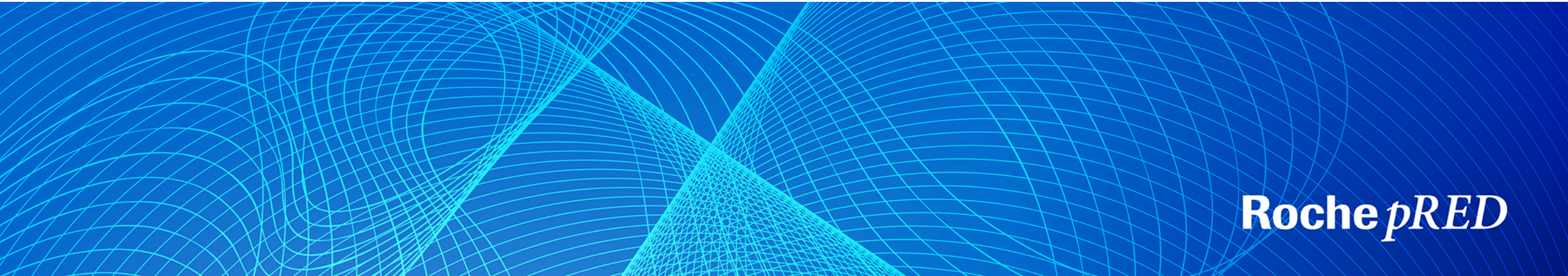
Mathematics In Drug Discovery: An Practitioner's View

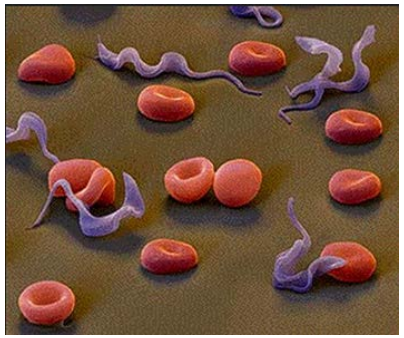
Dr. Jitao David Zhang, Bioinformatics and Computational Biology
Pharmaceutical Sciences, Pharma Research and Early Development, Roche Innovation Center Basel
Perlen-Kolloquium, University of Basel, October 25th, 2018



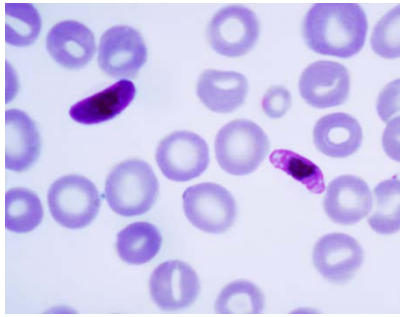
Contact the author

- Now is the best time in human history to fight diseases
- Mathematics approaches are indispensable to modern drug discovery
- Interdisciplinary mathematics will transform drug discovery in the coming decades

The history of *Homo sapiens* is a history of living with, understanding, and fighting diseases



Trypanosomes



Plasmodium

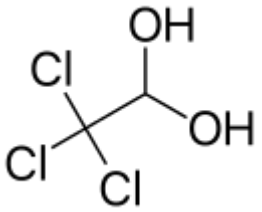
Tropical diseases

~500,000 years ago

A young patient of smallpox,
the first eradicated infectious disease

Hygiene, vaccination, and antibiotics

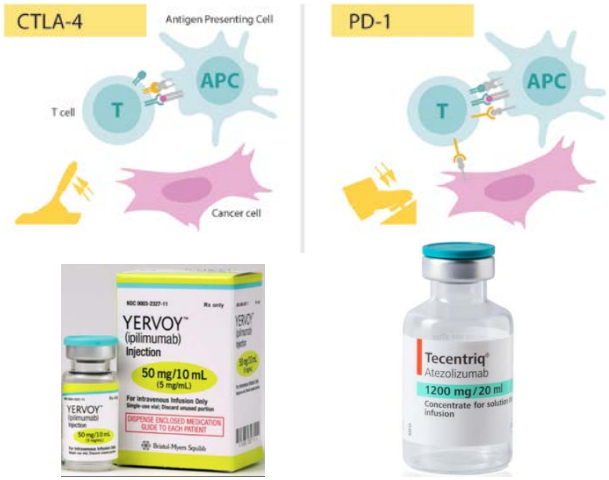
~250 years ago



Chloral hydrate,
the first synthesized drug

Pharmaceutical drugs

~150 years ago

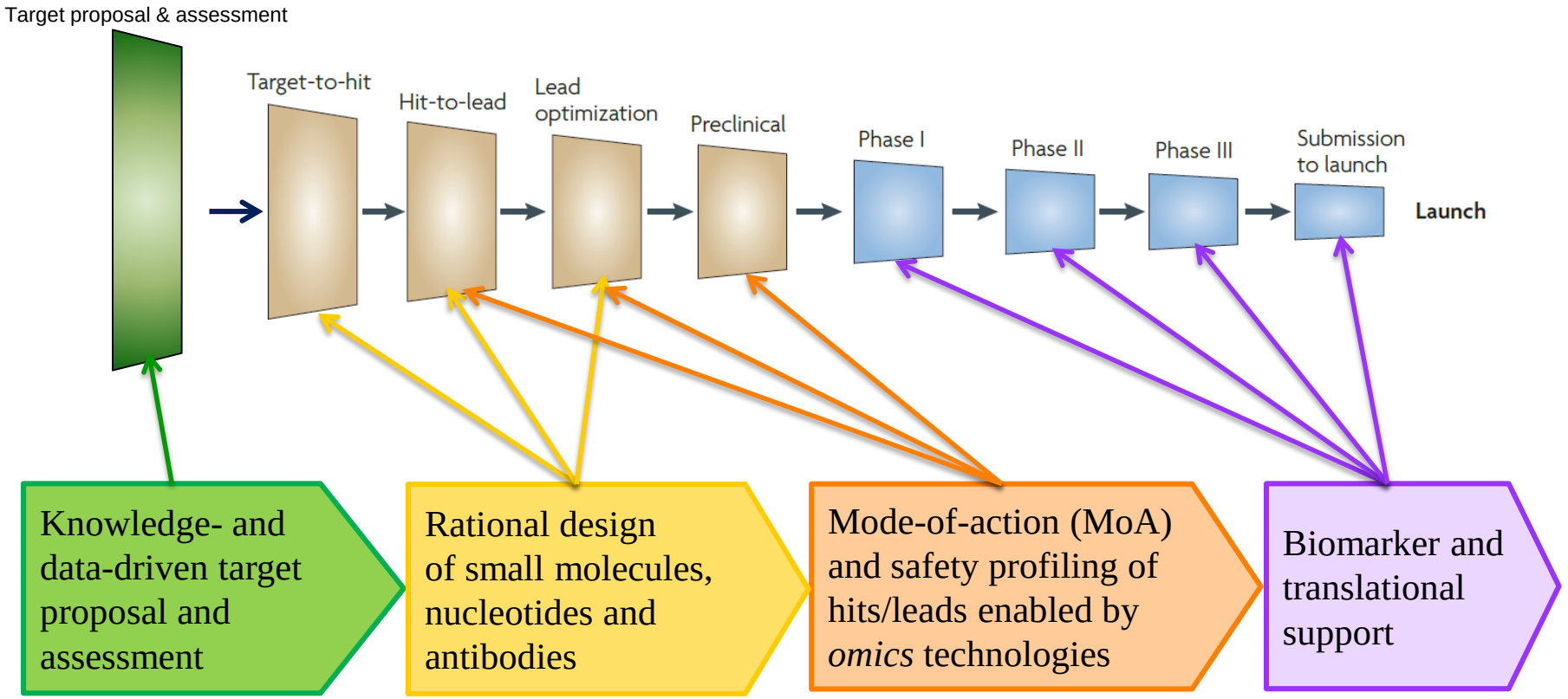


Nobel prize laureates 2018,
immune checkpoints,
and drugs targeting the pathways

Personalized precise healthcare

~20 years ago

Bioinformatics and computational biology, a branch of applied mathematics , are indispensable for modern drug discovery



Modified from Paul *et al.* "How to Improve R&D Productivity: The Pharmaceutical Industry's Grand Challenge." Nature Reviews Drug Discovery, 2010

Bioinformatics and computational biology have become indispensable for modern drug discovery

Applied mathematics empowers drug discovery by many ways

Applied mathematics *in drug discovery*
is not a definable scientific field
but a human attitude.

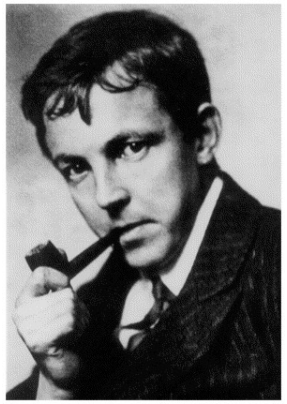
Richard Courant
(1888-1972)

Statistics, Data Mining and Machine Learning	Applied Combinatorics and Graph Theory	Stochastic Simulation	Geometric Modeling
Ordinary / Partial/ Stochastic Differential Equations	Network Analysis	Dynamical Systems	Molecular, Quantum, and Continuum Mechanics

Seemingly ‘pure’ mathematics significantly contributes to understanding biology, too

The mathematician’s patterns, like the painter’s or the poet’s, must be beautiful

Drug discoverer’s patterns, like the mathematician’s or painter’s or the poet’s, must be beautiful



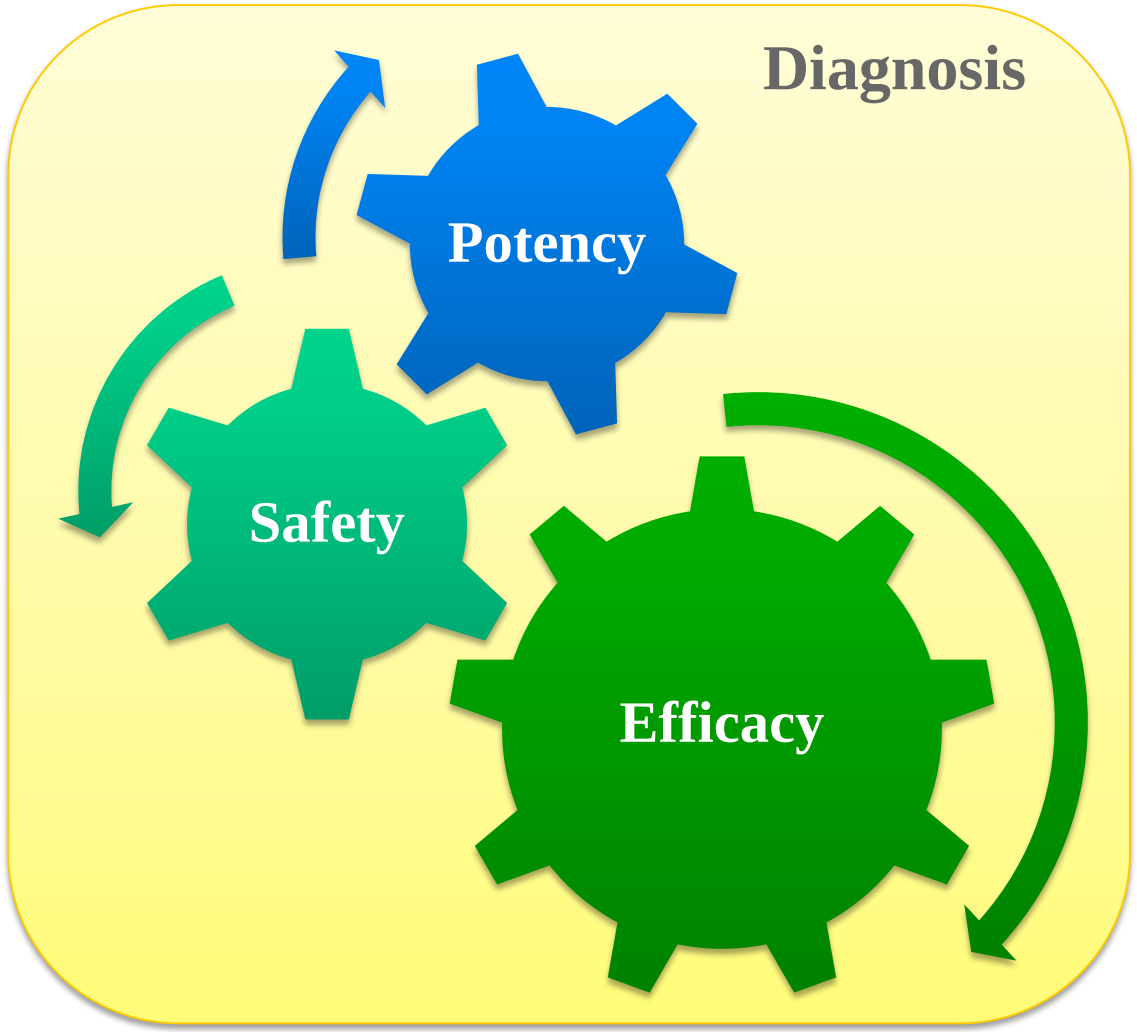
Godfrey Harold Hardy
(1877-1947)

It must be admitted that the biological examples … in the present paper are very limited. This can be ascribed … to the fact that biological phenomena are usually very complicated. … It is thought, however, that **the imaginary biological systems** … and **the principles …** should be of some help in interpreting real biological forms.

The chemical basis of morphogenesis, 1952

Alan Mathison Turing
(1912-1954)

Prerequisites to make a good drug that works



- **Potency**
- **Safety**
- **Efficacy**
- **Diagnosis** : doctors' judgement + *biomarkers*
  - *Biomarkers* are informative features derived from measurements of patient or patient material, *e.g.* blood chemistry, genetic make-up, imaging, *etc.*
- Other criteria: commercial rationale, development ability, intellectual property, *etc.*

Success in drug discovery is determined by potent, safe, efficacious drugs and accurate diagnosis

The essence & THE challenge of Drug Discovery

Constrained optimization and decision making based on incomplete, noisy and heterogeneous data, and limited prior knowledge.

Bioinformatics and computational biology in preclinical research contribute to making safe and efficacious drugs

Safety

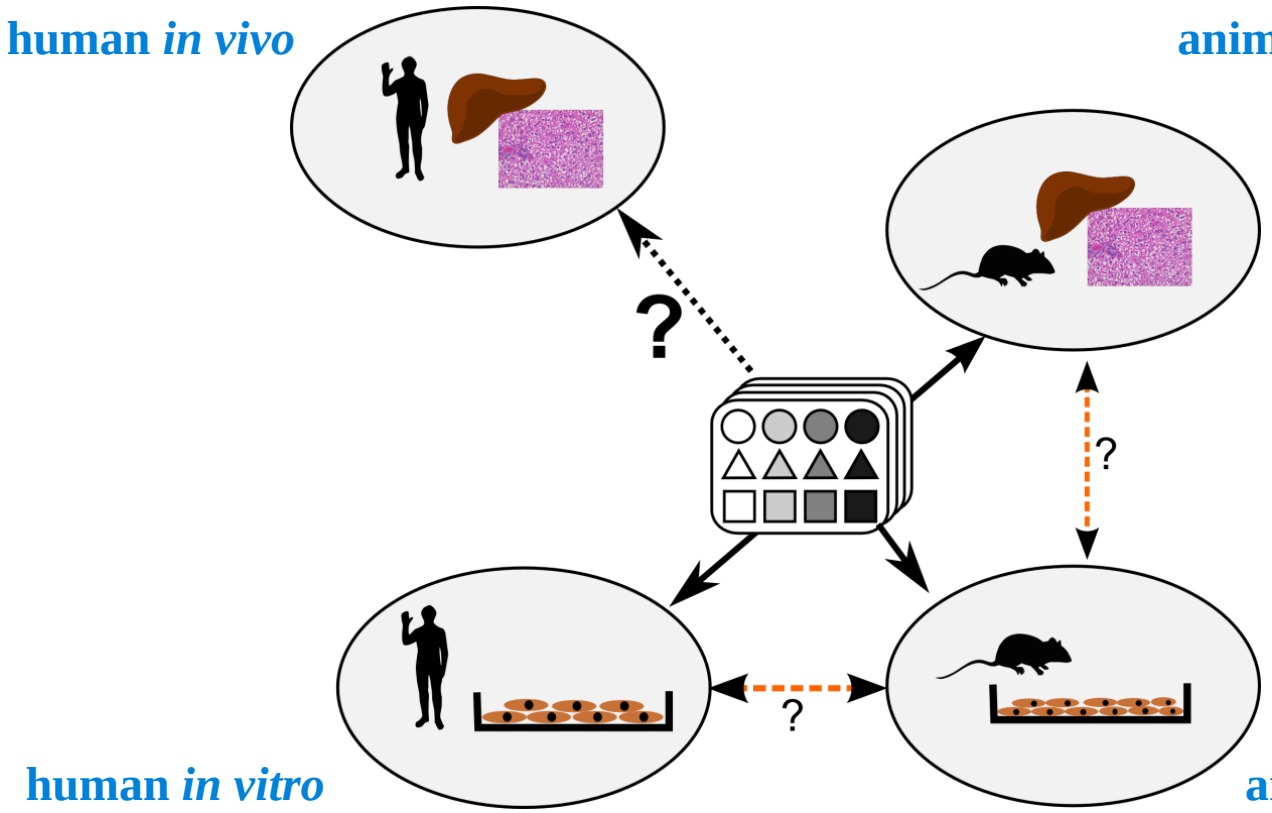
- Data mining reveals a network of early response genes as a consensus signature of drug-induced *in vitro* and *in vivo* toxicity. Zhang *et al.*, Journal of Pharmacogenomics, 2014.

Efficacy

- Molecular phenotyping combines molecular information, biological relevance, and patient data to improve productivity of early drug discovery. Drawnel and Zhang *et al*, Cell Chemical Biology, 2017.

We do research in computational biology at Roche and collaborate to make safe and efficacious drugs

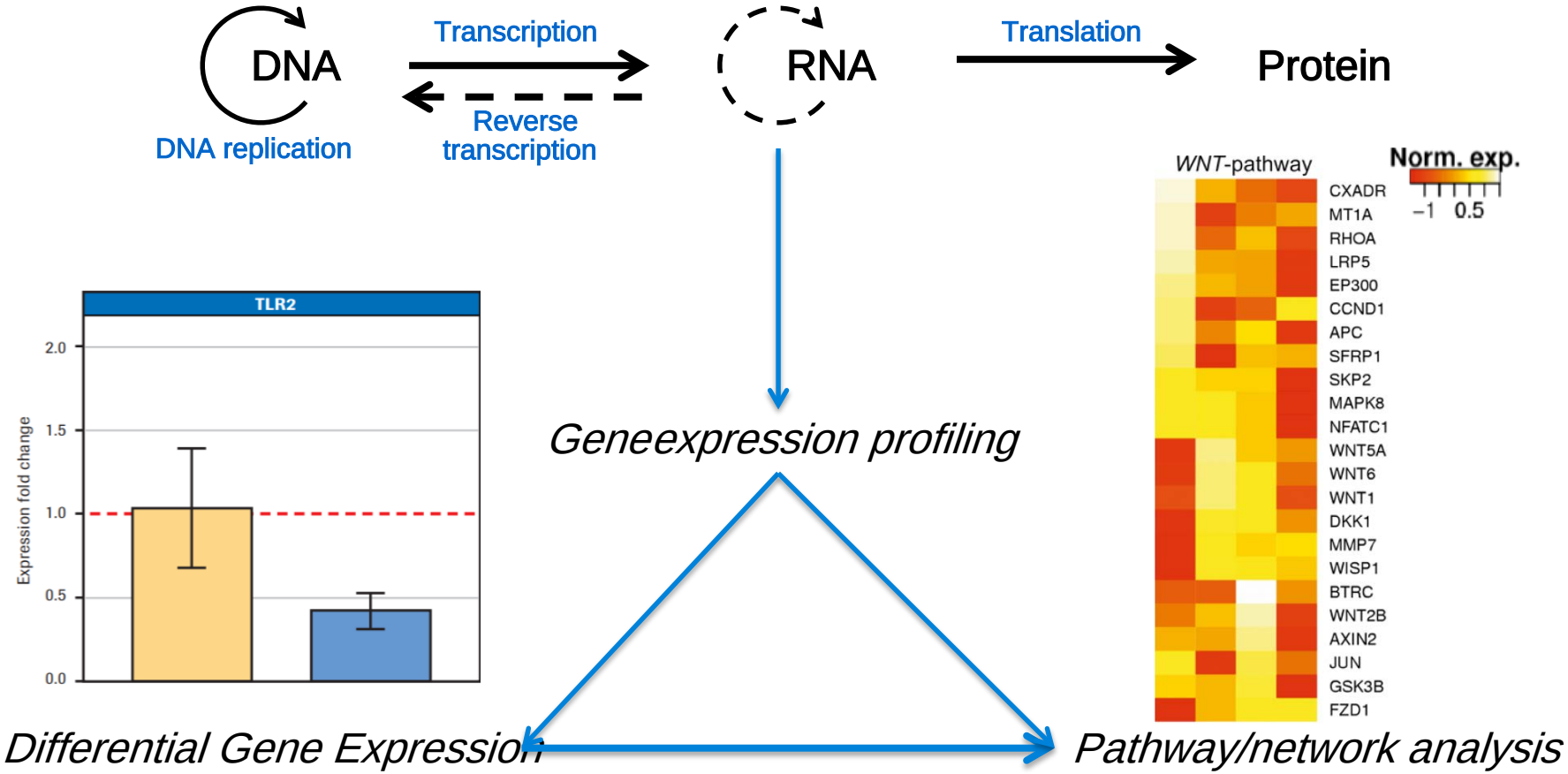
One challenge in drug discovery: non -clinical safety assessment



- 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

170
Compounds

> 2000
Cellular assays

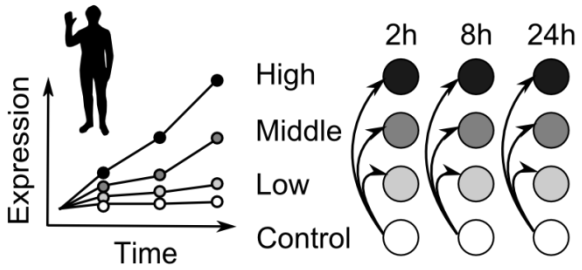
> 12000
Pathology records

> 24000
Expression profiles

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

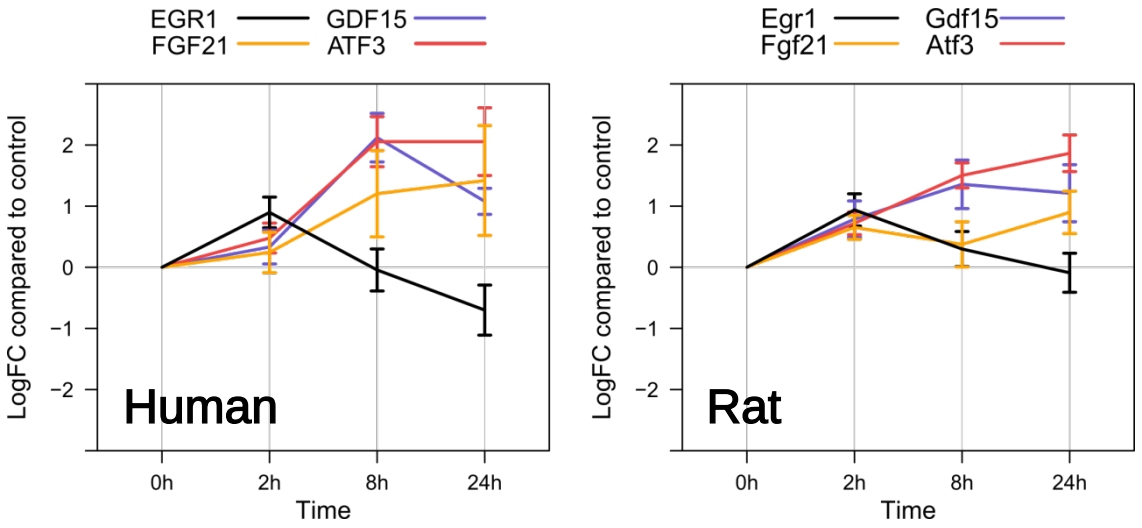
We built a computational pipeline to identify early signatures of toxicity

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

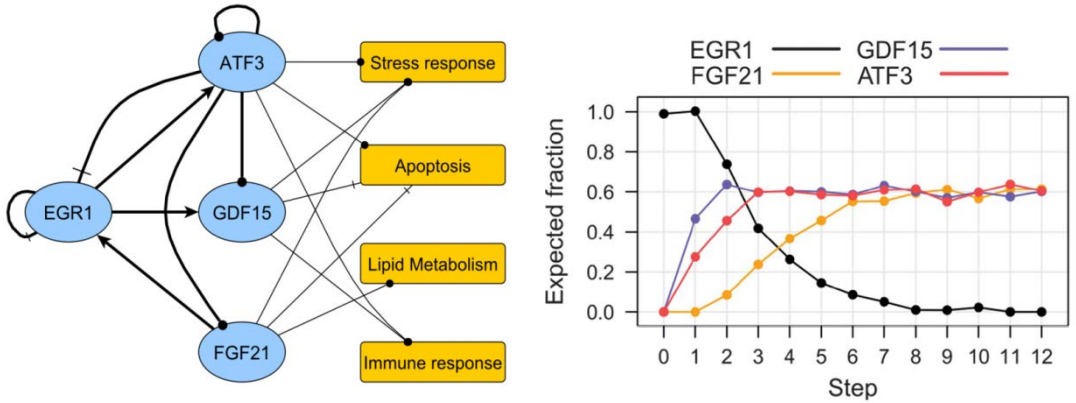


We integrate unsupervised learning, regression analysis, and network modelling to reach the goal

Conserved dynamics of the early signatures in human and rat primary hepatocytes is intrinsic to the network structure



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

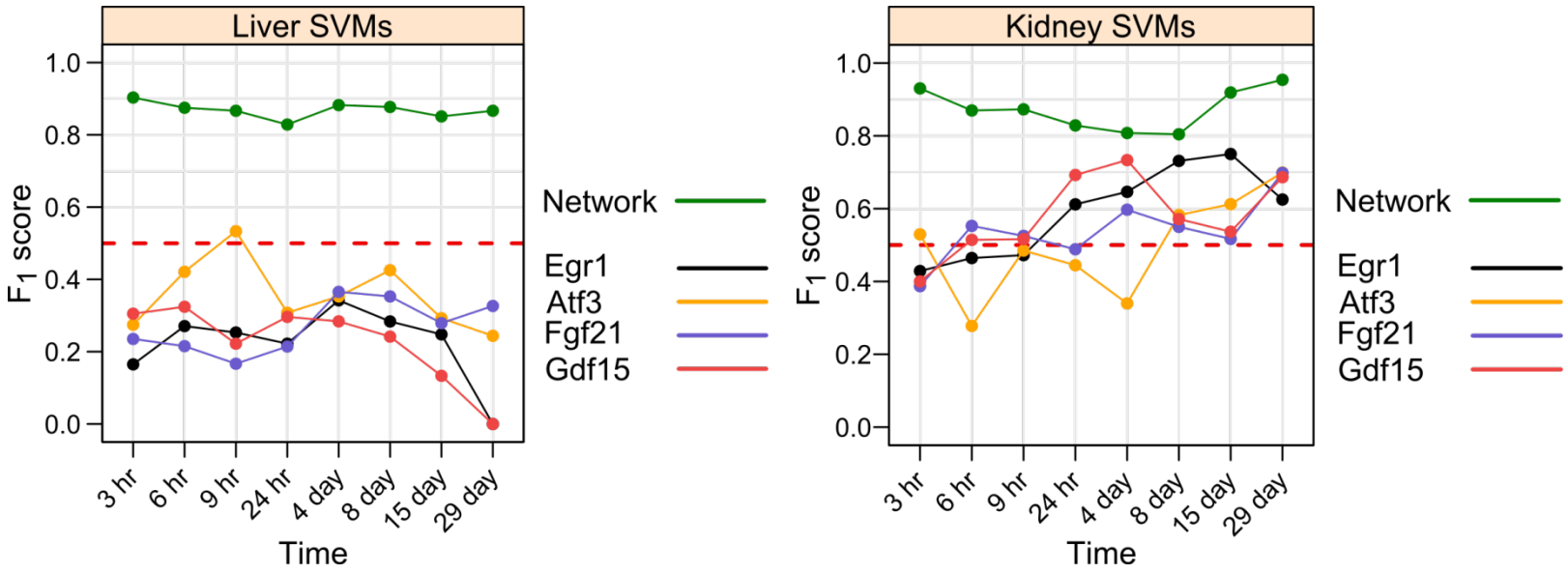


- The network structure was constructed by queries in interaction database and literature information.
- Boolean network simulation (Nikolaos Berntenis and Martin Ebeling, BMC Bioinformatics 2013) suggests that **the conserved dynamics of the network in human and rat is encoded in the conserved structure of the network**.

Integrated data analysis reveals an evolutionarily conserved network with intrinsic dynamics that responds early to drug-induced toxicity

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

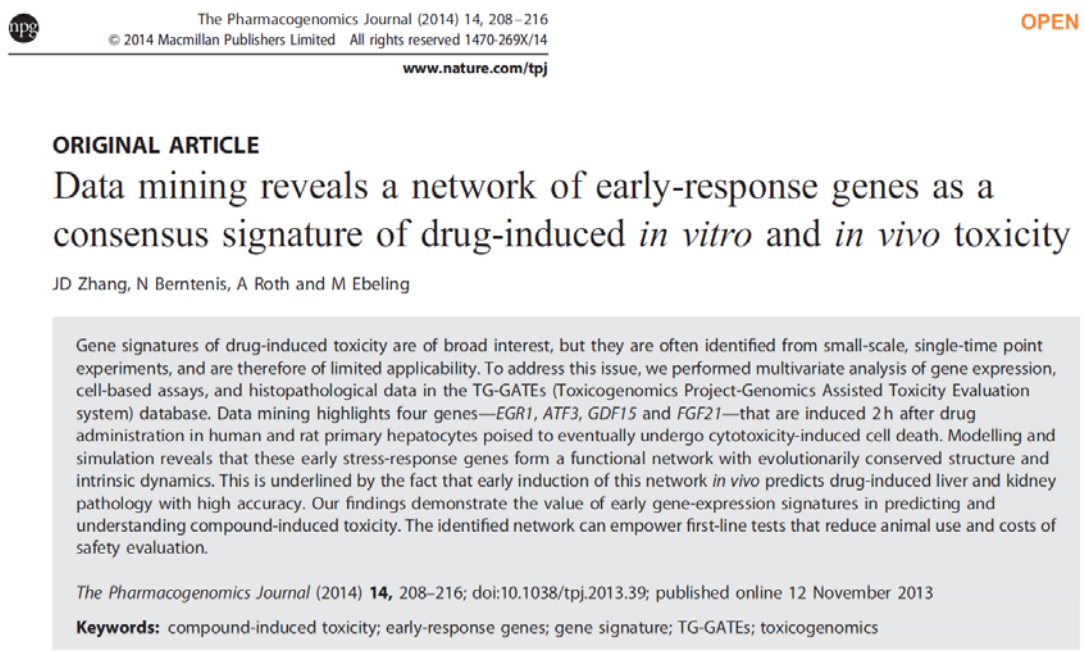
- Support Vector Machines (SVMs) were trained to predict*in vivo* pathology between 3h and 29d using gene expression changes of Egr1, Atf3, Gdf15, and Fgf21 at 3h.
- Profiles were randomly split into training samples (80%) and test samples (20%).
- SVMs are trained by 10-fold cross-validation in training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

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 Berntenis, 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.

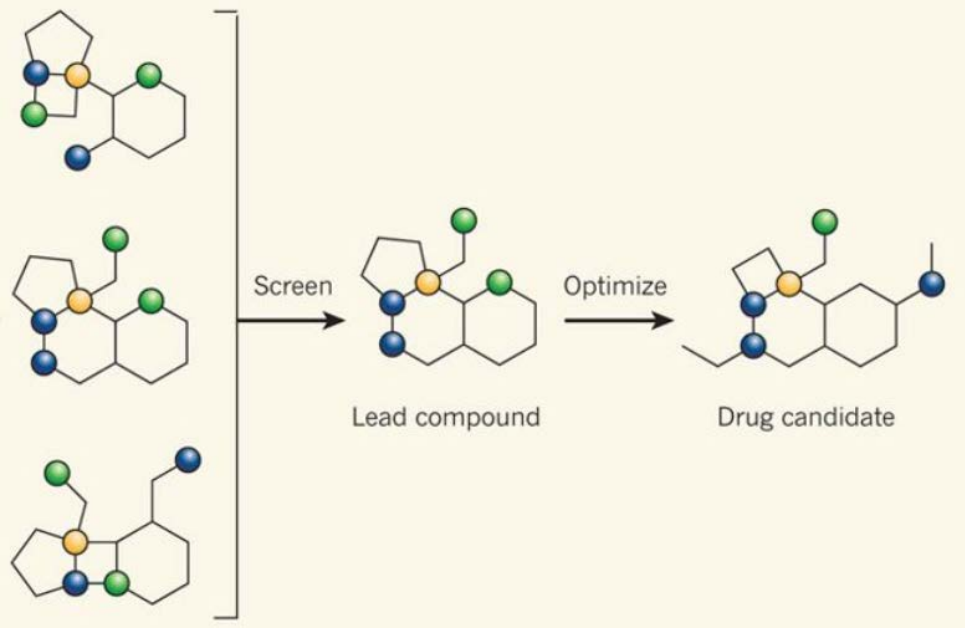
The Pharmacogenomics Journal (2014) **14,** 208–216; doi:10.1038/tpj.2013.39; published online 12 November 2013

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

Zhang *et al.*, J Pharmacogenomics, 2014

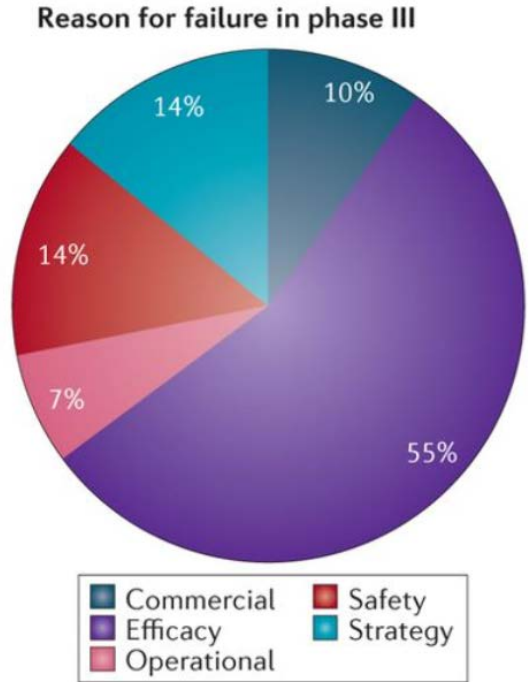
Computational biology and bioinformatics help identifying safer drugs

Another challenge in drug discovery: lack of tools to examine effects of compounds in human cells, tissues and organs



Hajduk *et al*, Nature, 2011

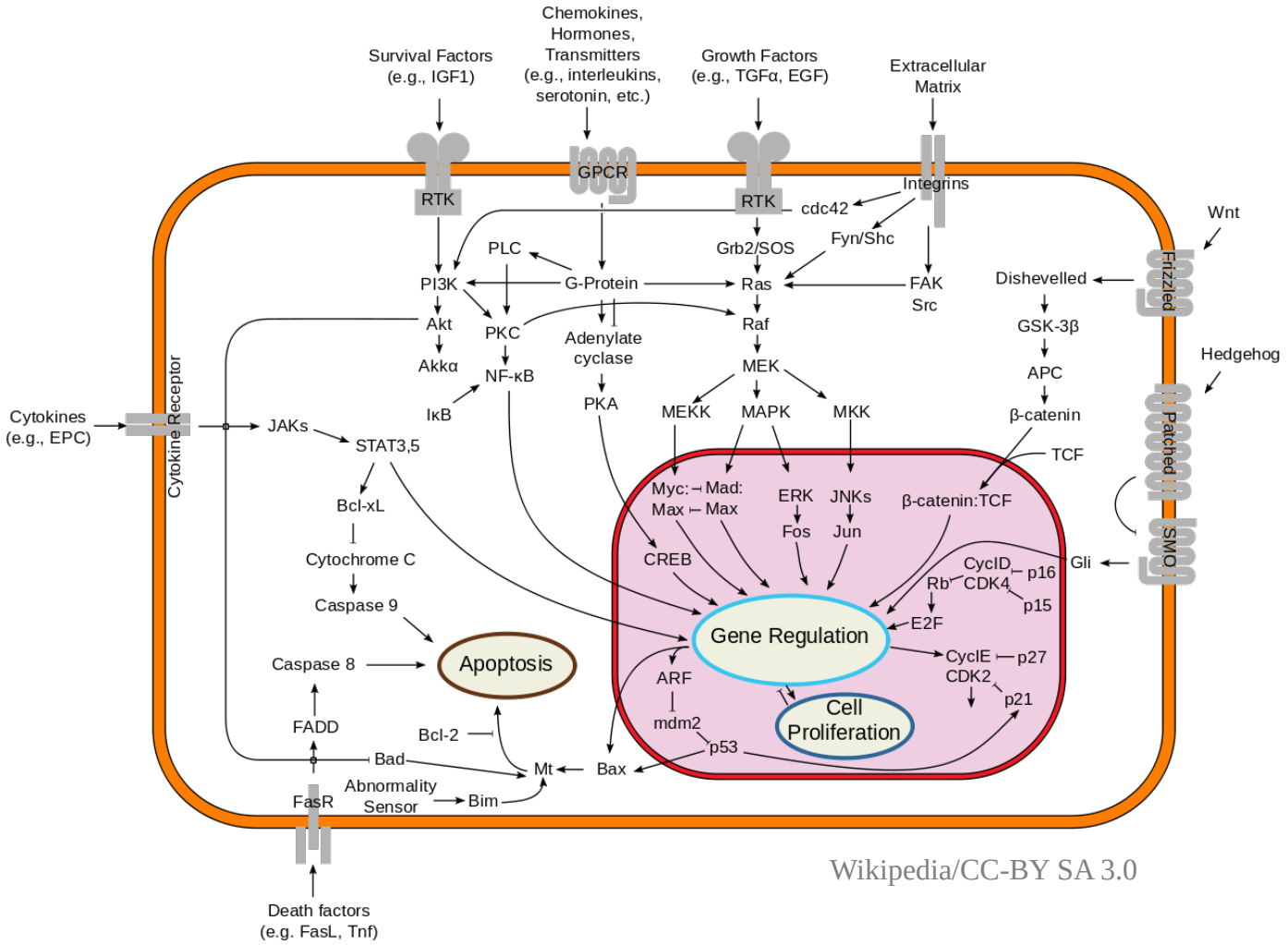
- Chemical tractability
- Potency
- Specificity
- Response in human cells?



Harrison, Nature Reviews Drug Discovery, 2016

Drug discovery may benefit from early assessment of pathway -level responses to drug candidates

What is a biological pathway, and why we care about it?

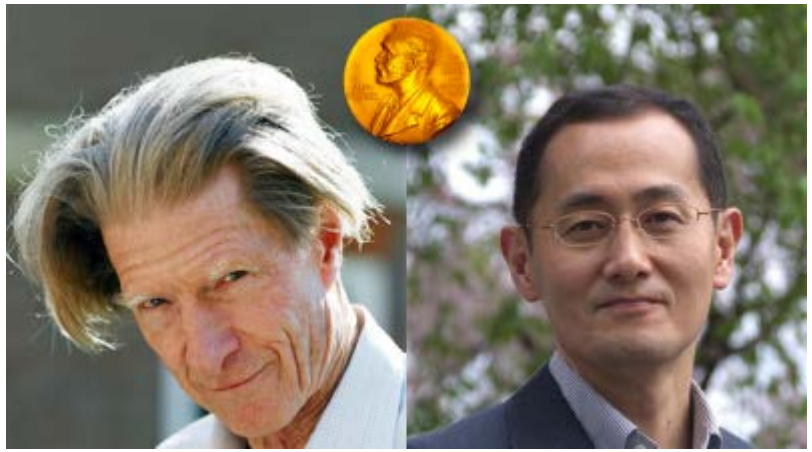


Wikipedia/CC-BY SA 3.0

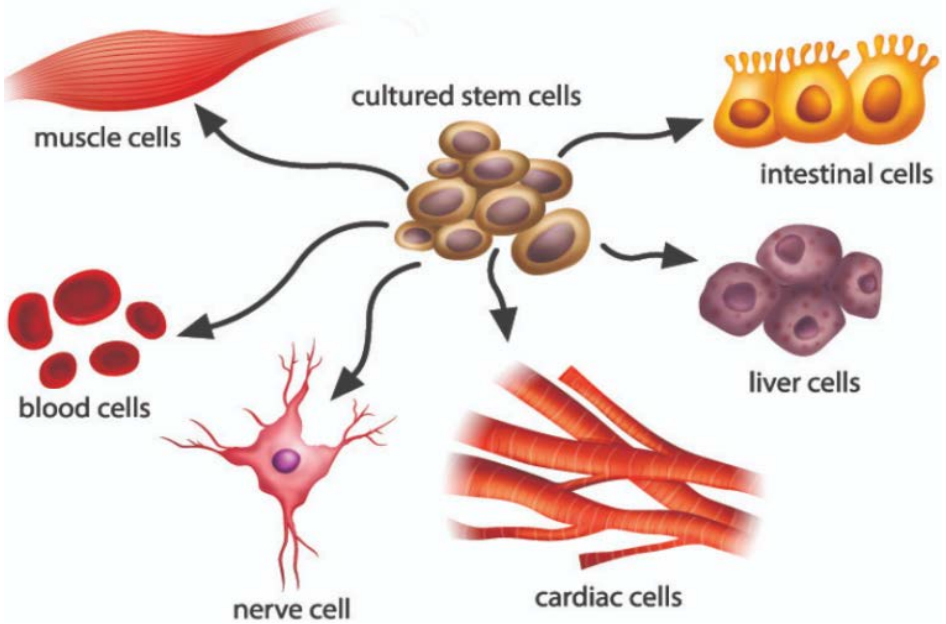
- A biological pathway is a series of actions among molecules in a cell that leads to a certain product or a change in the cell
- A pathway can trigger the assembly of new molecules, such as a fat or protein. Pathways can also turn genes on and off, or spur a cell to move
- Biological functions can be considered as sums of outputs of biological pathways.
- Pathway activation and inactivation leaves *fingerprints*, *i.e.* specific changes, in gene (mRNA) expression profiles of the cells. These fingerprints are sometimes called *gene signatures*

It is possible to infer the status of biological pathways by gene expression profiling

Stem cells, a revolutionary tool for drug discovery



John B. Gurdon and Shinya Yamanaka, Nobel Laureates

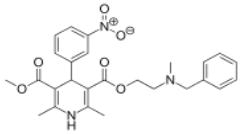


Stem cells

Stem-cell technology empowers *molecular phenotyping* that reveals cell-specific pathway responses of compounds

Molecular Phenotyping

Small molecule



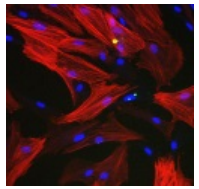
Antibodies

Antisense oligos

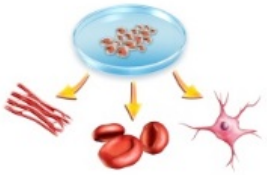
Therapeutic candidates

Early time point (3-12h)

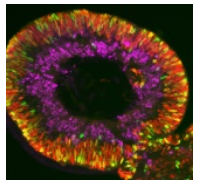
Human *in vitro* models



Cell lines/
primary cells

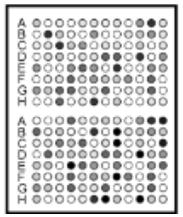


iPS-derived cells
(opt. genome editing)



Advanced
models

~ 1000 pathway reporter genes



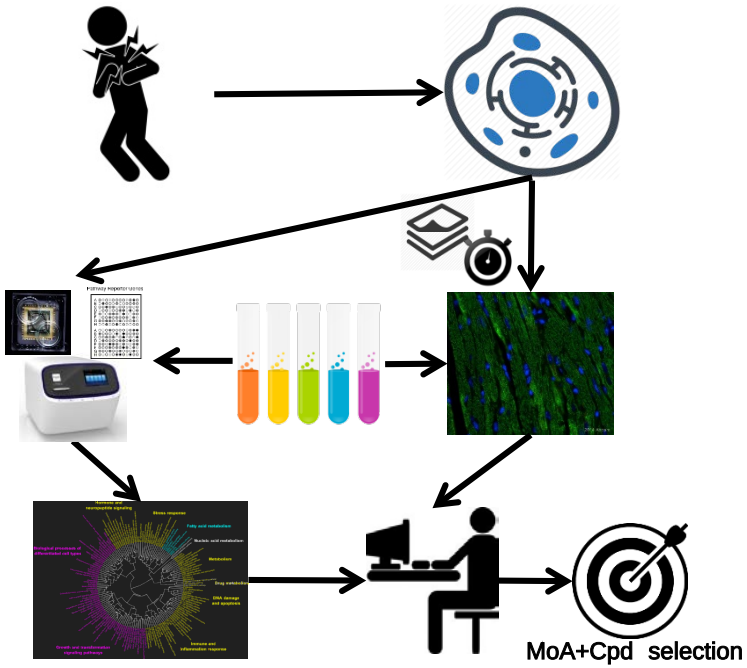
Next-generation sequencing

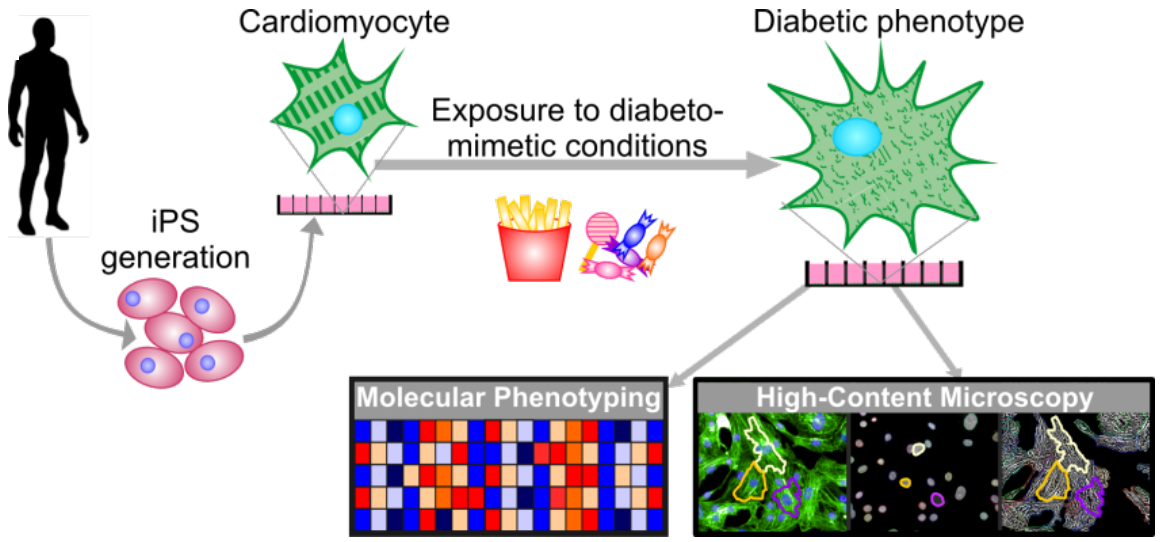
Molecular phenotyping

What pathways are perturbed in what cells by each compound?

Zhang *et al*, BMC Genomics, 2014
Zhang *et al.*, BMC Genomics, 2015

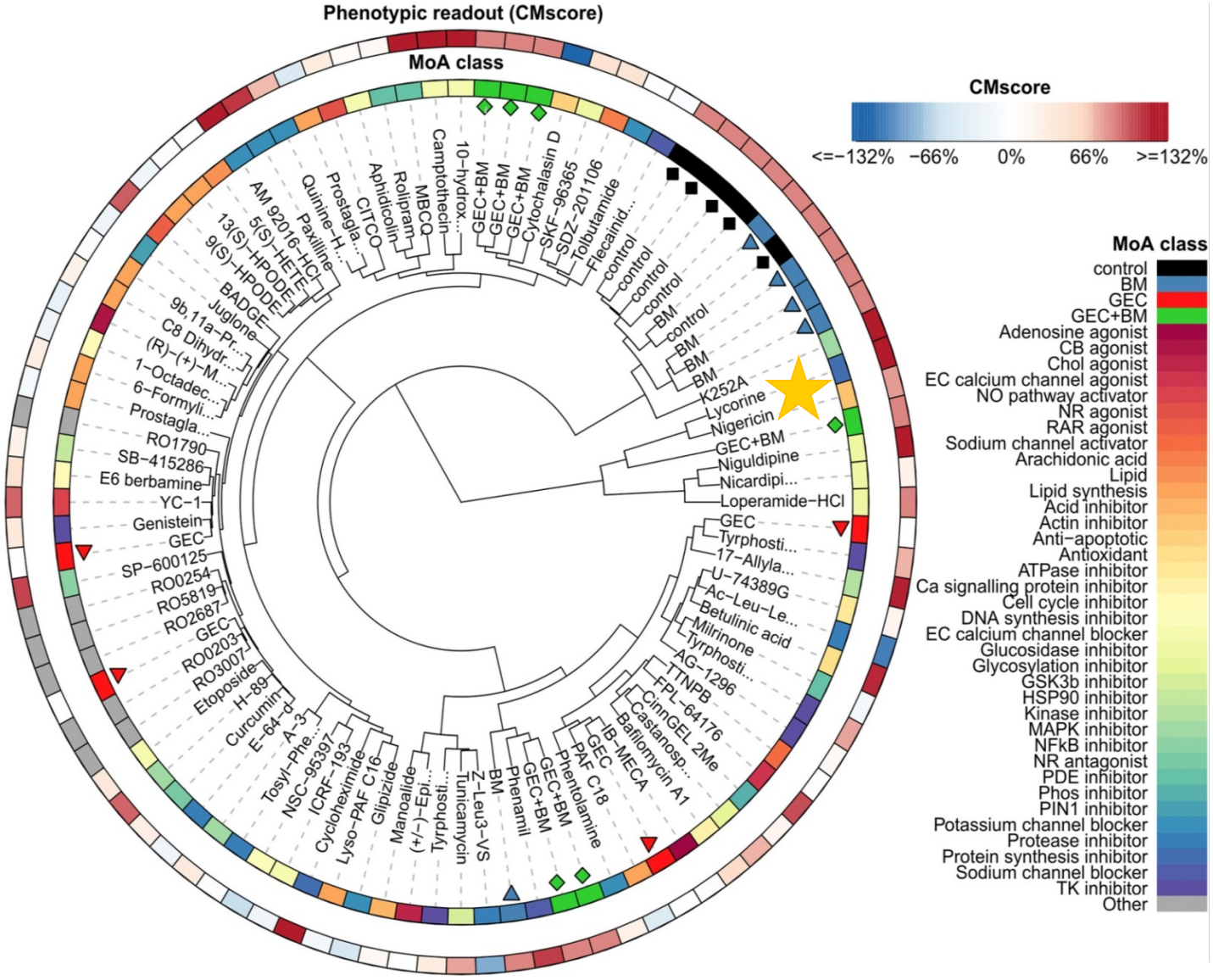
From phenotypic drug discovery to molecular -phenotypic drug discovery



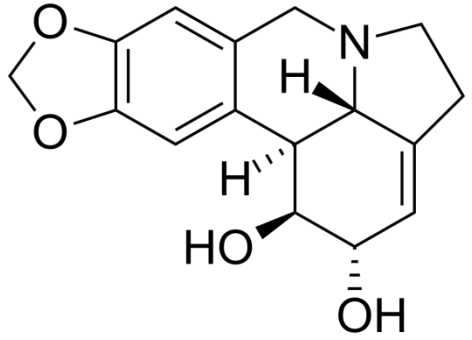


Molecular phenotyping reveals pathway modulation patterns of drug candidates that may inform candidate selection

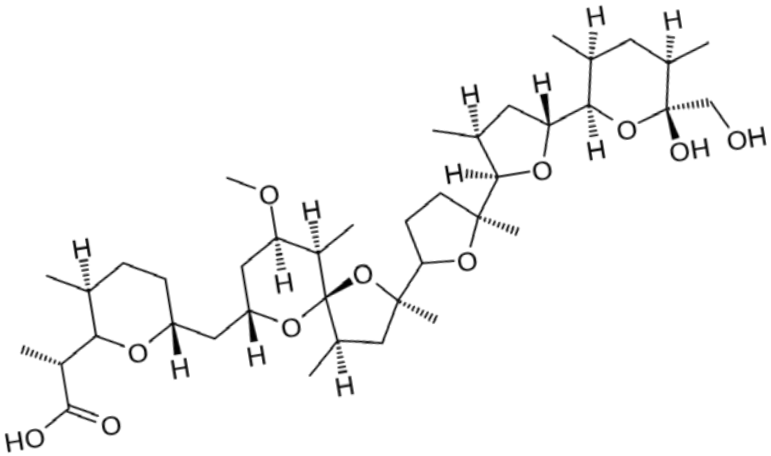
Integration of molecular, chemical, and phenotypic information



Lycorine: Protein synthesis inhibitor

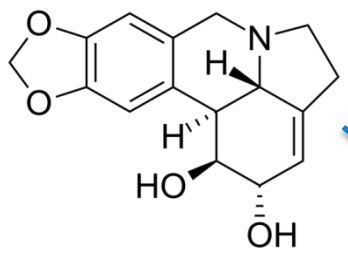


Nigericin : Potassiumionophore

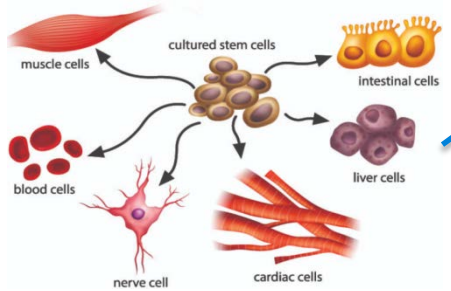


Molecular phenotyping clusters compounds based on pathway modulation profiles beyond phenotype or structure

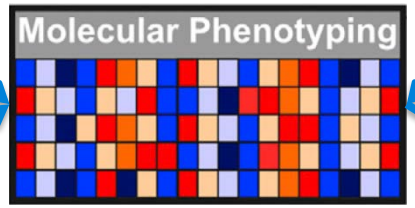
Molecular phenotyping links drug -induced pathway modulation with patient data to prioritise likely efficacious compounds



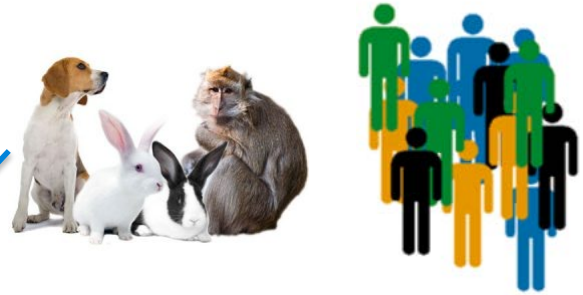
Compounds



Stem cells



Validate



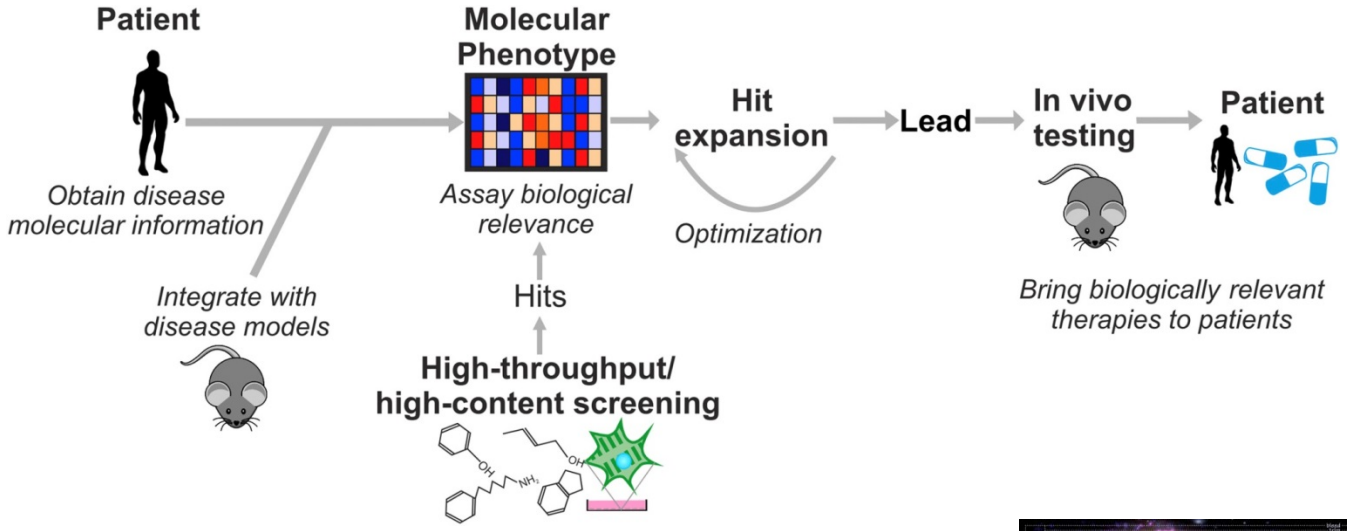
Data from animal models and patients

Efficacious drugs for the correct patient group

Molecular phenotyping reveals compounds with desired pathway profiles that are relevant for patients

Summary of the molecular phenotyping pilot study

- Molecular phenotyping characterizes and clusters compounds by pathway modulation patterns besides structure or induced phenotype.
- Molecular phenotyping brings biological relevance to screening assays by integrating patient information.
- MP is relevant for both target based and phenotypic drug discovery (Moffat *et al.*, Nature Reviews Drug Discovery, 2017; Comess *et al.*, J Med Chem, 2018).



Resource

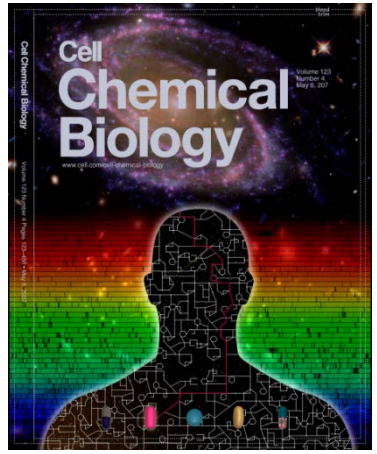
Cell Chemical Biology

Molecular Phenotyping Combines Molecular Information, Biological Relevance, and Patient Data to Improve Productivity of Early Drug Discovery

Graphical Abstract

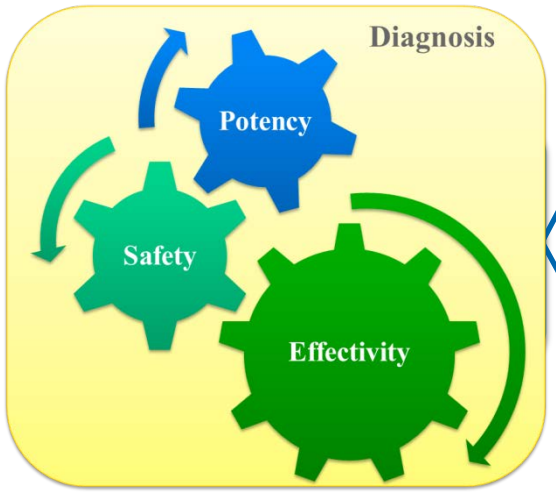
Authors

Faye Marie Drawnel, Jitao David Zhang, Erich Küng, ..., Thomas Singer, Martin Ebeling, Marco Prunotto



Molecular phenotyping integrates molecular information, biological relevance, and patient data for drug discovery

Bioinformatics and computation biology support disease understanding and drug development in many ways



Disease understanding

Drug development

- **Analysis and modelling of data from patients or disease models**
- **Biological network analysis**
- **Translational biomarker identification**
- **Molecular phenotypic drug discovery**
- ***Omics* data mining, integration, and modelling**

Selected publications

- Gatti *et al*, bioRxiv, 2017
- Grabole *et al*, Genome Med, 2016
- Van Vries *et al*, J Infect Dis, 2015

- Moisan *et al*, Nature Cell Biology, 2014
- Uhlmann *et al*, Mol Syst Biol, 2012
- Zhang & Wiemann, Bioinformatics, 2008

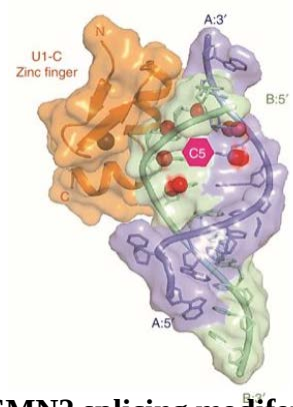
- Mueller *et al*, J Hepatol, 2017
- Haller *et al*, Int. J. Cancer, 2015
- Adam *et al*, Urol Onco, 2013

- Drawnel *et al*, Cell Chem Biol, 2017
- Moisan *et al*, Mol Ther Nuc Acids, 2017
- Zhang *et al*, BMC Genomics, 2015

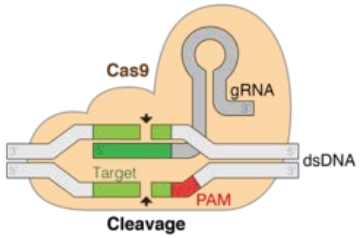
- Zhang *et al*, BMC Genomics, 2017
- Xu *et al*, Cancer Cells, 2015
- Zhang *et al*, J Pharmacogenomics, 2014

Bioinformatics and computational biology are integral to disease understanding and drug development

Now is the best time in human history to fight diseases



SMN2 splicing modifer
Naryshkin *et al.*, Science 2014; Sivaramakrishnan & McCarthy *et al.*, Nat Comm, 2017



CRISPR-CAS9 gene editing system

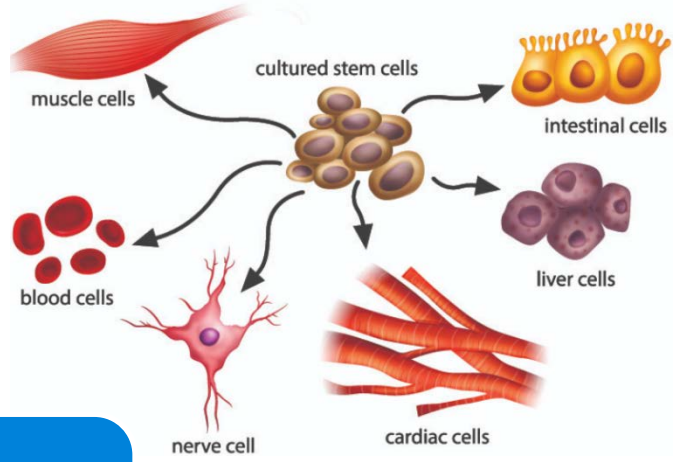
More biological, chemical, and medicinal knowledge

New disease-modelling systems

Digitalization of molecular mechanisms in living organisms

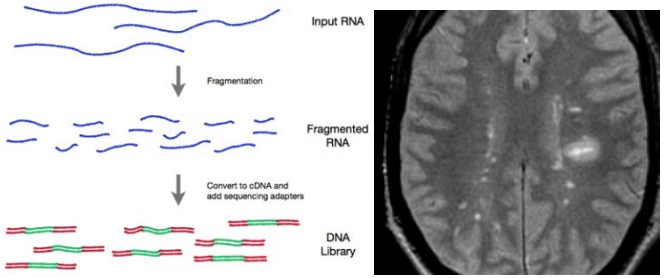
Better algorithms, models, and more computing resources

New therapeutic modalities



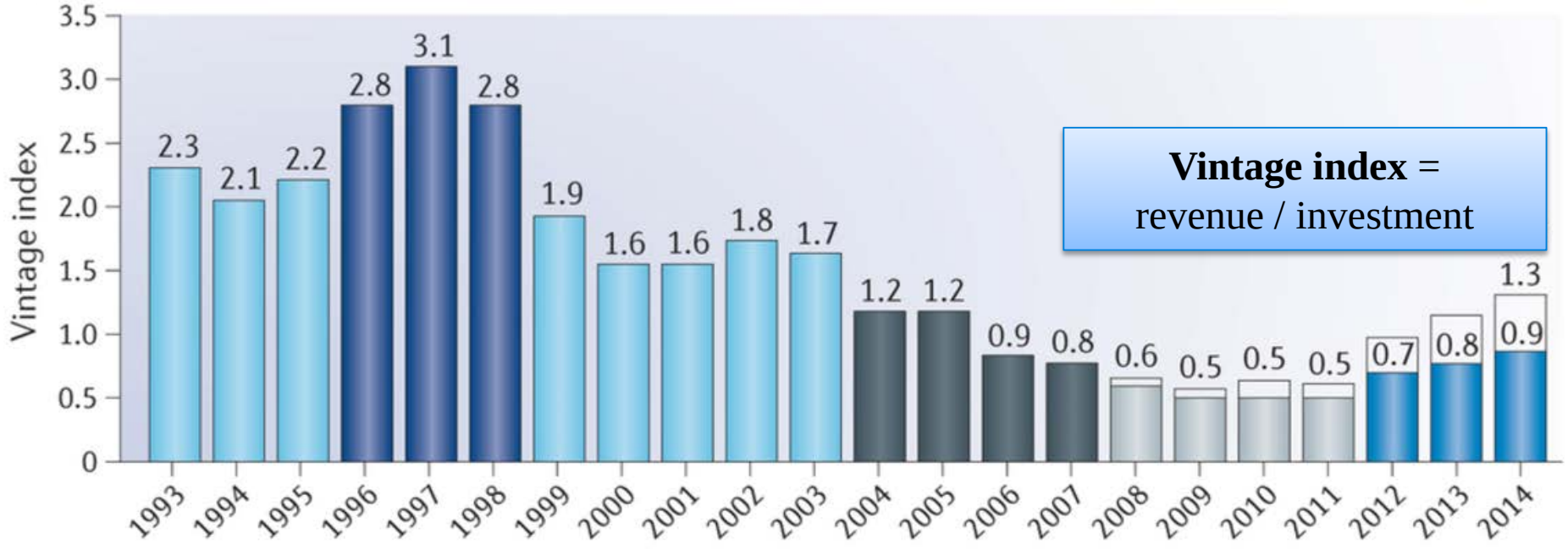
Stem cells

Comprehensive Sensing



Gene expression profiling and imaging

Increasing cost and decreasing return of investment in drug discovery



Modified from Smietana *et al.* "Improving R&D Productivity." Nature Reviews Drug Discovery, 2015

Finding new drugs has become more challenging and expensive

危机

~~*n.* crisis~~

Danger + Opportunity

Selected potential breakthroughs in coming decades

Potency

- Quantum computing for structure search and drug design: scalable algorithms and software is missing

Safety

- Lack of predictive models for some applications, and suboptimal translatability of current *in silico* and *in vitro* models

Efficacy

- Lack of personalised prescription and dosing for most drugs

Drug discovery as a process

- Quantification of information robustness
- Rationalization of decision making

Disease Biology

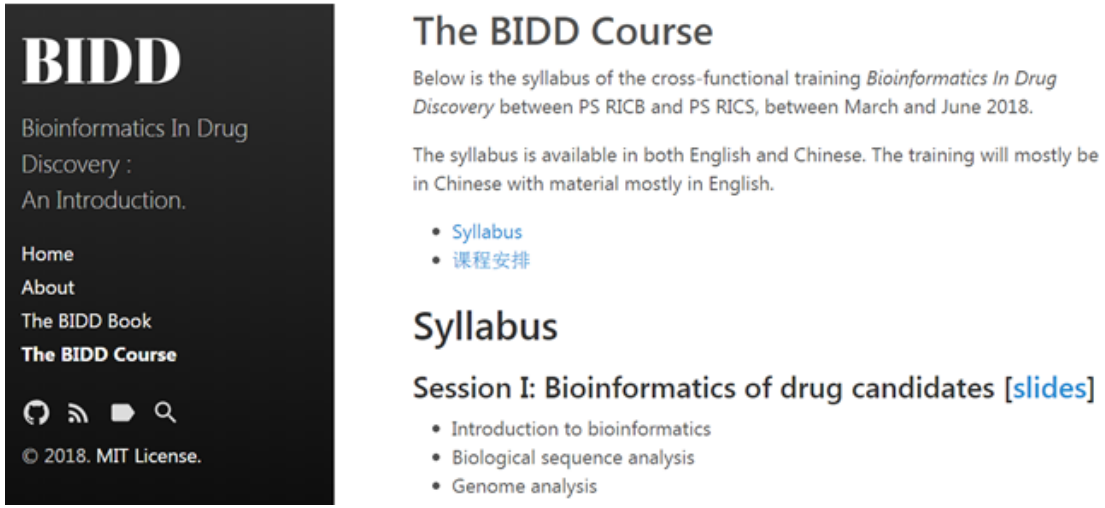
- How individual cells communicate, collaborate, and regulating each other to achieve homeostasis, and how to regain homeostasis if it is lost
- How genome sequence, structure, and variants orchestrate to function
- Identify targets for diseases for which samples are difficult to get, *e.g.* Alzheimer and Parkinson's Disease
- How to effectively identify and explore the druggable subset of the chemical space *in silico*, with new *in vitro* or *ex vivo* systems, and synthetic biology

Interdisciplinary research, including mathematics and informatics, is called to tackle these challenges

The course series *Bioinformatics In Drug Discovery* populated concepts and principles among Roche colleagues

- **Background:** Between April and August 2018, my colleagues and I co-organised the cross-functional training course *Bioinformatics In Drug Discovery* (BIDD).
  - The audience was ~20 scientists of mixed background (pharmacologists, toxicologists, biologists, etc.)
  - The course was given in six remote sessions à 90 minutes.
- **Topics covered:**
  1. Bioinformatics of drug targets and drug candidates
  2. Gene expression and regulation
  3. Functional analysis based on sequencing and gene expression
  4. Bioinformatics of Proteomics & Metabolomics
  5. Bioinformatics of Genomics & Genetics
  6. Statistics, machine learning, and data integration
- **Lessons learned:** It is both challenging and rewarding to bring bioinformatics and computational biology thinking to other scientists. As a team, we are convinced that the gain of productivity will pay out.

The BIDD-course website



Screenshot of the course website

New rounds of the BIDD course and other relevant training programs are being scheduled. Material will be re-used, and enhancement is planned.

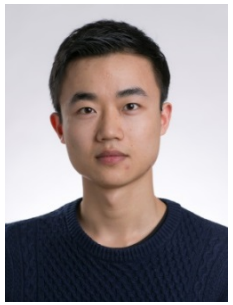
We encourage students and young professionals of mathematics and informatics joining us!

Gregor Sturm, LMU/TUM Bioinformatics

- BioQC algorithm and software (Zhang *et al*, BMC Genomics, 2016)
- A prevalence study of tissue heterogeneity in gene-expression data (Sturm & Zhang, in preparation)

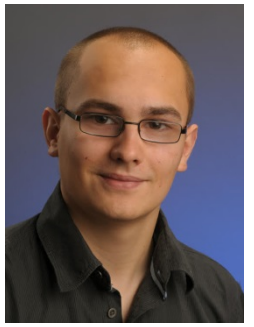
Dr. Simon Gutbier, Roche

Post-doc project *IMmune Pathway Characterizaton with Tool Compound Screening (IMPACTS)* to identify druggable pathways in microglia carrying a genetic risk factor of Parkinson's Disease using stem-cell technology, compound screening, and molecular phenotyping. Co-supervisiors: C. Patsch, M. Britschgi (Roche), and S. Cowley (U Oxford)



Tao Fang, ETH Comp Biol. & Bioinformatics

- Prediction of drug-induced liver and kidney toxicity in rat using gene-expression and drug-target-interaction data and deep neural networks (master thesis, publication in preparation). Co-supervisor: M. D. Robinson (UZH)
- A novel approach for competitive gene-set enrichment analysis (publication in preparation)



Rudolf Biczok, KIT Informatics

- A database system for differential gene expression analysis, and its application in pathway/network inference (master thesis). Co-supervisor: A. Stamatakis (KIT/HITS)
- Data mining, modelling, and integration for the IMPACTS PostDoc project (see above)

We offer internship, master, and post-doc positions for young scientists interested in drug discovery

Conclusions and perspectives

- Best time in human history to join the fight against diseases.
- The central challenge of drug discovery is constrained optimization and decision making based on incomplete, noisy and heterogeneous data, and limited prior knowledge .
- Interdisciplinary research , especially applying mathematical approaches and tools to biological, chemical and medicinal questions, is imperative to fill the knowledge gaps and to make potent, safe, and efficacious drugs and to perform accurate diagnosis.
- Mathematics and informatics will continue transforming drug discovery
  - From correlation to causation
  - From qualitative description to quantitative prediction
  - From trial-and-error to systematic understanding
  - From population inference to individual prediction and continuous intervention
  - From observations to engineering and synthesis of the biological system
- I argue for a unified framework of research, training and mentoring for bioinformatics and computational biology, as a branch of applied mathematics, for drug discovery in the coming decades

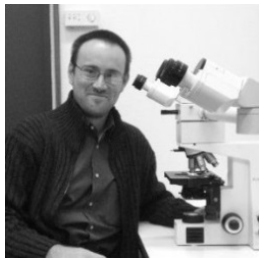
Acknowledgements



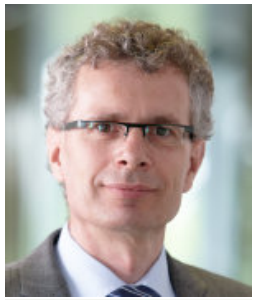


F. Hoffmann-La Roche Ltd	
Clemens Broger †	**Faye Drawnel**
Martin Ebeling	Markus Britschgi
Manfred Kansy	Roland Schmucki
Fabian Birzele	Martin Stahl
Kurt Amrein	Isabelle Wells
Annie Moisan	**Lu Gao**
Luca Piali	Lue Dai
John Young	Ravi Jagasia
Lisa Sach-Peltason	**Marco Prunotto**
Mark Burcin	**John Moffat**
Christoph Patsch	Gang Mu
Michael Reutlinger	Jianxun Jack Xie
Matthias Nettekoven	Filip Roudnicky
Andreas Dieckmann	Isabelle Wells
Klas Hatje	Iakov Davydov
Laura Badi	**Ulrich Certa**
Tony Kam-Thong	Detlef Wolf
Corinne Solier	Ken Wang
Thomas Singer	Nikolaos Berntenis





External to Roche
Stefan Wiemann
Wolfgang Huber
Ozgür Sahin
Agnes Hovrat
Katharina Zweig
Sally Cowley
Alexandros Stamatakis
Michael Prummer
Mark D. Robinson
Michael Hennig
Florian Haller
Jung Kyu Canci
Verdon Taylor
Maria Anisimova
Lorenzo Gatti
Erhard van der Vries
Ab Osterhaus
Nevan Krogan
Oliv Eidam





Doing now what patients need nex