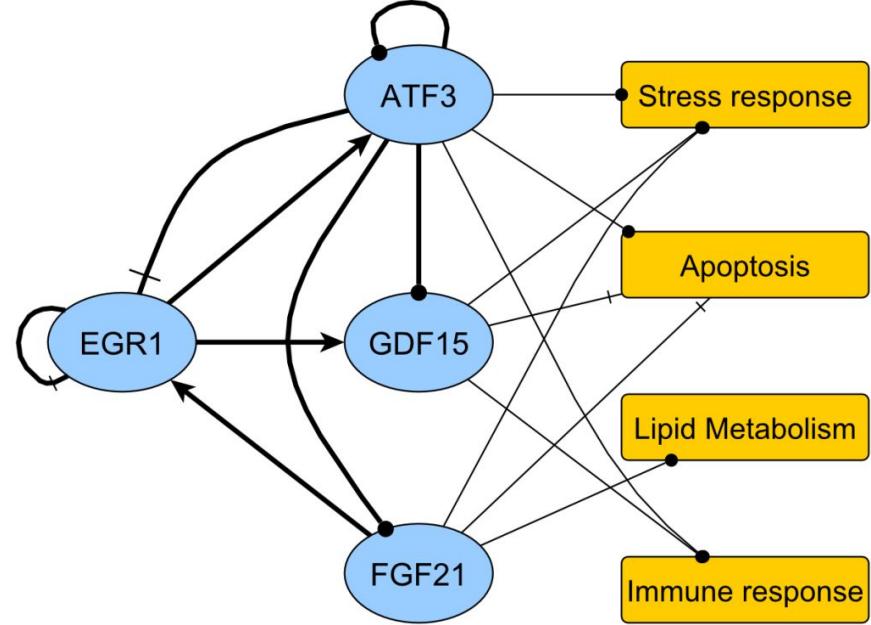
The genes form a functional network of early stress response with conserved structure and conserved dynamics



The early-response signature network, with downstream effects described in literature and annotated in functional databases

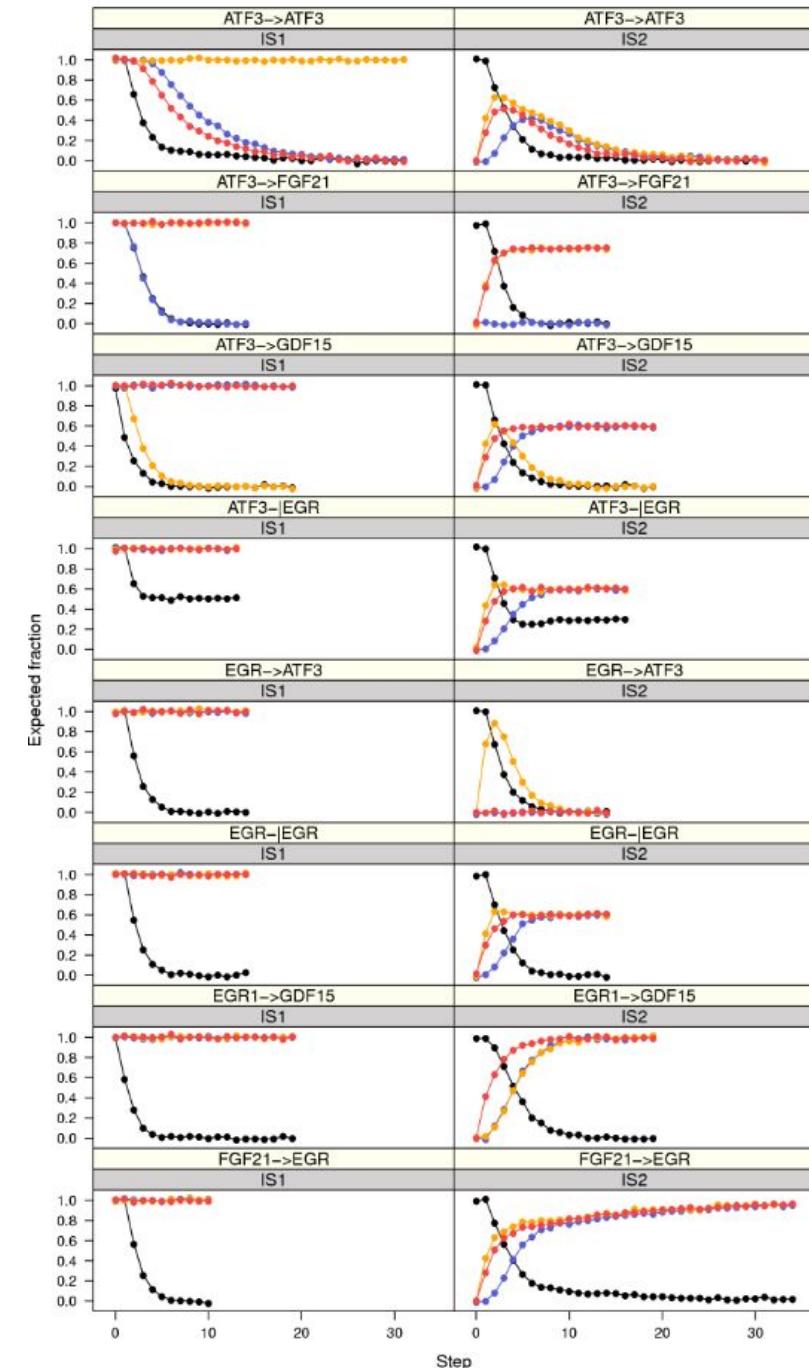
Literature search and functional annotation helped us realize the genes form a functional network

Boolean network modelling



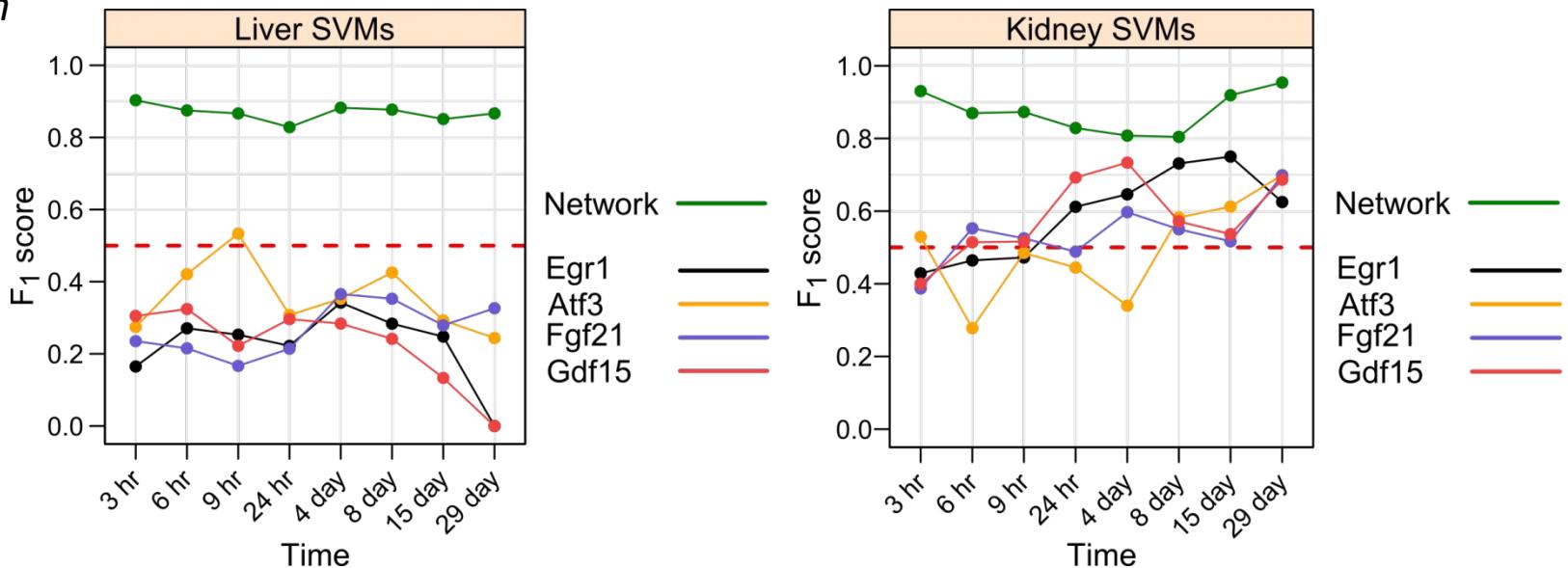
- Boolean network simulation results (Nikolaos Berntenis and Martin Ebeling, BMC Genomics 2013) support the hypothesis that **the conserved dynamics of the network in human and rat is encoded in the conserved structure of the network.**
 - Permutation results suggest that **ATF3 is an important node to maintain the network dynamics.**

Boolean network modelling revealed that the dynamics is intrinsic to the network



The network finding was translated from *in vitro* to *in vivo*, and from liver to kidney

- Support Vector Machines (SVMs) were trained to predict *in vivo* pathology between 3h and 29d using gene expression changes of Egr1, Atf3, Gdf15, and Fgf21 at 3h.
- Profiles were randomly split into training samples (80%) and test samples (20%).
- SVMs are trained by 10-fold cross-validation in training samples. Then they are tested on test samples, which mimic new, unseen data.



The predictive power of the network is translated from *in vitro* to *in vivo*, and from liver to kidney

Summary of the work with TG-GATEs

- A novel computational pipeline identified four genes - EGR1, ATF3, GDF15, and FGF21 - that are induced as early as 2h after drug administration in human and rat primary hepatocytes poised to eventually undergo cell death.
- Boolean network simulation reveals that the genes form a functional network with evolutionarily conserved structure and dynamics.
- Confirming *in vitro* findings, early induction of the network predicts drug-induced liver and kidney pathology *in vivo* with high accuracy.
- The findings are not only useful for safety assessment, but also inspired the molecular-phenotyping platform.



The Pharmacogenomics Journal (2014) 14, 208–216
 © 2014 Macmillan Publishers Limited All rights reserved 1470-269X/14
www.nature.com/tpj

OPEN

ORIGINAL ARTICLE

Data mining reveals a network of early-response genes as a consensus signature of drug-induced *in vitro* and *in vivo* toxicity

JD Zhang, N Berntsen, A Roth and M Ebeling

Gene signatures of drug-induced toxicity are of broad interest, but they are often identified from small-scale, single-time point experiments, and are therefore of limited applicability. To address this issue, we performed multivariate analysis of gene expression, cell-based assays, and histopathological data in the TG-GATES (Toxicogenomics Project-Genomics Assisted Toxicity Evaluation system) database. Data mining highlights four genes—*EGR1*, *ATF3*, *GDF15* and *FGF21*—that are induced 2 h after drug administration in human and rat primary hepatocytes poised to eventually undergo cytotoxicity-induced cell death. Modelling and simulation reveals that these early stress-response genes form a functional network with evolutionarily conserved structure and intrinsic dynamics. This is underlined by the fact that early induction of this network *in vivo* predicts drug-induced liver and kidney pathology with high accuracy. Our findings demonstrate the value of early gene-expression signatures in predicting and understanding compound-induced toxicity. The identified network can empower first-line tests that reduce animal use and costs of safety evaluation.

The Pharmacogenomics Journal (2014) **14**, 208–216; doi:10.1038/tpj.2013.39; published online 12 November 2013

Keywords: compound-induced toxicity; early-response genes; gene signature; TG-GATES; toxicogenomics

Zhang *et al.*, J Pharmacogenomics, 2014

Computational biology and bioinformatics help identifying safer drugs

Looking around and looking forward

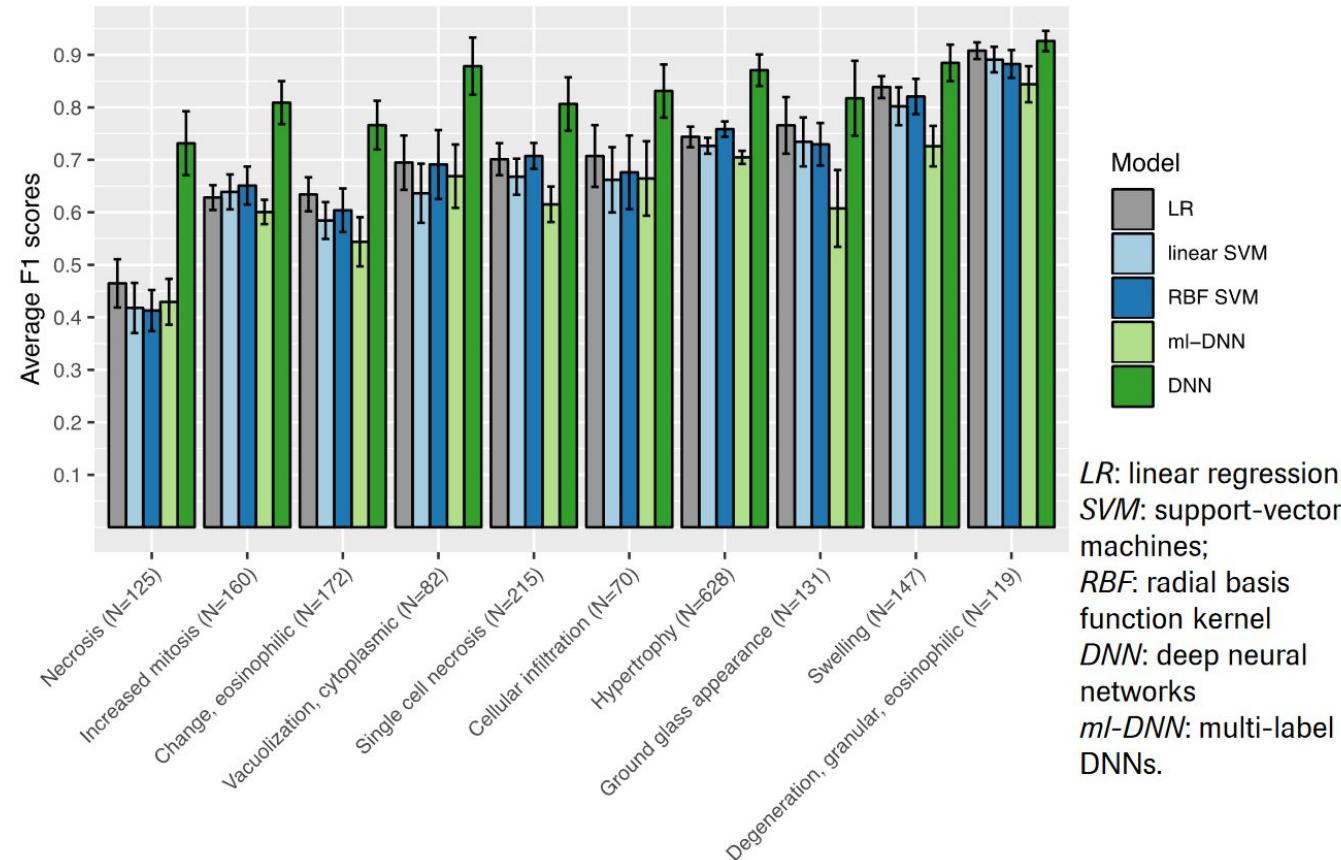
- **Selected further work by external groups**
 - Sutherland *et al.* (Lily), PLOS Comp Biol 2016, confirmed the difficulty to directly translate between different systems
 - El-Hachem *et al.* (U Montreal), Environ Health Perspect 2016, confirmed that early toxicological response occurring in animals is recapitulated in human and rat primary hepatocyte cultures at the molecular level
 - Thiel *et al.*, (RWTH Aachen), PLOS Comp Biol 2017, used physiologically-based pharmacokinetic modeling to characterize the transition from efficacious to toxic doses.
 - Shimada & Mitchison (Harvard), Mol Sys Bio 2019, used machine learning to characterize system-level response to drugs and toxicants in TG-GATEs, and pinpointing underlying molecular mechanisms.
- **What we are doing**
 - Gain molecule-level understanding of drug-induced histopathology
 - Apply the knowledge to accelerate development and reduce attrition rate of new drugs
 - Leveraging stem-cell technology and omics for drug discovery & personalized safety

The four-gene network is not the end, but a start, for the community and for us

An application of supervised machine learning for drug-induced liver histopathology prediction

		Predicted labels		
		1	0	
Actual labels (observations)	1	True Positive (TP)	False Negative (FN)	Recall=TPR (True Positive Rate) $TPR = \frac{TP}{TP+FN}$
	0	False Positive (FP)	True Negative (TN)	Specificity = $\frac{TN}{TN+FP}$ False Positive Rate: $FPR = \frac{FP}{FP+TN}$
	Precision $\frac{TP}{TP+FP}$	False Negative Rate $\frac{FN}{TN+FN}$	Accuracy $\frac{TP+TN}{TP+TN+FP+FN}$	

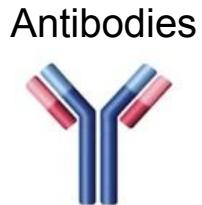
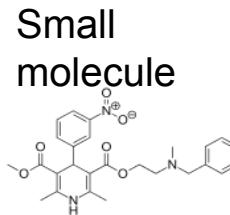
$$F_1 = \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}}$$



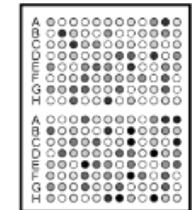
Performance of different machine learning algorithms in the task of drug-induced liver histopathology prediction using differential gene expression data. ISMB (2019) Poster by Fang *et al.* F_1 score is the harmonic mean of precision and recall.

Molecular Phenotyping

A workflow to quantify expression of pre-defined pathway reporter genes at early time points after perturbation to infer pathway activities, which may predict late-onset cellular phenotypes



~1000 pathway reporter genes



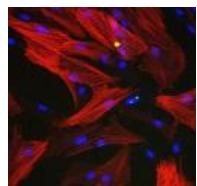
Next-generation sequencing



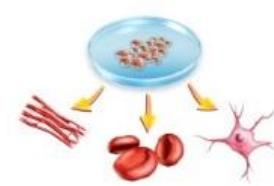
Therapeutic candidates

Early time point (3-12h)

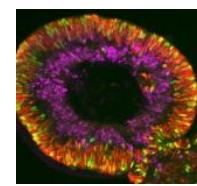
Human *in vitro* disease models



Cell lines/
primary cells



iPS-derived cells
(opt. genome
editing)



Micro-phy
siological
systems

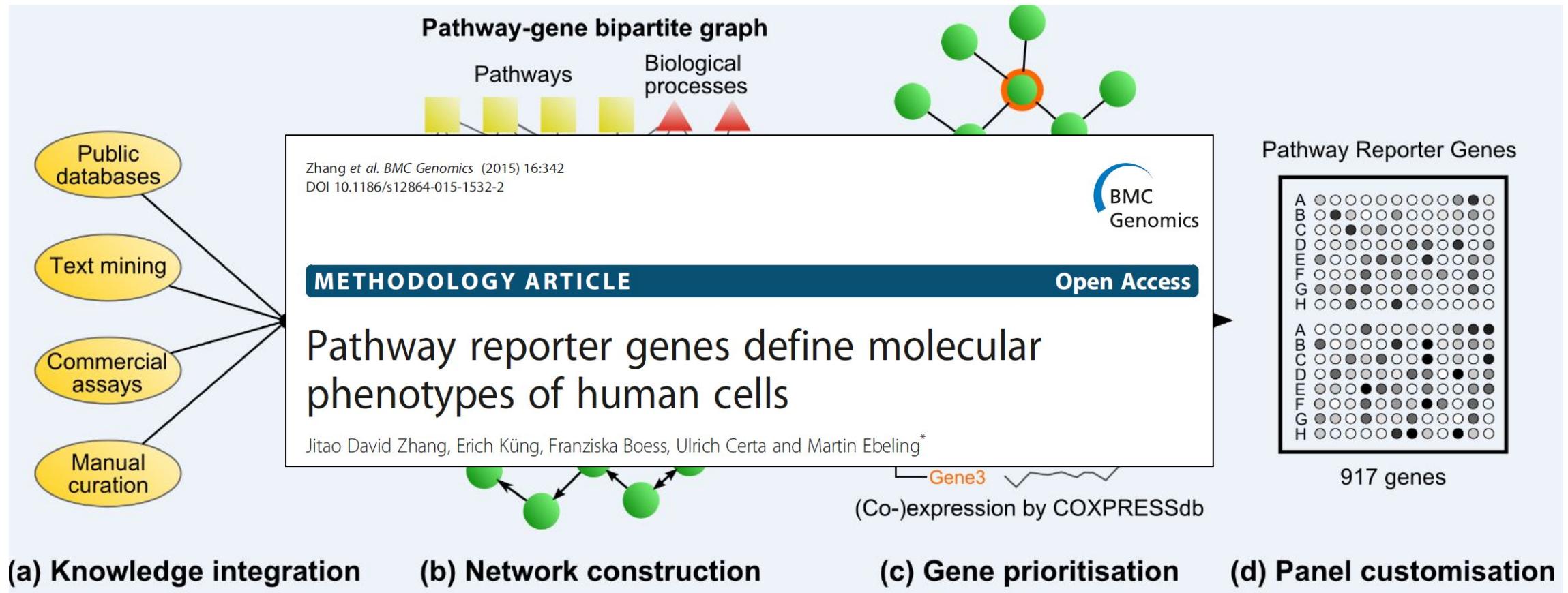
Molecular phenotyping

What pathways are perturbed by each compound?

Applications as screening tool:

- Cluster compounds based on pathway profiles
- Detect false-positive hits in a phenotypic screen
- Correlate pathway activity with phenotypic readouts

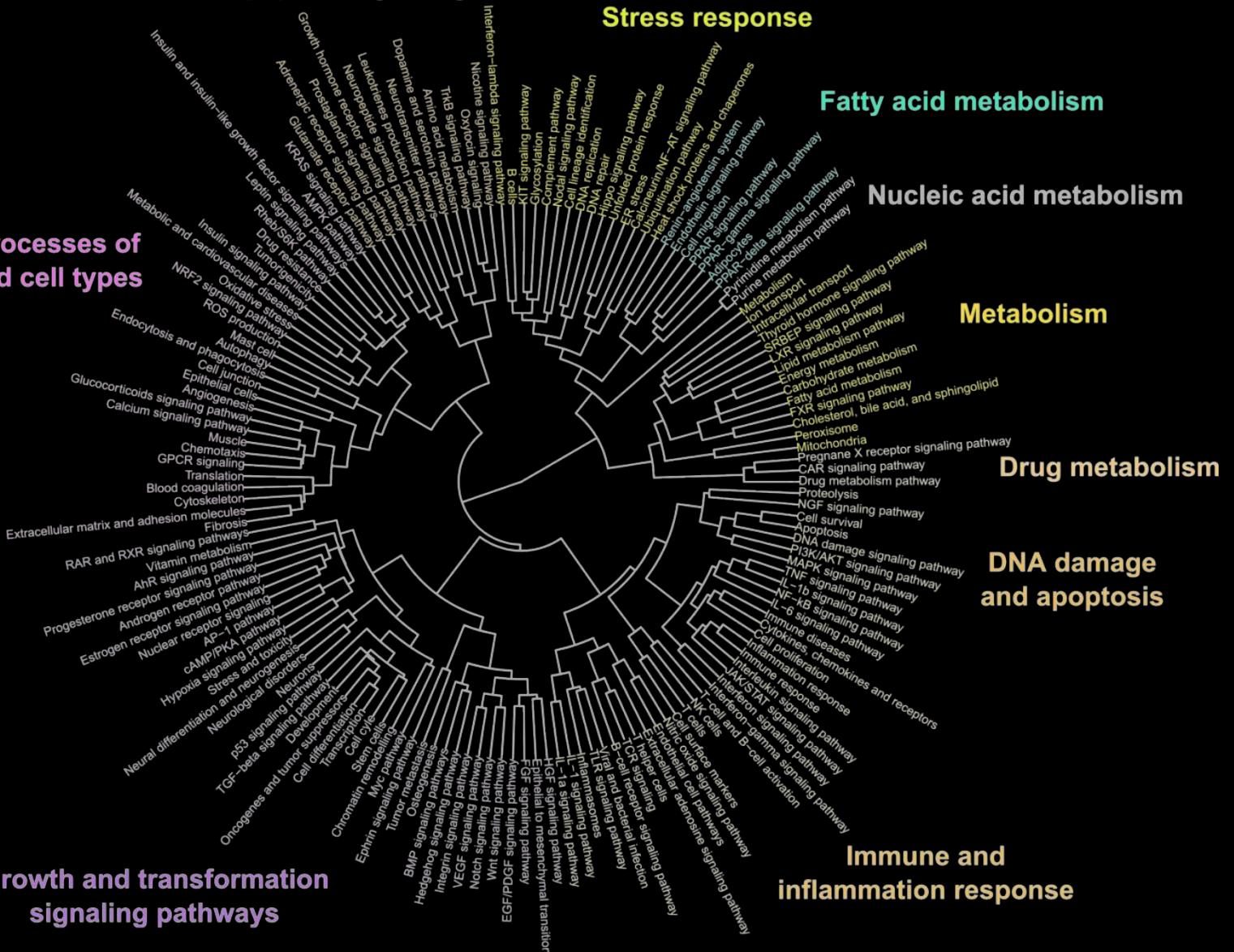
Pathway Reporter Genes



Hormone and neuropeptide signaling

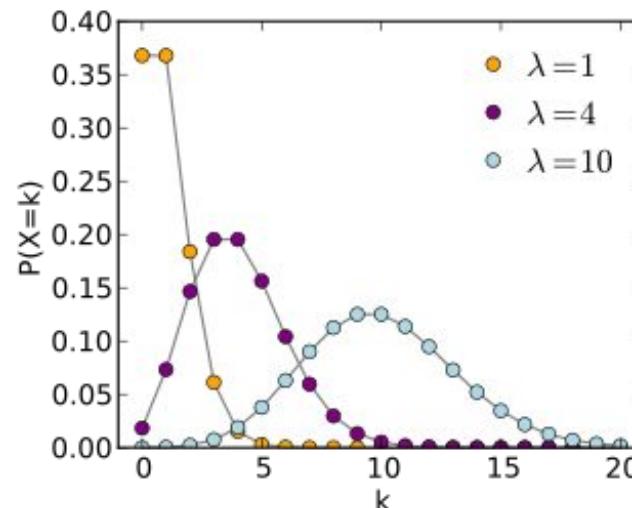
Biological processes of differentiated cell types

Growth and transformation signaling pathways



Difference in statistical modelling of microarray data and next-generation sequencing count data

- Microarray data: log-normal distributed, for instance implemented in the *limma* package of R/Bioconductor.
- Bulk RNA-sequencing data: Negative-Binomial distributed (or Poisson with overdispersion), for instance implemented in both *edgeR* and *DESeq2* package of R/Bioconductor.
- Single-cell data: some authors recently suggest that negative-binomial or Poisson distribution suffices if the cell population is homogenous (Kim, Tae Hyun, Xiang Zhou, and Mengjie Chen. 2020. “[Demystifying ‘Drop-Outs’ in Single-Cell UMI Data.](#)” *Genome Biology* 21 (1): 196), though many tools assume zero-inflated negative-binomial model.

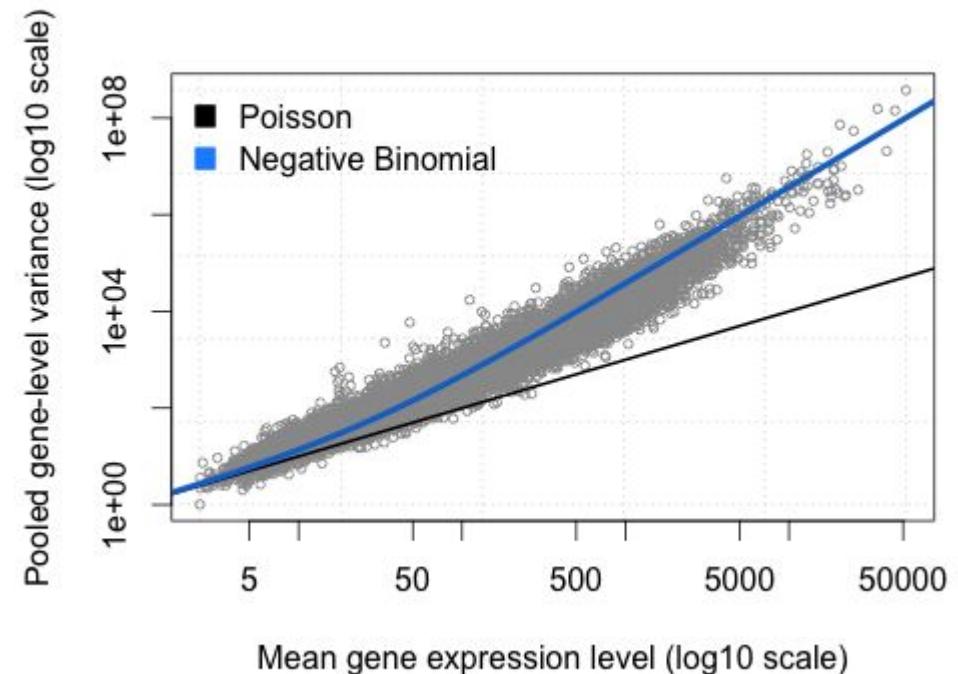


Poisson distribution with three rate parameters, from [Wikimedia](#), reused with the CC Attribution 3.0 license

From Poisson distribution to Negative Binomial Distribution

Two definitions of Negative-Binomial distribution

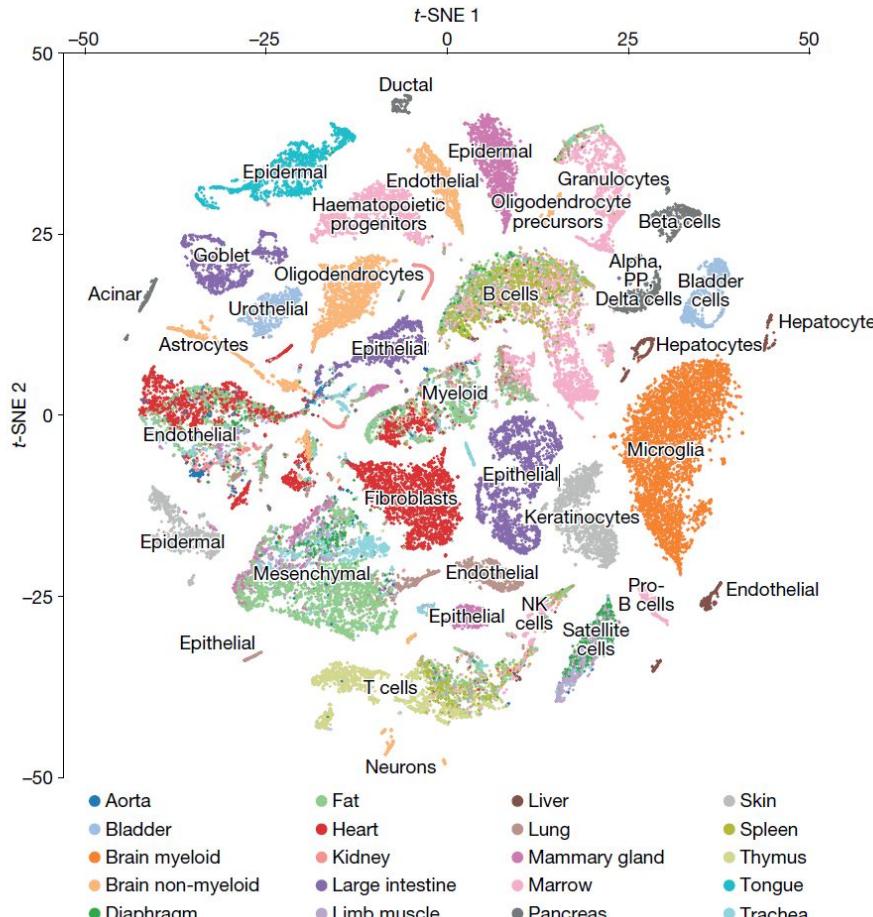
1. The number of failures seen before getting n successes (the inverse of *Binomial Distribution*, which the number of successes in n independent trials)
2. Poisson-Gamma mixture distribution, weighted mixture of *Poisson* distributions, where the rate parameter has an uncertainty modelled by a *Gamma* distribution.



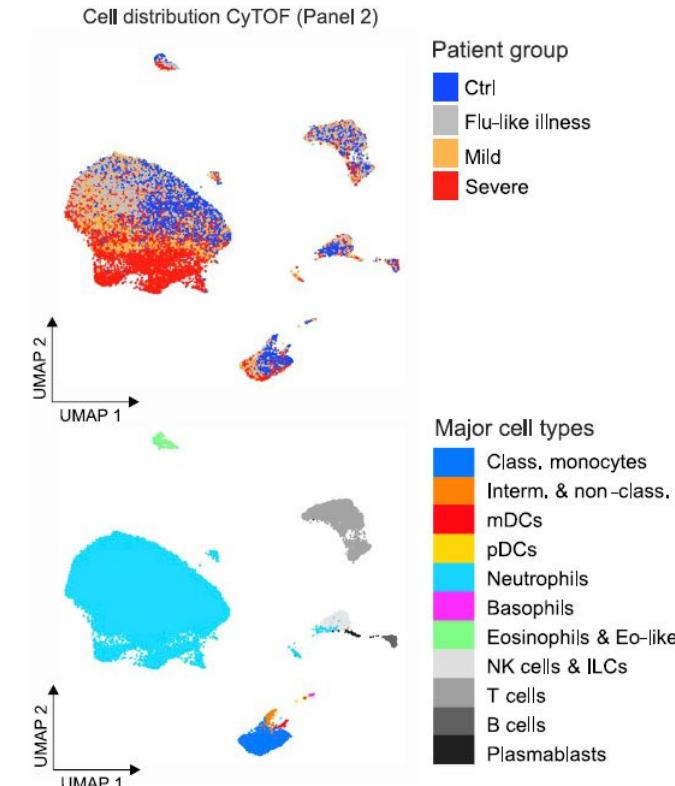
Credit of Jesse Lipp,
bioramble.wordpress.com

Commonly used dimensionality reduction techniques

- Principal component analysis (PCA)
- t-SNE (t-distributed Stochastic Neighbor Embedding)
- UMAP (Uniform Manifold Approximation and Projection) [A great talk by Leland McInnes, the developer of UMAP, a mathematician, Ph.D. In Profinite Lie Rings]
- For a recent overview of dimensionality reduction techniques and their applications in biology, see Nguyen, Lan Huong, und Susan Holmes. "Ten Quick Tips for Effective Dimensionality Reduction". *PLOS Computational Biology* 15, Nr. 6 (20. Juni 2019): e1006907.

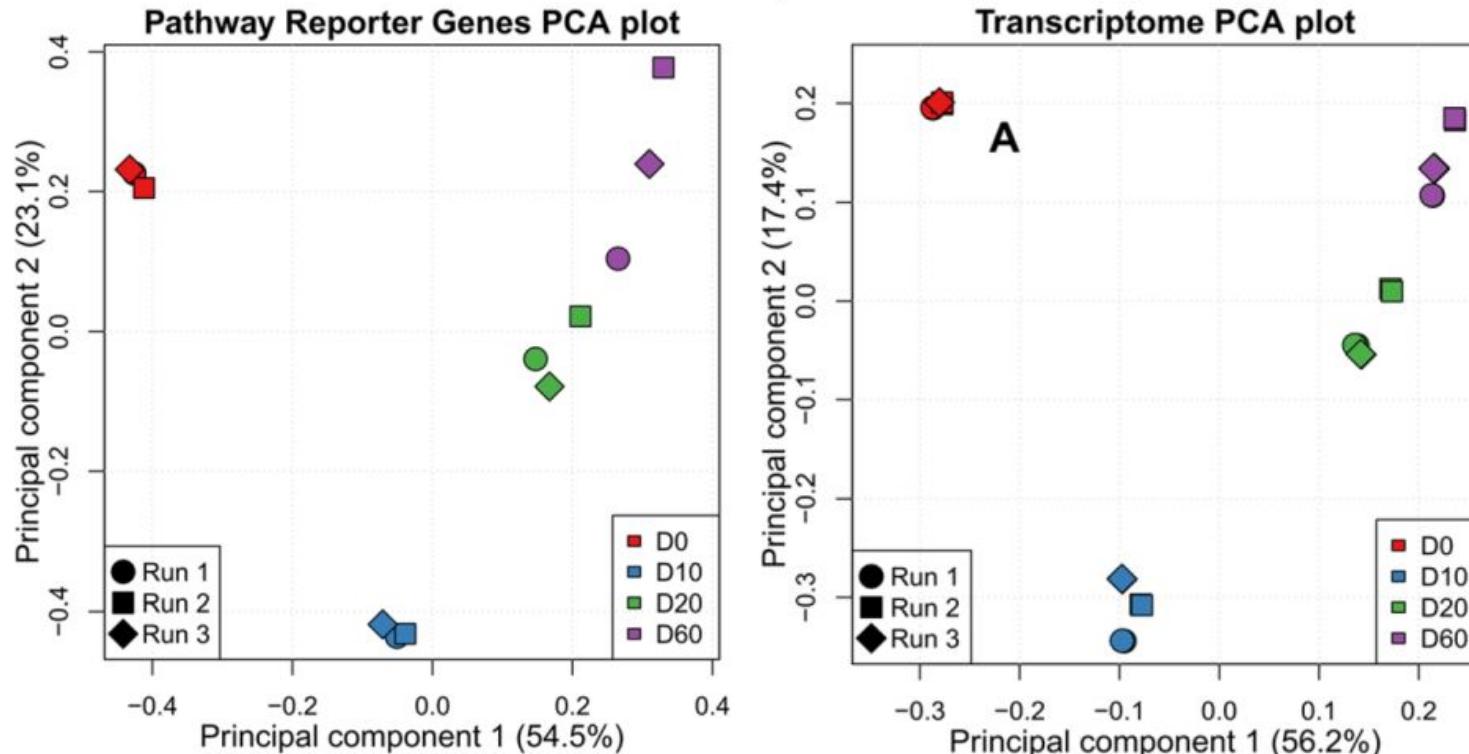
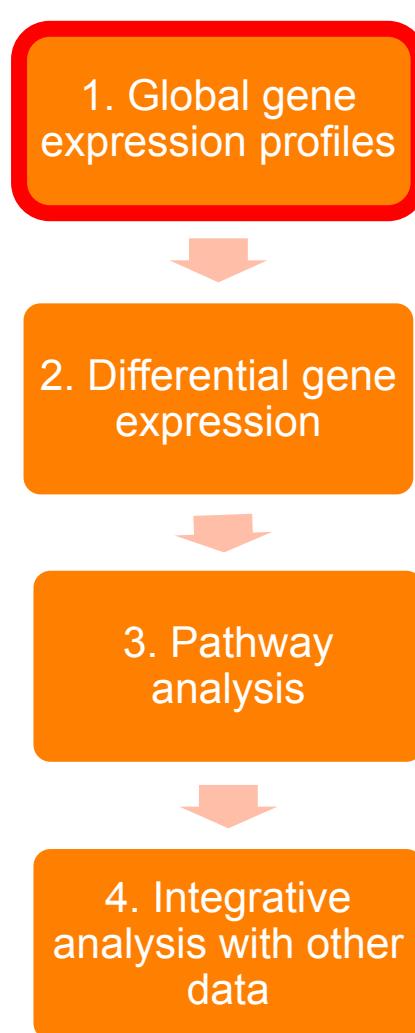


The Tabula Muris Consortium. 2018. "Single-Cell Transcriptomics of 20 Mouse Organs Creates a Tabula Muris." *Nature* 562 (7727): 367.



Schulte-Schrepping, Jonas, Nico Reusch, Daniela Paclik, Kevin Baßler, Stephan Schlickeiser, Bowen Zhang, Benjamin Krämer, et al. 2020. "Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment." *Cell* 182 (6): 1419-1440.e23.

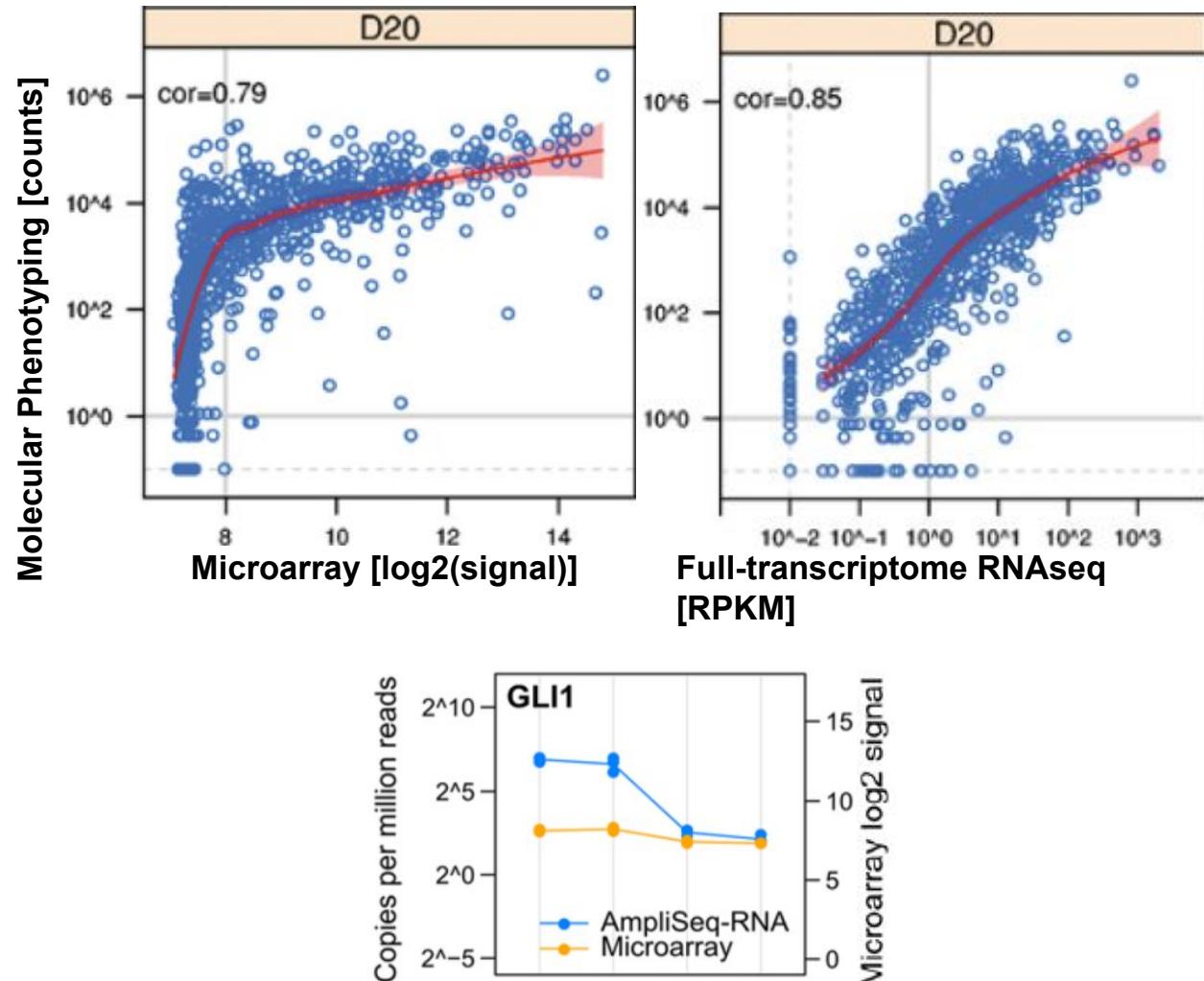
Data Analysis



Pathway reporter genes faithfully capture global gene expression patterns

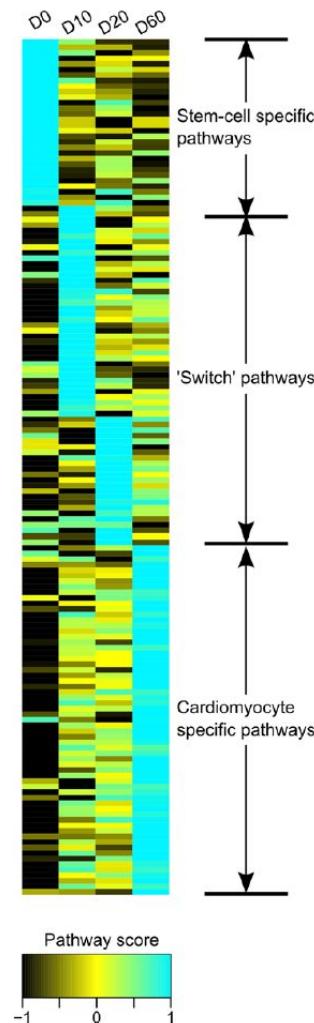
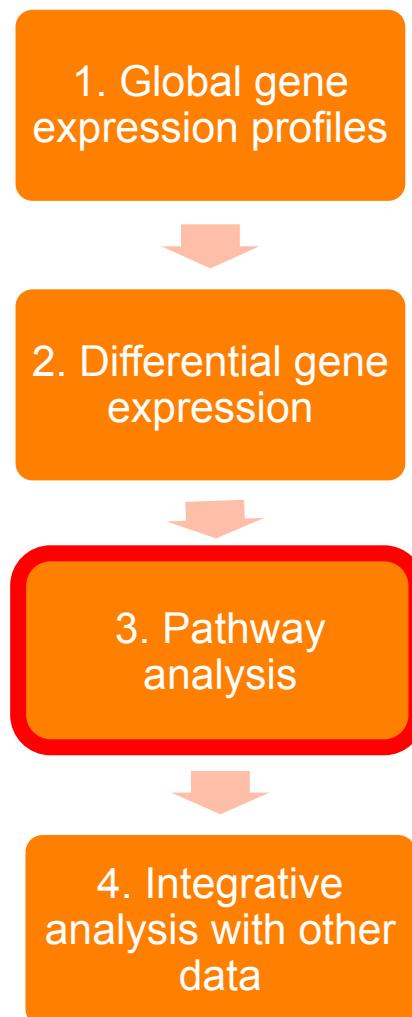
Data Analysis

1. Global gene expression profiles
2. Differential gene expression
3. Pathway analysis
4. Integrative analysis with other data



AmpliSeq, like RNA-seq, shows higher dynamic range than hybridization-based platforms

Data Analysis



Activity patterns of 154 human metabolic and signaling networks during differentiation of induced pluripotent stem-cells (iPS) into cardiomyocytes.

Cyan: Pathway is activated

Black: Pathway is suppressed

Pathway reporter genes inform about pathway activity patterns

Data Analysis

1. Global gene expression profiles



2. Differential gene expression

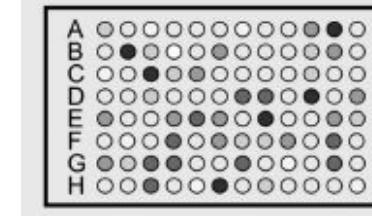


3. Pathway analysis

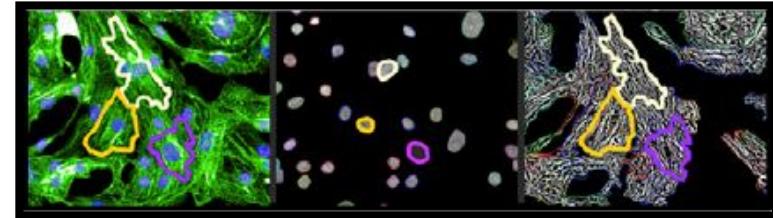


4. Integrative analysis with other data

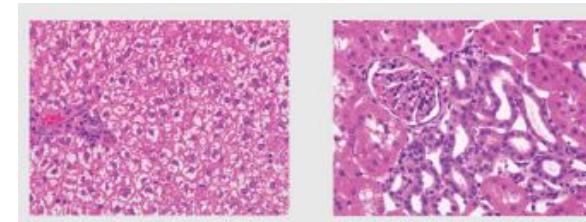
In vitro assay readouts



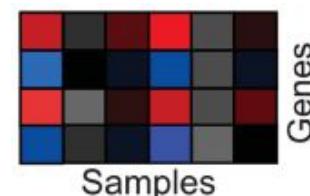
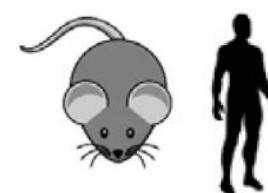
High-content microscopy



Histopathology records



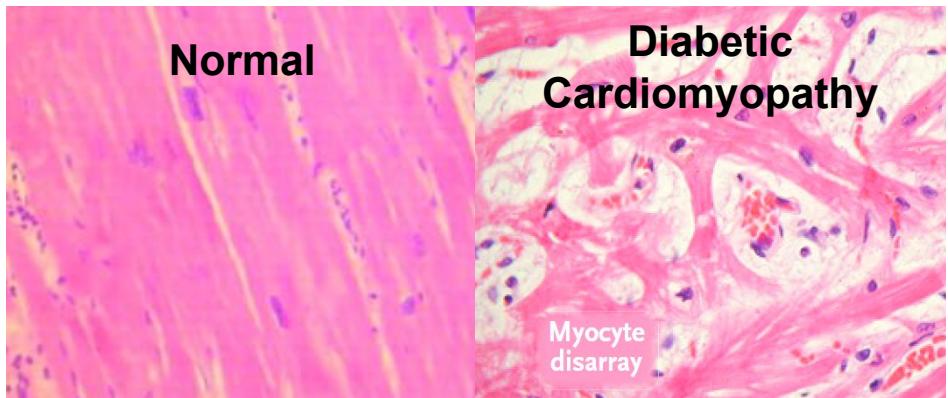
Gene signatures



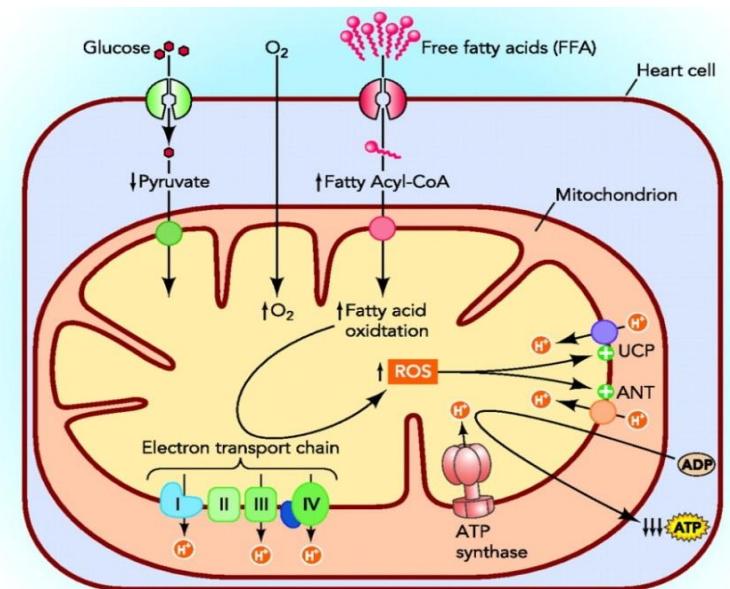
Molecular phenotyping allows data integration

Cell-based model of diabetic cardiomyopathy for phenotypic screening

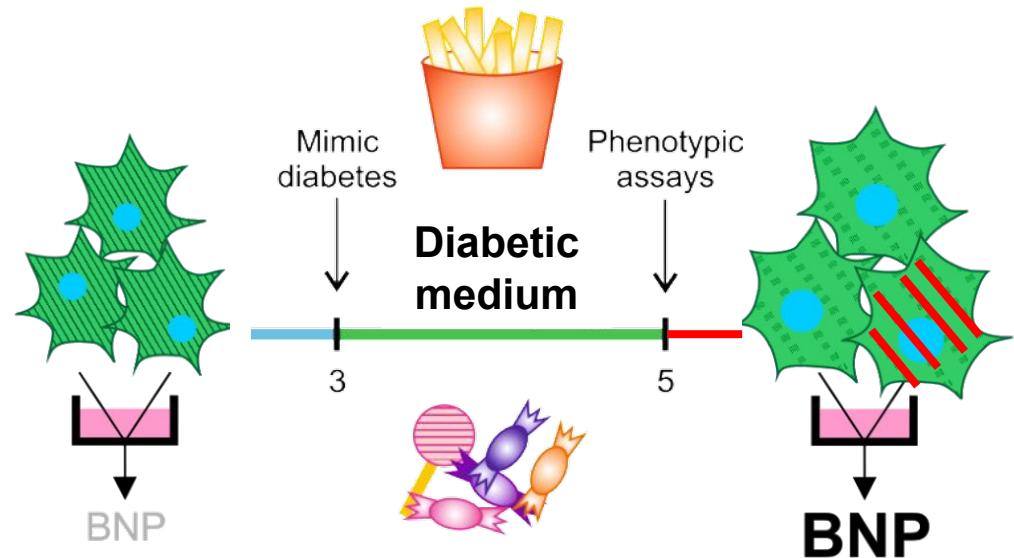
Metabolic dysfunction promotes cardiomyopathy



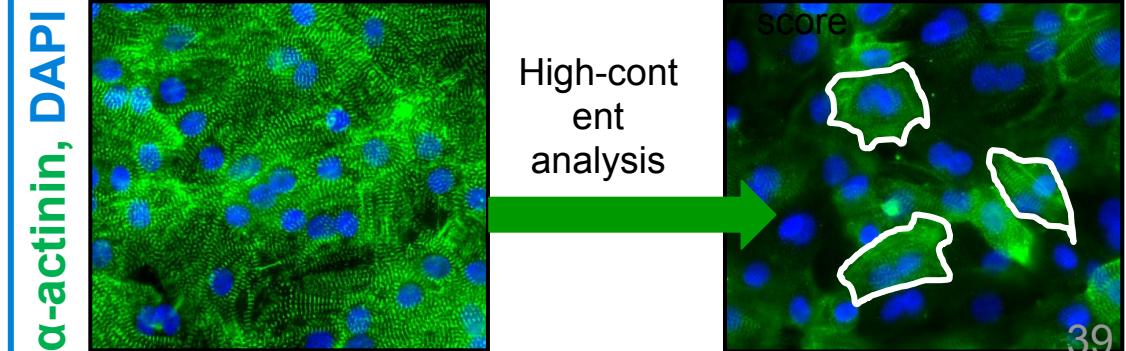
Diabetic cardiomyocyte metabolism



iPS-derived cardiomyocyte model

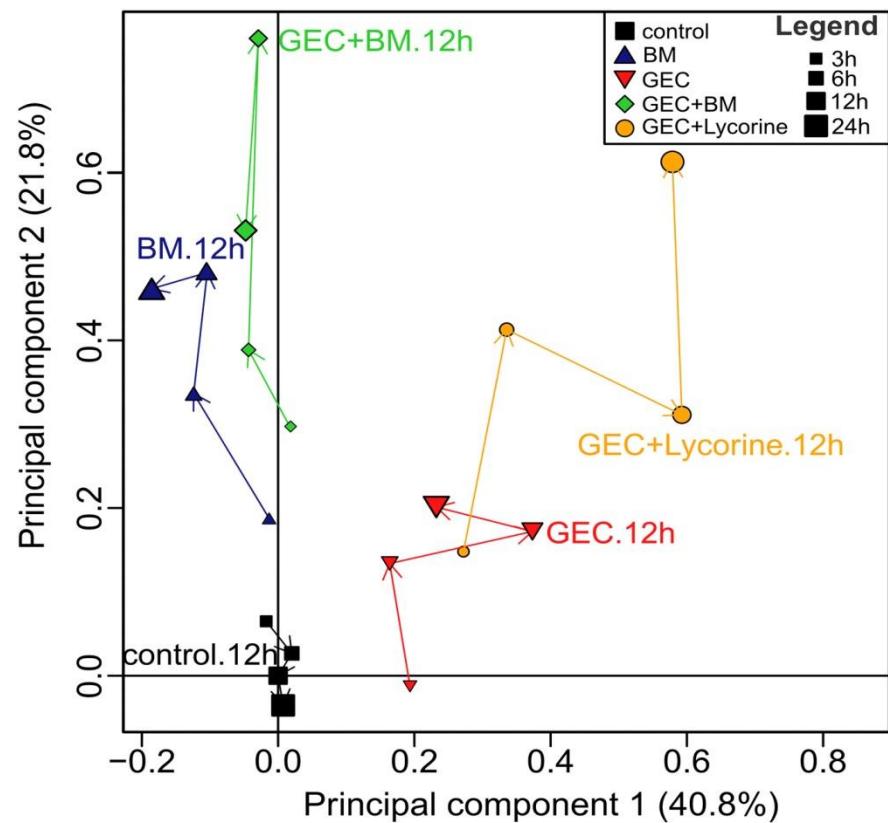


Glucose, Insulin, Fatty Acids
Cortisol, Endothelin-1



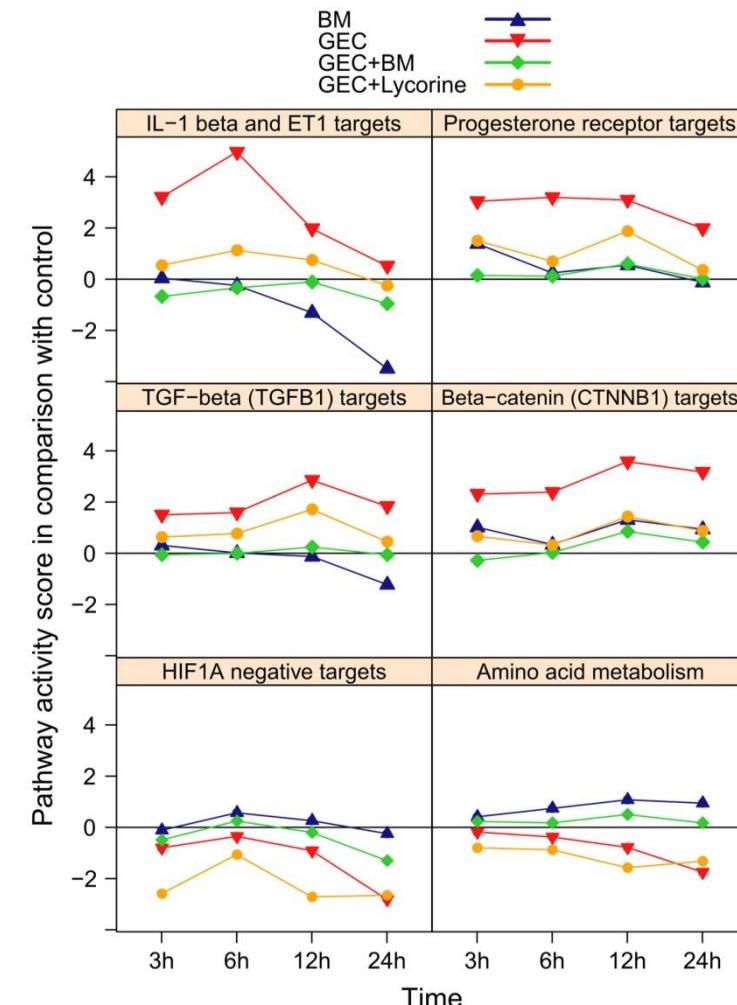
Determining the optimal time-point for molecular phenotyping

Maximal dynamic range at 12 hours



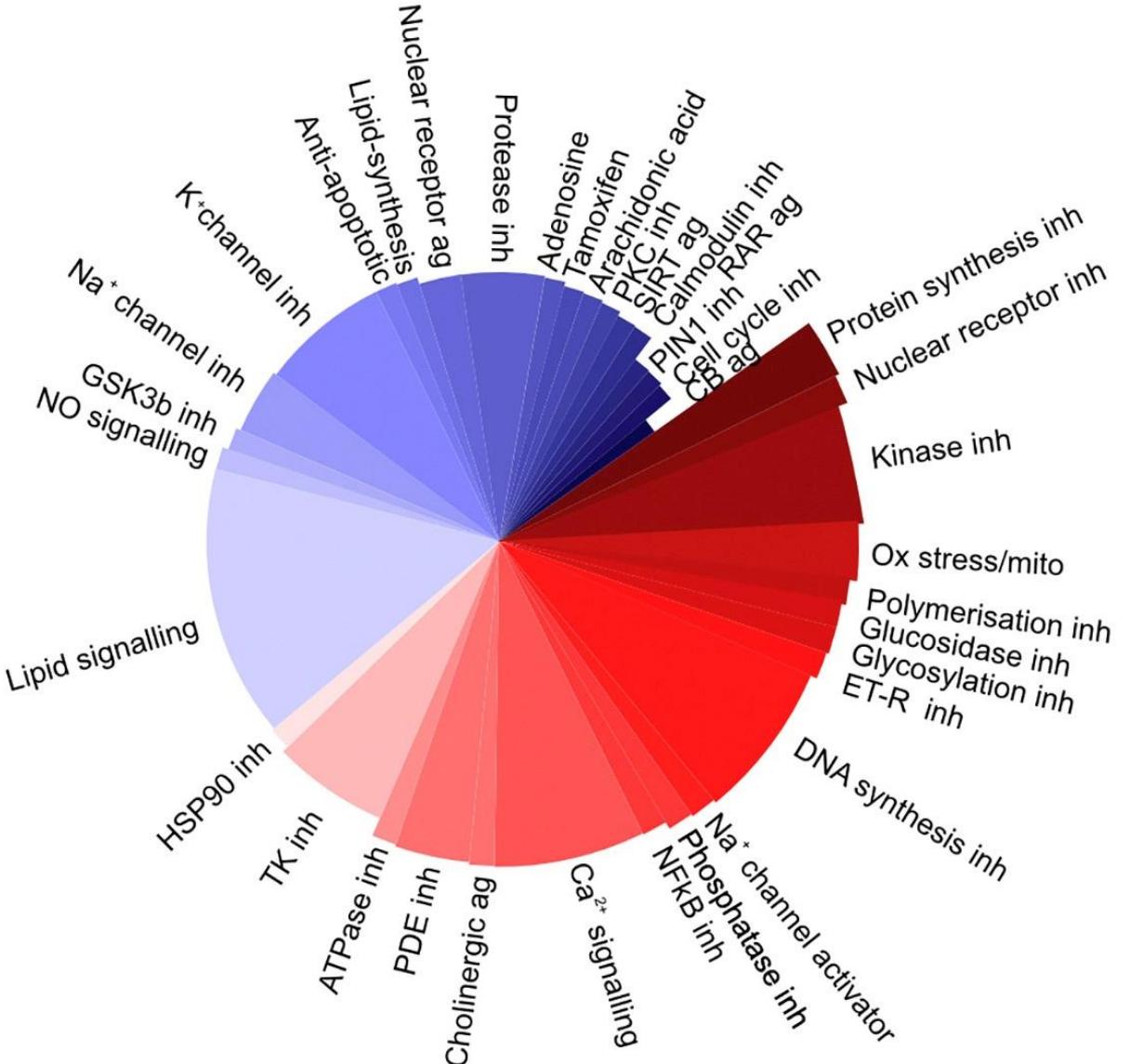
Phenotypic screening performed after 48 hours

Lycorine and positive control manipulate similar pathways

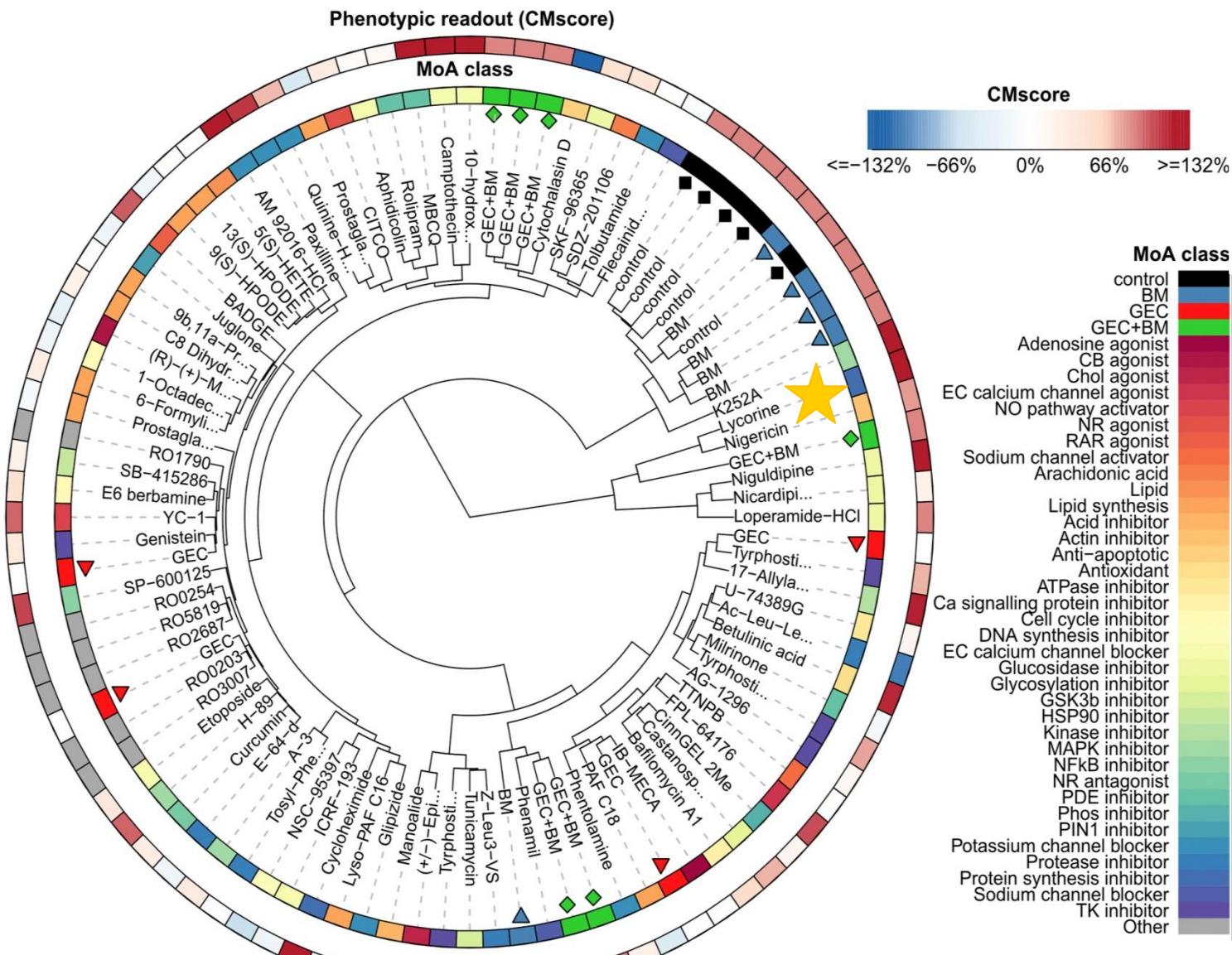


Mechanistic enrichment of screening data

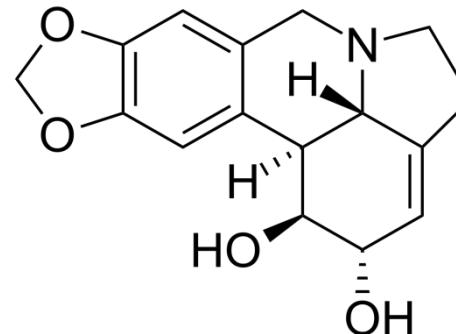
Integration of molecular and phenotypic information



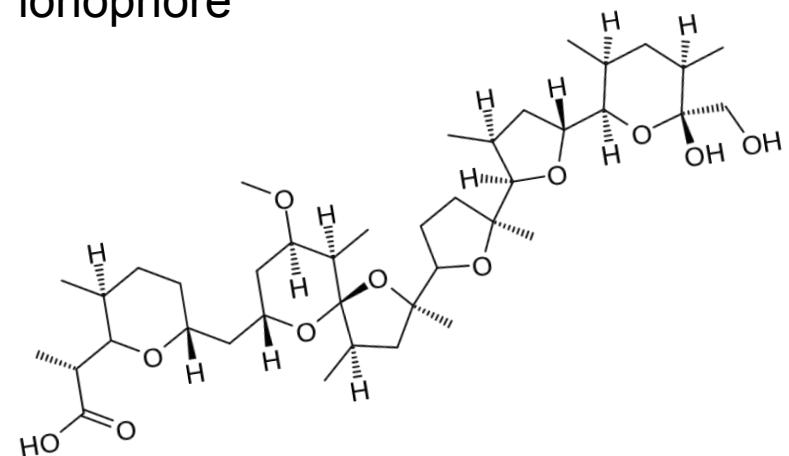
Integration of molecular and phenotypic information



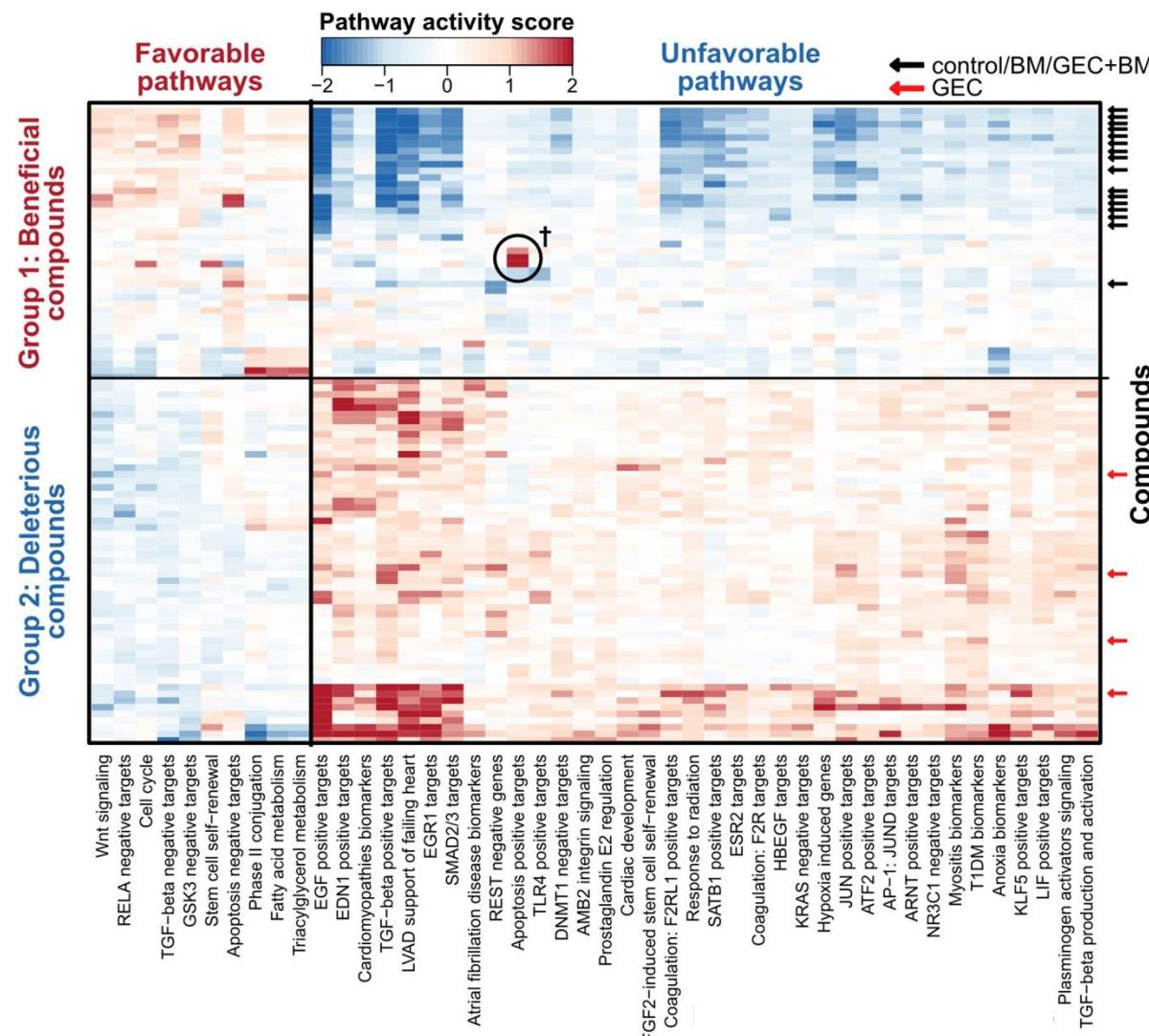
Lycorine: Protein synthesis inhibitor



Nigericin: Potassium ionophore



Clustering analysis of compounds and pathway responses



Pathways associated with CM score

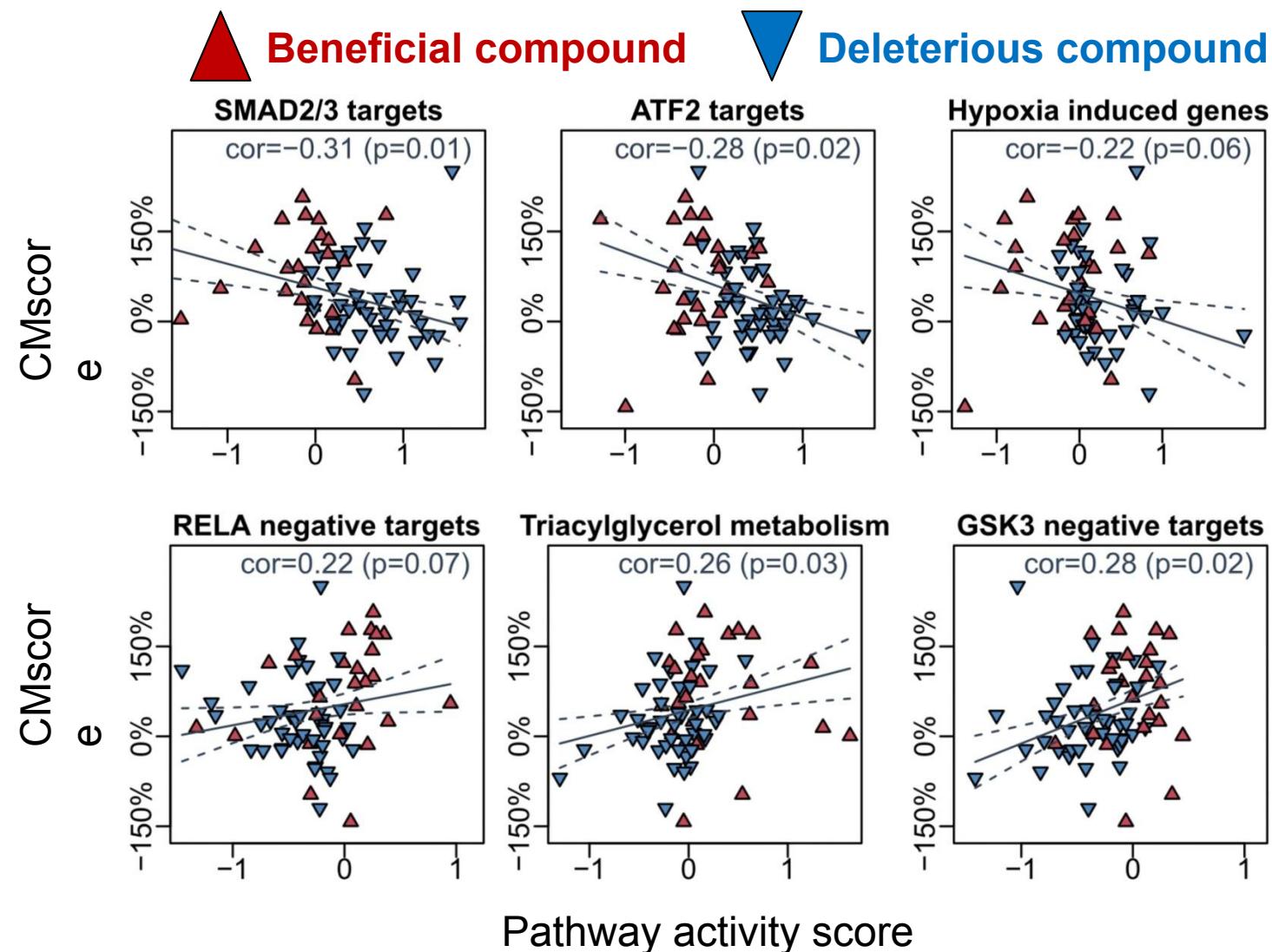
Beneficial compounds regulate **favorable** pathways positively and **unfavorable** pathways negatively

Beneficial compounds have **higher CMscore**

Deleterious compounds regulate **unfavorable** pathways positively and **favorable** pathways negatively

Deleterious compounds have **lower** CMscore

Beneficial compounds generate specific pathway signatures



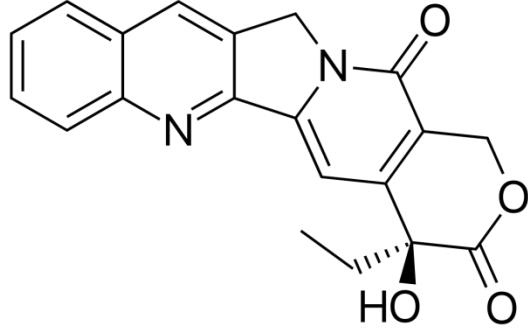
Beneficial compounds negatively regulate

Beneficial compounds positively regulate

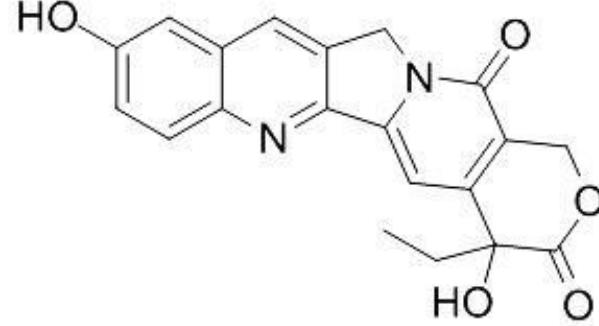
Pathway signatures can be monitored during screening campaigns for maintained beneficial mechanistic effects

Molecular phenotyping allows filtering of undesirable molecules

Camptothecin

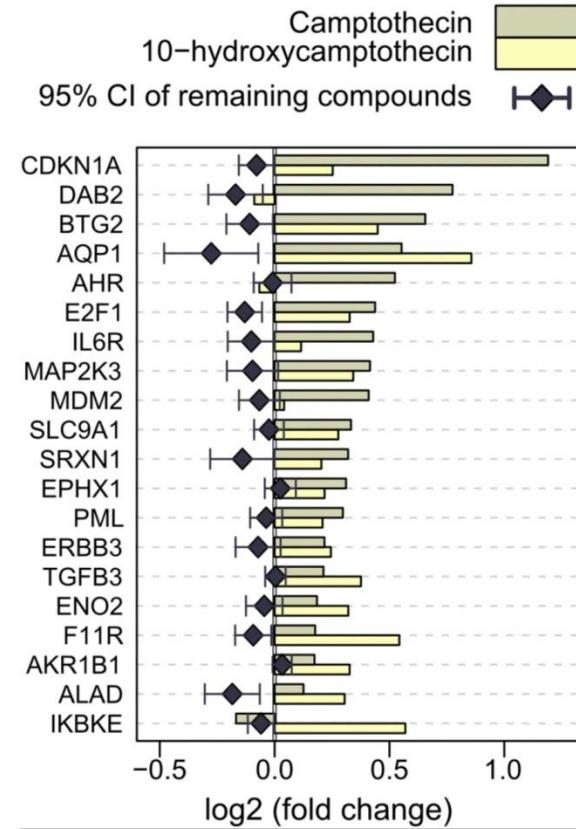


10-hydroxycamptothecin



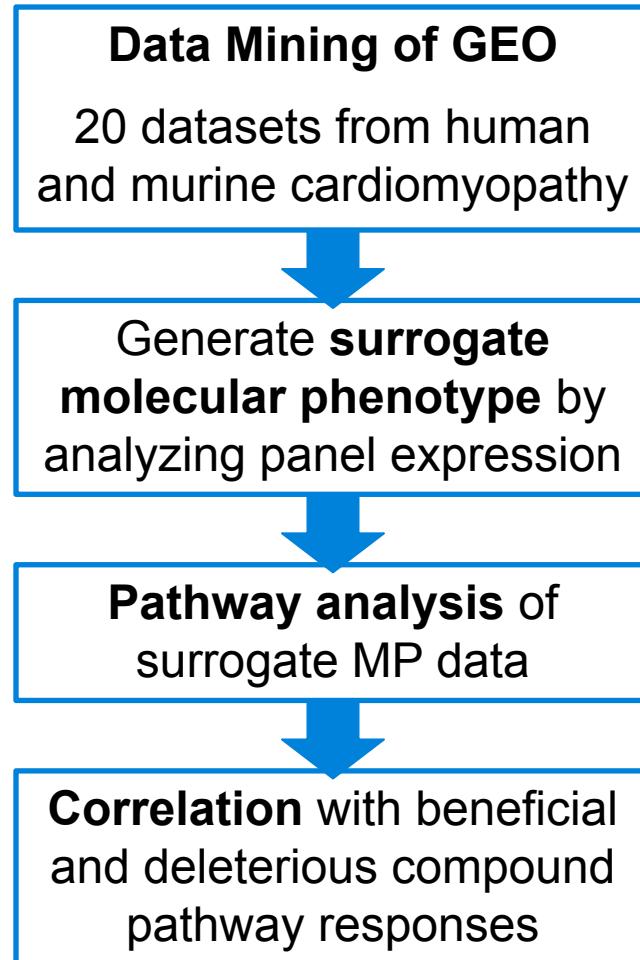
- Topoisomerase inhibitors
- Produced high CMscore in the phenotypic assay
- Identified as ‘hits’
- Cluster with beneficial compounds

Induce target genes of apoptosis

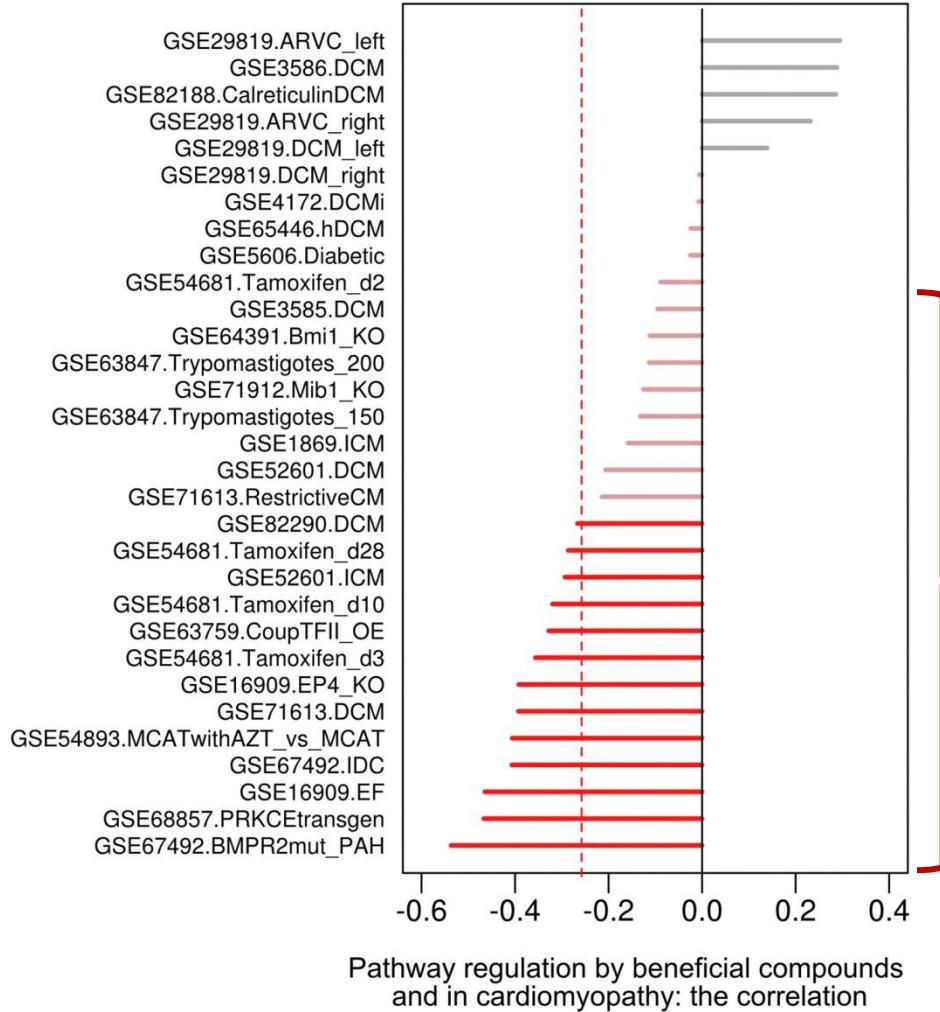


Compounds with undesirable pathway profiles can be eliminated from further testing

Beneficial compound signatures are downregulated in cardiomyopathy samples

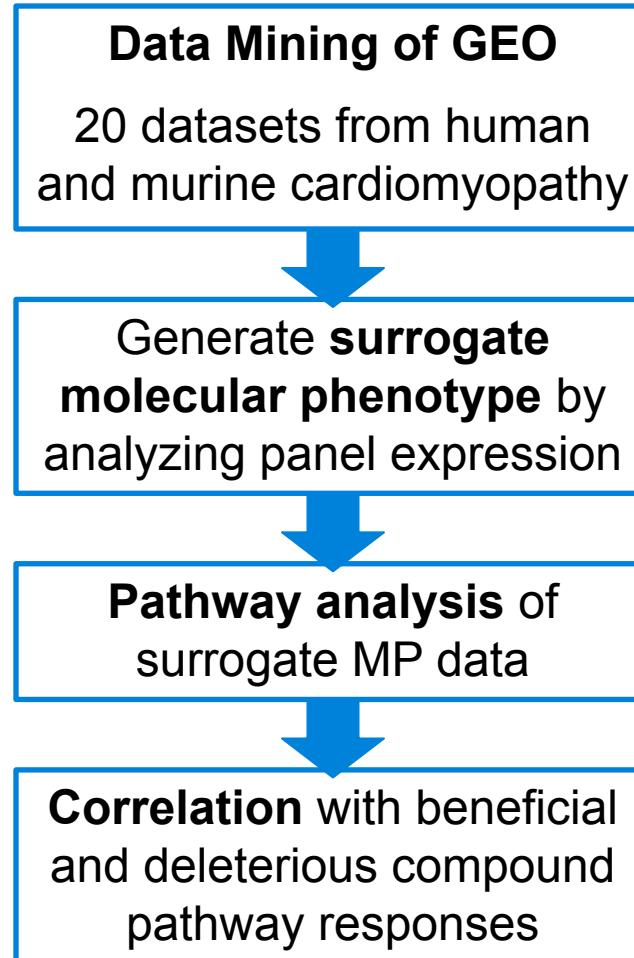


Human and animal cardiomyopathy studies

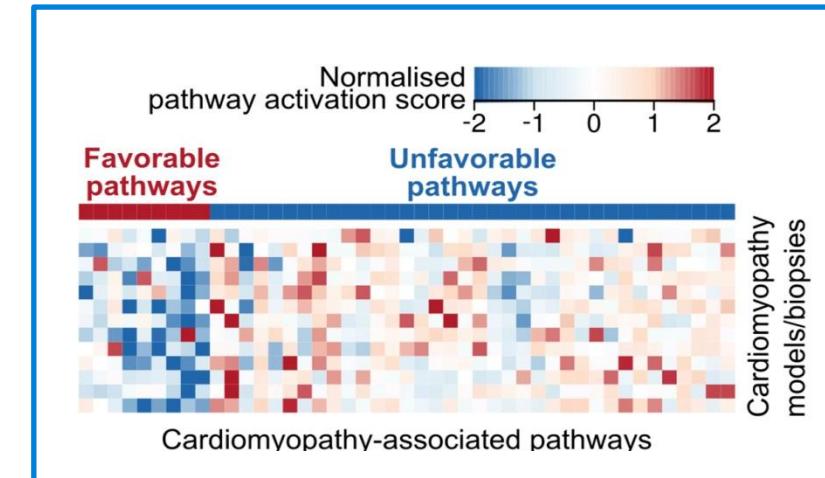


GEO=[NCBI Gene Expression Omnibus](#)

Beneficial compound signatures are downregulated in cardiomyopathy samples



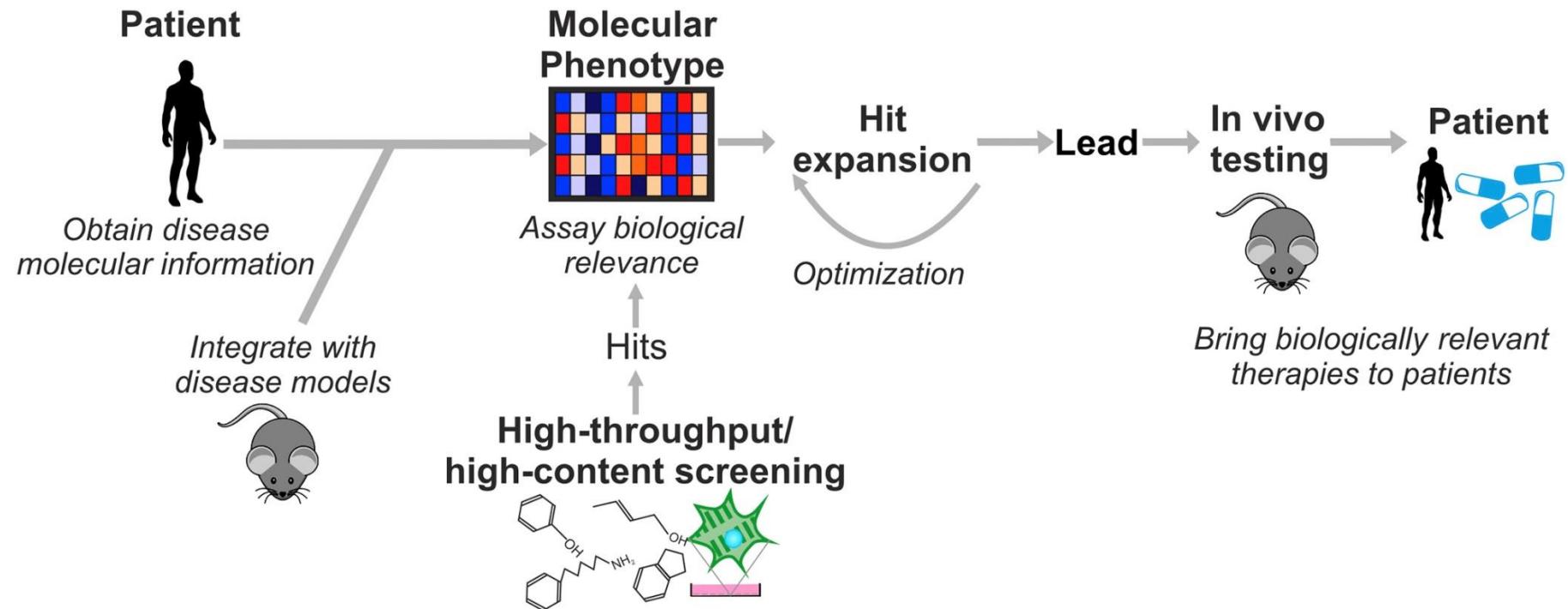
Favorable pathways tend to be down-regulated in patients/animal models



Unfavorable pathways tend to be up-regulated in patients/animal models

Molecular phenotyping can enrich screening campaigns to select compounds with profiles with biological relevance to patients

Molecular Phenotyping empowers Phenotypic Drug Discovery

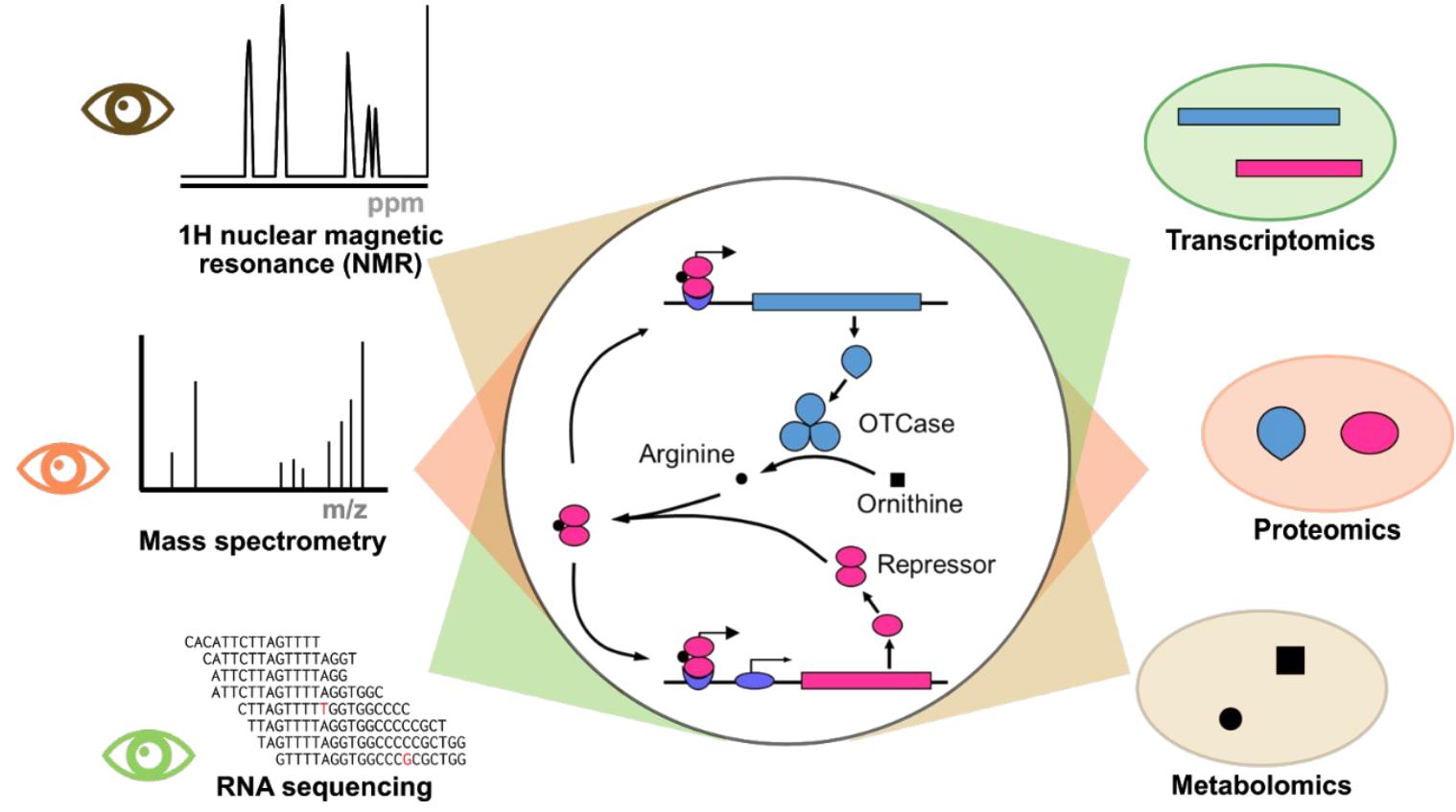


Molecular phenotyping

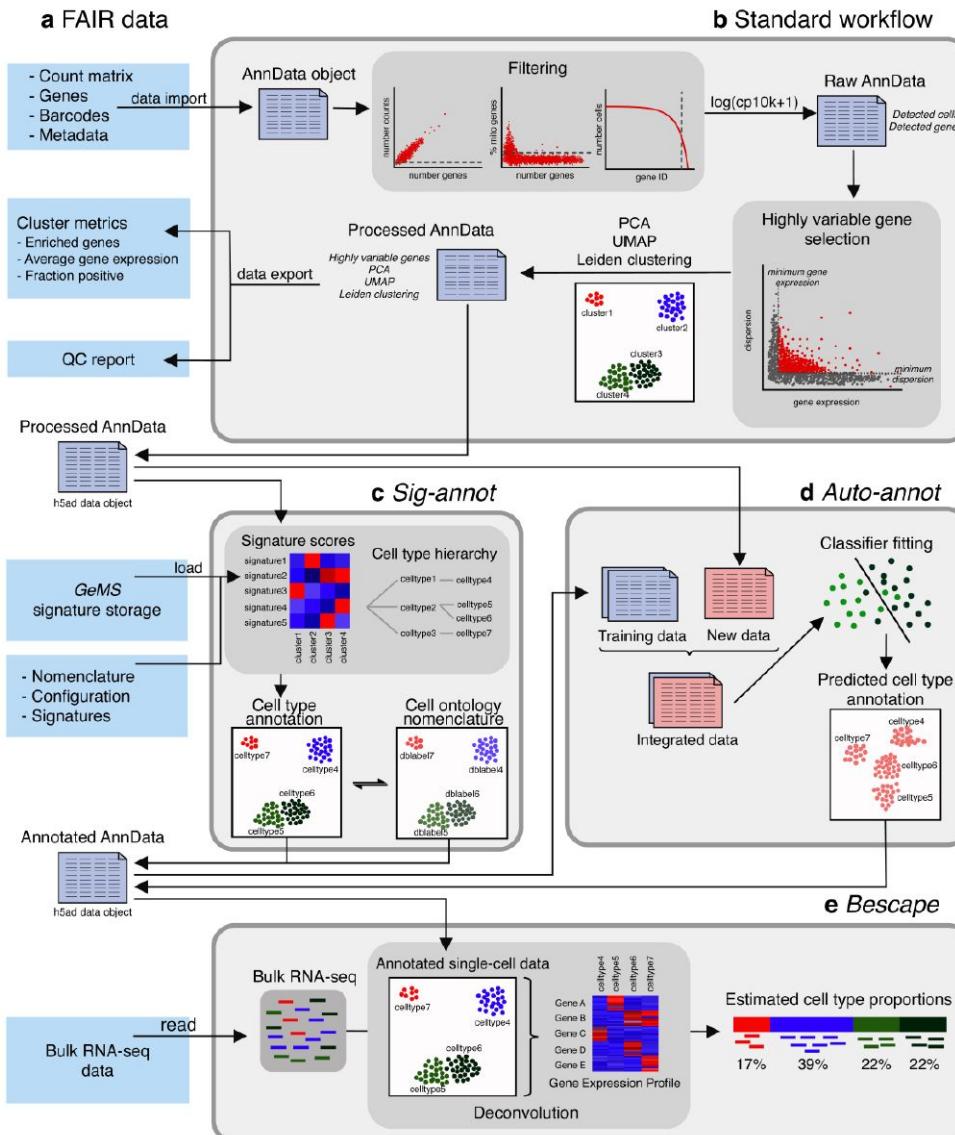
- (a) provides mechanistic validation of hits in successive screening campaigns,
- (b) enables undesirable and false-positive hits to be eliminated, and
- (c) brings biological relevance to screening assays by integrating patient information

Summary

- Gene expression profiling: a case study of omics and cellular modelling
- Applications for drug safety: TG-GATEs
- Applications for drug mechanism: molecular phenotyping
- Current research topics
 - Single-cell sequencing
 - Spatial-transcriptomics
 - Genome editing
 - Microbiome
 - High-content cellular imaging
 - Integrative modelling

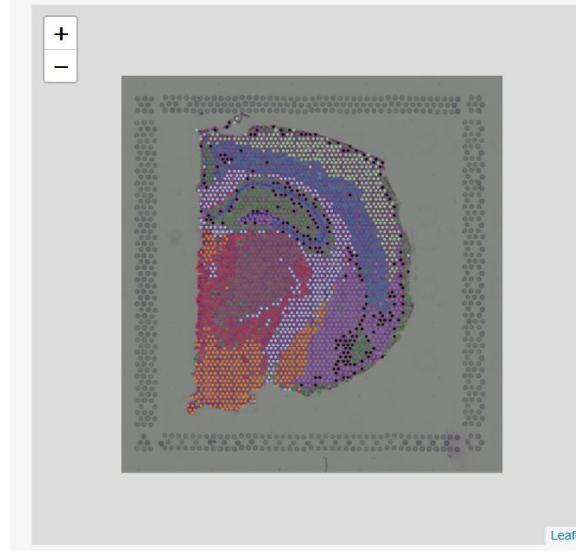


From single-cell analysis to spatial-transcriptomics



Visualization Controls

Use the sliders under the tissue image to adjust how you visualize and combine the tissue image and the gene expression data. Colors represent clusters identified by differentially expressed genes.



Gene Identification

By placing the pointer above a gene name within the table, spots in the tissue image will be colored based on the expression of that gene.

Alternatively, by placing the pointer above a value within the table, you can observe the expression of a specific gene with the spots from an individual cluster highlighted.

Cluster	1	2	3	4	5	6	7	8	9
Nptxr	3.5	1.1	2.2	4.7	1.8	4.1	1.6	1.2	3.1
Agt	0.59	2.0	3.6	1.2	0.78	0.42	3.2	2.6	0.61
Ttr	3.5	4.9	3.0	4.8	3.1	2.6	2.7	2.8	3.8
Pmch	0.77	1.8	4.5	1.5	0.56	0.79	1.3	0.63	0.66
Camk2n1	5.7	3.7	3.9	5.8	3.9	6.8	4.3	4.2	5.1
Olfm1	5.6	2.8	3.5	5.9	3.1	5.5	3.2	3.6	4.5
Pcp4	4.9	2.8	4.2	4.4	2.0	2.5	3.3	6.2	3.0
Prkcd	0.50	0.81	0.72	0.51	0.48	0.31	1.6	4.7	0.43
Cck	5.5	2.4	2.3	4.8	3.0	5.2	2.3	4.4	4.2
Nnat	2.2	2.7	5.2	3.8	1.6	1.2	3.0	1.8	2.4
Plp1	4.8	7.9	4.8	3.9	3.6	3.1	6.0	6.2	3.7
6330403K07Rik	3.5	2.0	5.3	3.6	1.4	3.3	3.4	1.5	2.2
Ctxn1	4.5	1.7	3.2	4.7	1.6	3.9	1.3	1.1	3.0
Atp1a1	4.2	2.7	3.1	4.2	1.8	4.9	2.3	2.0	3.0

Left:Mädler, Sophia Clara, Alice Julien-Laferrière, Luis Wyss, Miroslav Phan, Albert S. W. Kang, Eric Ulrich, Roland Schmucki, et al. 2020. "[Besca, a Single-Cell Transcriptomics Analysis Toolkit to Accelerate Translational Research.](#)" BioRxiv, September, 2020.08.11.245795.

Top: Spatial resolution of gene expression, which can be important for future digital pathology, source: <https://www.10xgenomics.com/spatial-transcriptomics/>

Summary and Q&A