sometimes, slides are about the research of the lecturer to illustrate a point

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

today: how to design a sysbio exp

# Challenge 2
# Conclusion

Dr. Peter Blattmann

blattmann@imsb.biol.ethz.ch

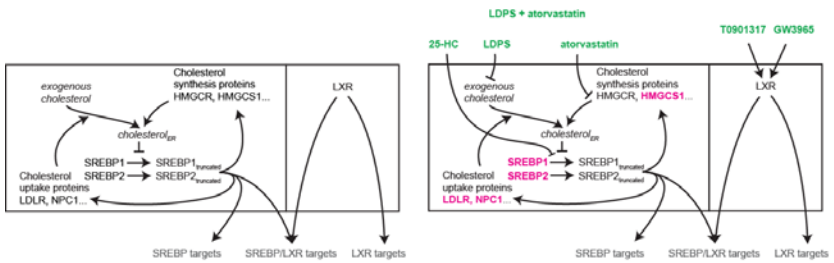
# Learning goals

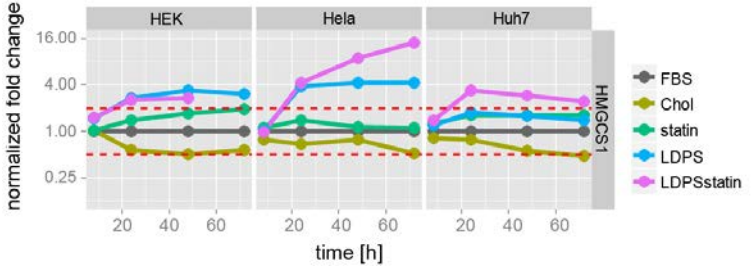
- Understand the importance of different considerations when designing a systems biology project
- Understand the rationale and main features of the presented research project
- Be able to critically judge the results and understand what follow-up experiments would be needed
- Be able to understand the differences between the challenge 1 and challenge 2

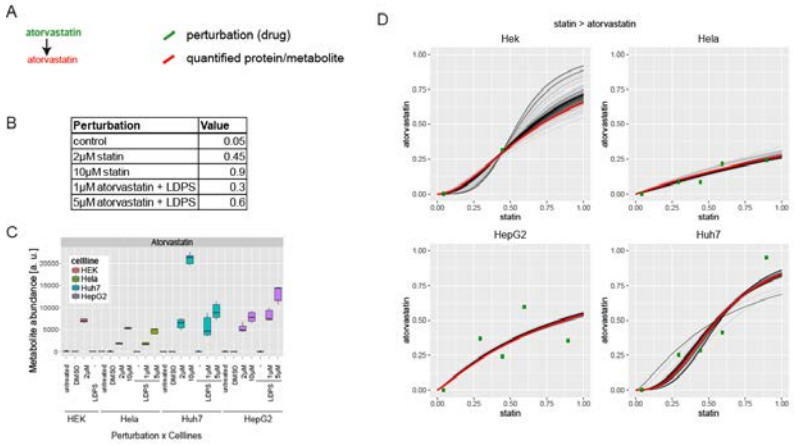
# Outline

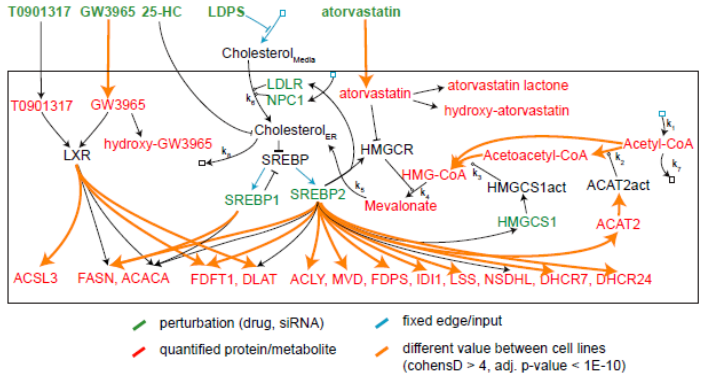
- Questions to the presented research project
- Project proposals from students
- Discussion of research projects
  - What are the similarities/differences between this project and the proposed projects?
  - What would be possible next steps?
- Difference between Challenge 1 and Challenge 2
- From systems biology towards personalized medicine

# Question to the presented research project




OPEN ACCESS
CellPress

Cell Systems
Article

## Systems Pharmacology Dissection of Cholesterol Regulation Reveals Determinants of Large Pharmacodynamic Variability between Cell Lines

Peter Blattmann,[1,7,*] David Henriques,[2] Michael Zimmermann,[1,6] Fabian Frommelt,[1] Uwe Sauer,[1] Julio Saez-Rodriguez,[3,4] and Ruedi Aebersold[1,5,*]

https://doi.org/10.1016/j.cels.2017.11.002

# Student submissions

- ....

sysbio exp:
general outlook/idea
choose model organism. define network and nodes. define how to perturb the system.
define data gathering method: applications of proteomics, transcriptomics, genomics, metabolomics etc.
=> which method and how exactly would you extract the object of interest?
(perhaps, you can also use cell biological methods such as FACS or microscopy based approaches, but mainly, use systems biology approaches)
how to analyze the obtained data? what to do with that data, how to proceed with experiment.
dont forget to mention the controls
how can we generate a model from the data? compare measured data with predicted data using statistics and ML
data gathering and evaluation can also be done using databases
make  a list of methods available. make a list of mathematical tools available and how they are applied in sysbio (like boolean ODE modeling)

# Systems Biology Project

Most Systems Biology project follow this or a similar rationale:

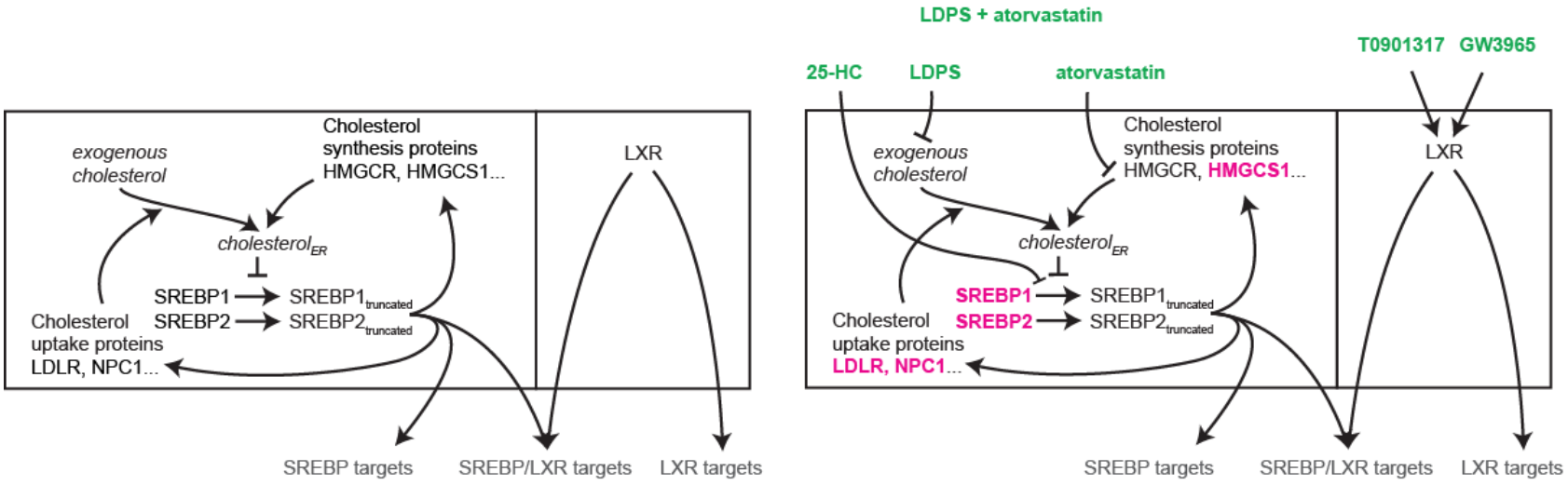
1. Define the network (nodes and edges)
2. Perturb the network while measuring its response
3. Integrate the measured data into the network
4. Validate the model

(Ideker et al. 2001)

possible exam question (self): define nodes and edges for different levels (metab, protein, neural, signalling, RNA, DNA, miRNA, phosphoproteomics, transcriptomics etc)

# Define the network (nodes and edges)

- How can you define a network (nodes, edges)?
- Where can one find information about nodes and edges?
  from databases
- What are important considerations when building a network?
  what kind of data do we get (time series, specific measurements etc)
  what do we even seek? (dissociation constants, whether binding occurred etc)

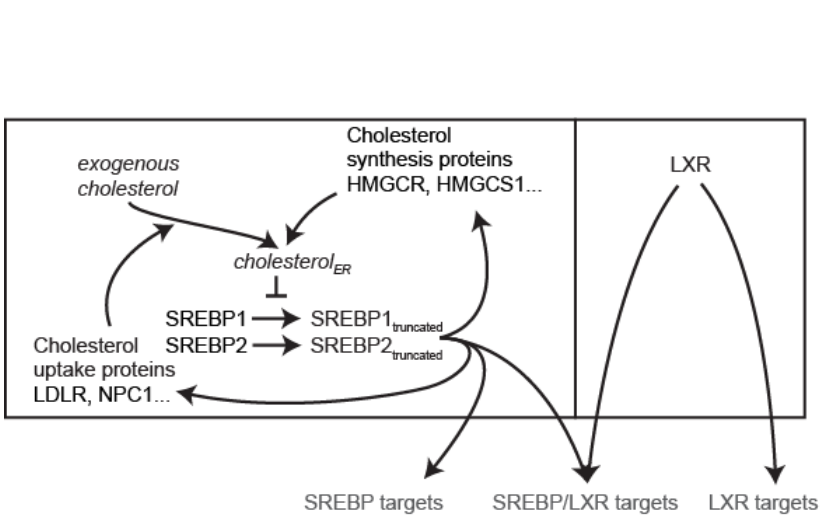


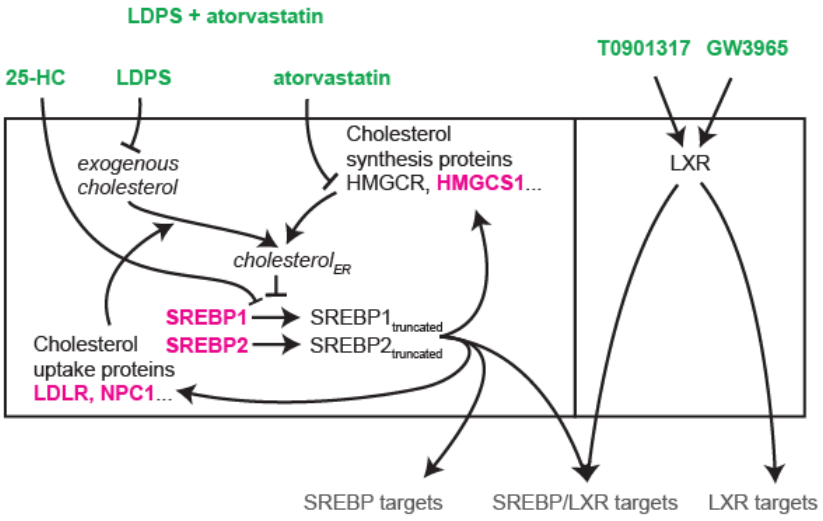
# Perturbations

- How can one perturb a biological system?
- What is important to consider when selecting perturbations?

DNA change: CRISPR, homolg recomb; RNAi, small molecules, but also external factors like pH, temp etc

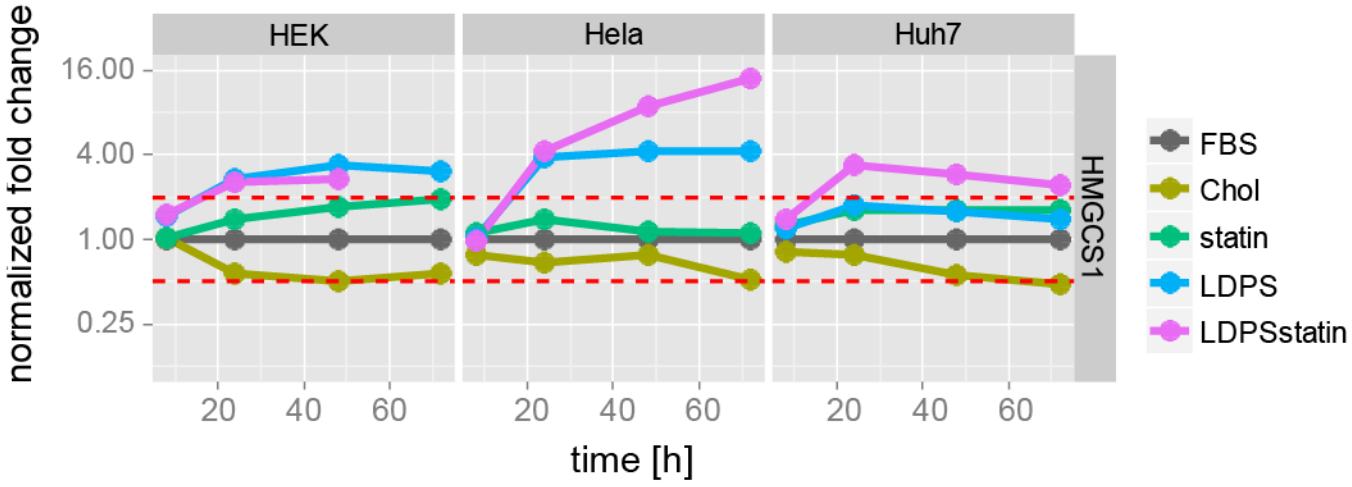
need to know timescale (how long does it take a pertub to happen), is it reversible?, sometimes sevreral pertub needed, since there are redundant pw or synthetic lethality can happen. with several pertub we can perturb the network in different ways and get several results



# Measuring the response - Timecourse

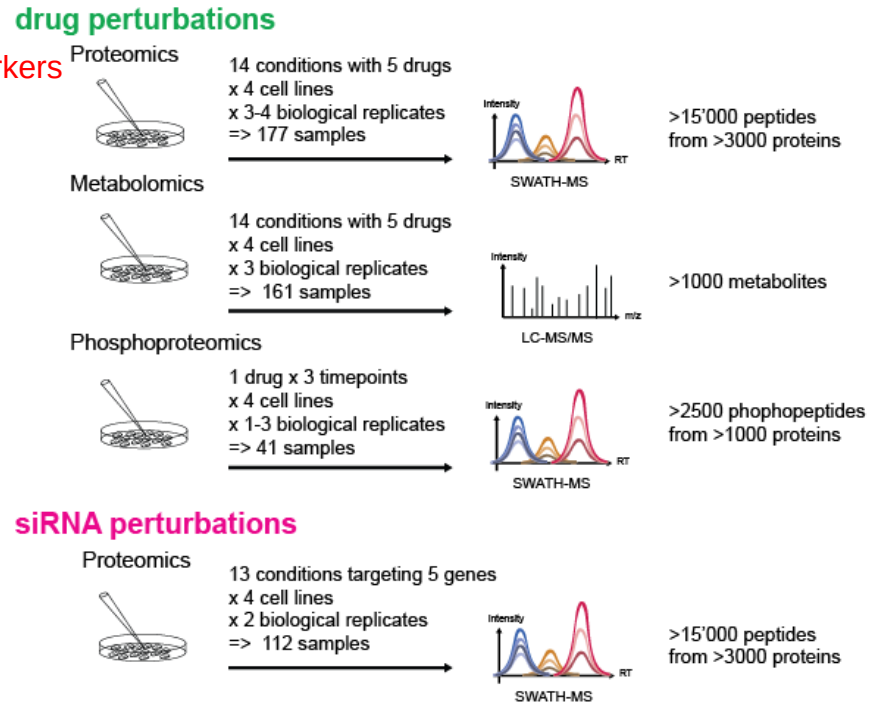
- Timecourse measurements, benefit and drawback?
- How to normalize fold change?



# Measuring the response

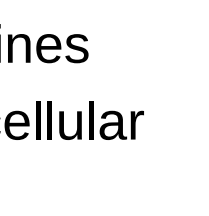
- Untargeted versus targeted measurements
- Number of samples (the number of samples grows expontentially if performing replicates, different cell lines, timepoints etc…)

measure on genotype level:
sequencing (genotype of cell)
epigenetic state: look for epigenetic markers across different cell lines

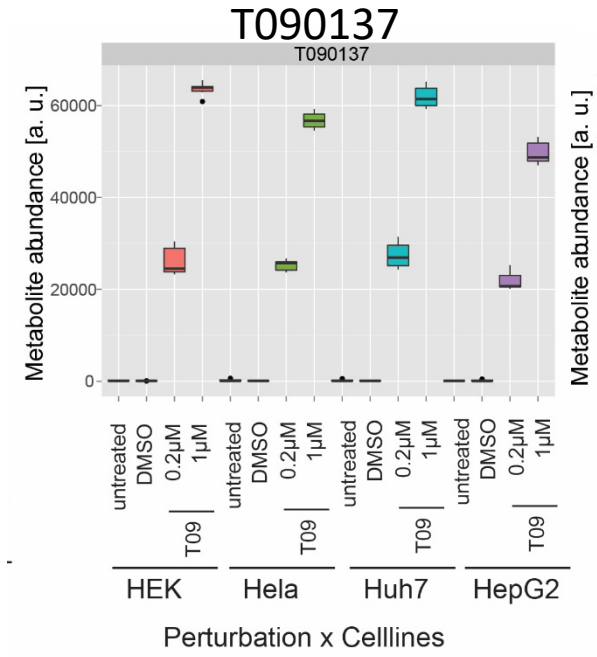


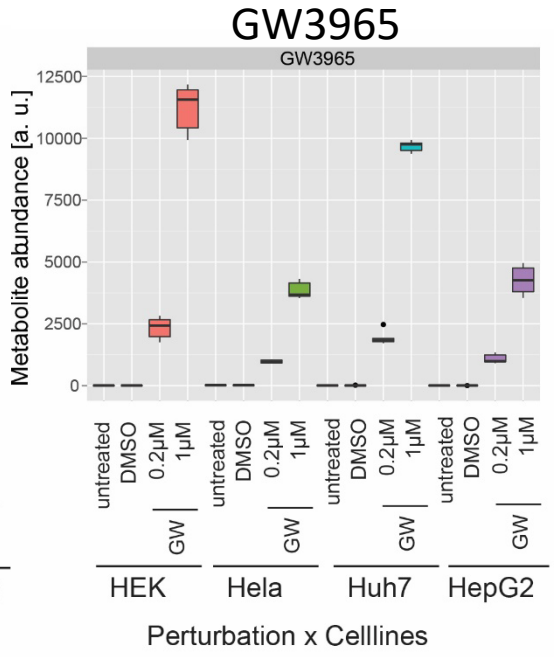
# Results - intracellular drug concentration

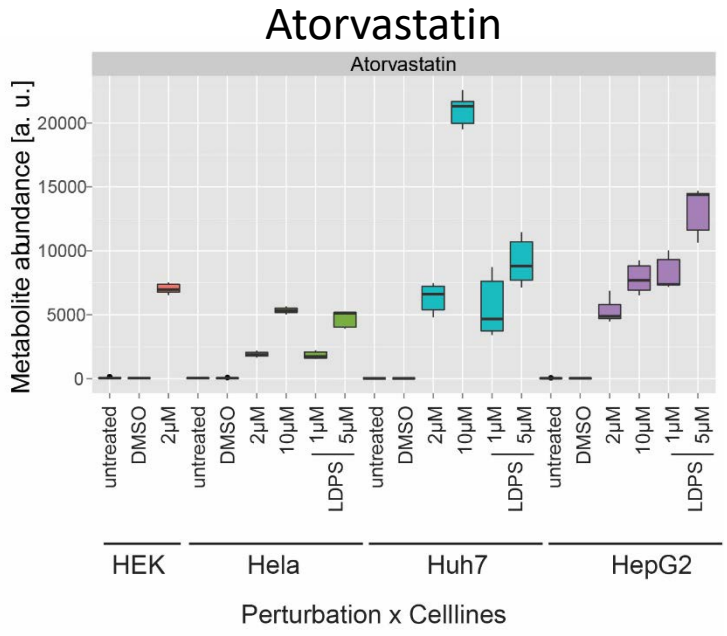
- All drugs show dose-dependent concentrations
- T090137 shows similar concentrations in all cell lines
- For other drugs (GW3965, atorvastatin) the intracellular concentration is cell line and media dependent



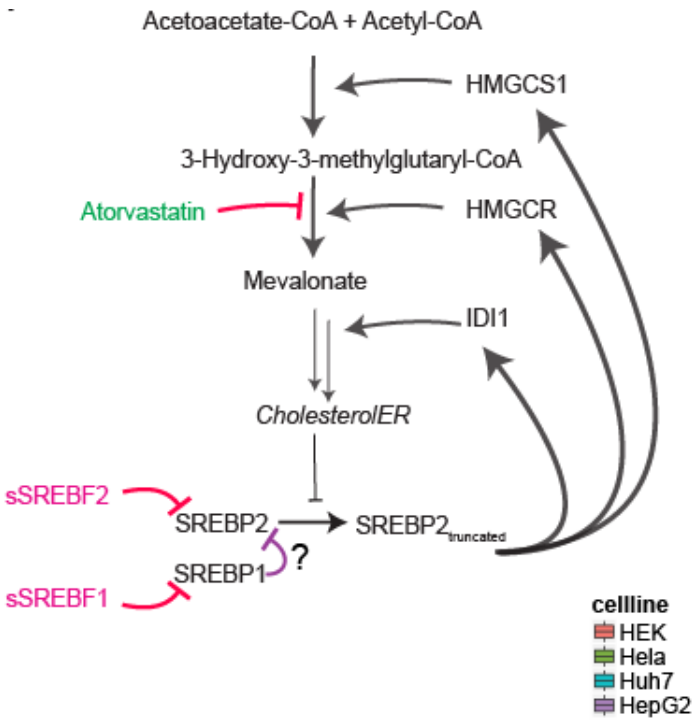

n=6; 3 biological replicates, 2 injections

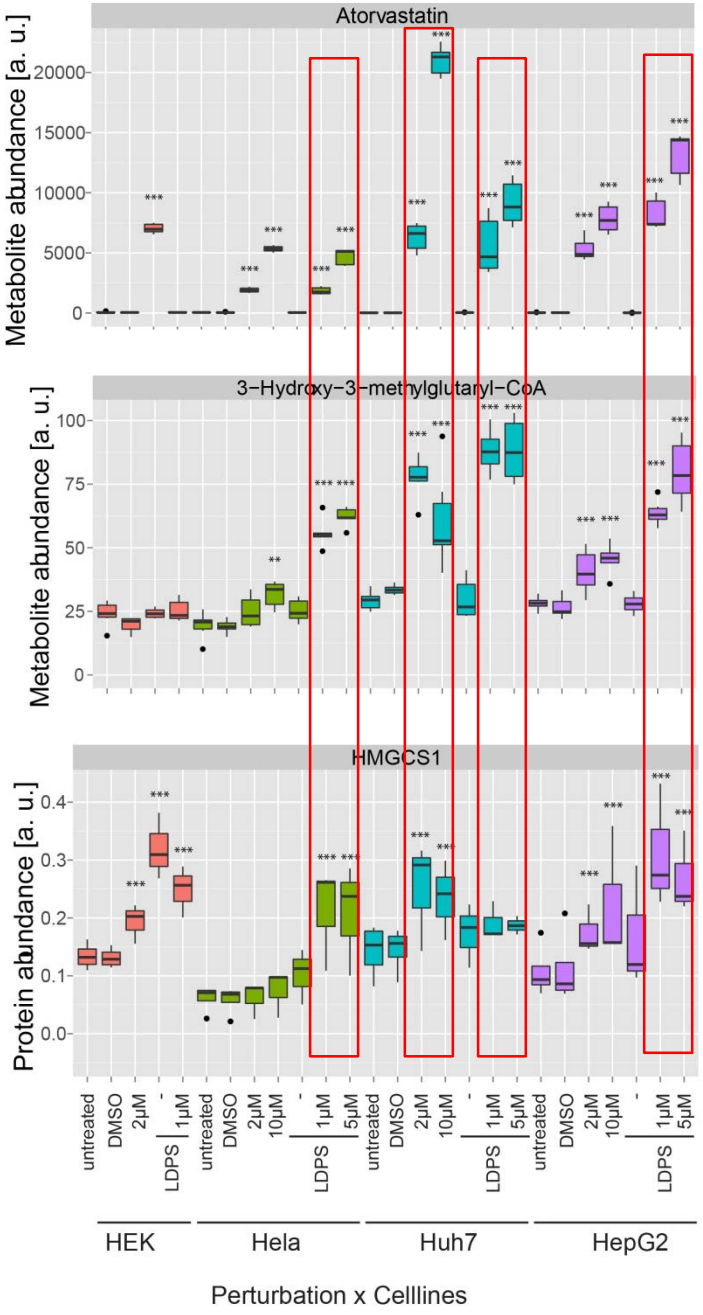


# Results – SREBP feedback

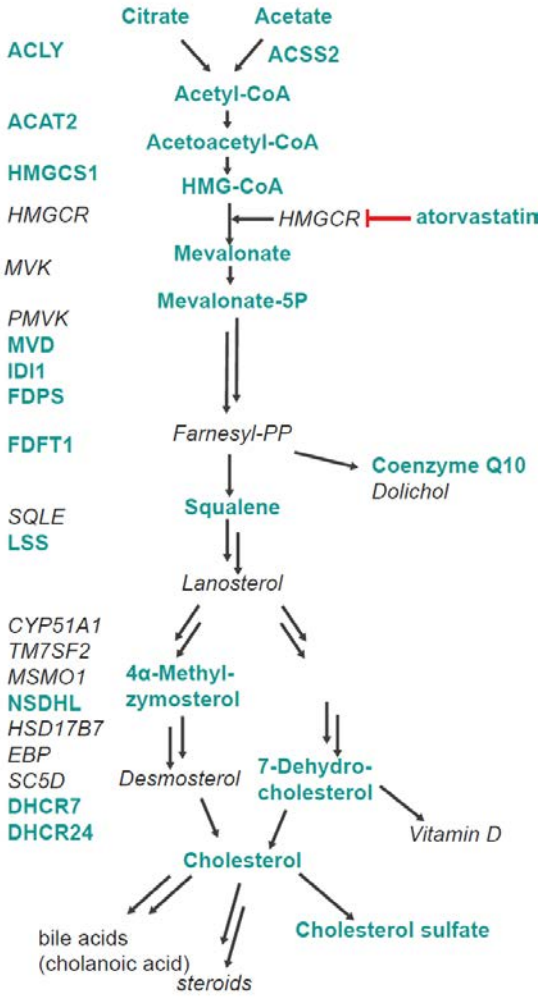


# Results – profiling a whole pathway

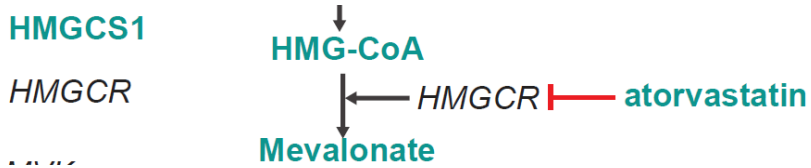
| Protein | SwissProt ID | nPeptides | nFragments | controls | | | | GW3965 | | | | T090137 | | | | 25HC | | | | atorvastatin | | | | LPDS | | | | LPDS+ atorvastatin | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | HEK | Hela | Huh7 | HepG2 | HEK | Hela | Huh7 | HepG2 | HEK | Hela | Huh7 | HepG2 | HEK | Hela | Huh7 | HepG2 | HEK | Hela | Huh7 | HepG2 | HEK | Hela | Huh7 | HepG2 | HEK | Hela | Huh7 | HepG2 |
| ACLY | P53396 | 7 | 35 | | | | | | | | | | | | ↑ | | | | | | | | ↑ | | | | | | ↑ | | ↑ |
| ACAT2 | Q9BWD1 | 6-7 | 31-33 | | | | | | | | | | | | | | | | | | | | ↑ | | | | | | ↑ | | ↑ |
| HMGCS1 | Q01581 | 6-7 | 31-33 | | | | | | | | | | | | | ↓ | | | | ↑ | | ↑ | ↑ | ↑ | | | | ↑ | ↑ | | ↑ |
| MVD | P53602 | 5 | 24-25 | | | | | | | | | | | | | | | | | | | | ↑ | | | | | | ↑ | | ↑ |
| IDI1 | Q13907 | 6 | 30 | | | | | | | | | | | | | ↓ | | | | | | | ↑ | ↑ | ↑ | | | | ↑ | | ↑ |
| FDPS | P14324 | 7 | 35 | | | | | | | | | | | | | | | | | | | | ↑ | | ↑ | | | | ↑ | | ↑ |
| FDFT1 | P37268 | 7 | 34-35 | | | | | ↑ | | ↑ | ↑ | | | | ↑ | | | | ↓ | | | ↑ | ↑ | | ↑ | | ↑ | ↑ | ↑ | ↑ | ↑ |
| LSS | P48449 | 6-7 | 30-35 | | | | | | | | | | | | | | | | | | | | | | ↑ | | | ↑ | ↑ | ↑ | ↑ |
| NSDHL | Q15738 | 3 | 15 | | | | | | | | | | | | | | | | | | | | | | | | | ↑ | ↑ | ↑ | ↑ |
| DHCR7 | Q9UBM7 | 2 | 8-10 | | | | | | | | | | | | | | | | | | | | | | | | | | ↑ | | ↑ |
| DHCR24 | Q15392 | 3 | 17 | | | | | | | | | | | | | | | | | | | | | | | | | | ↑ | | |

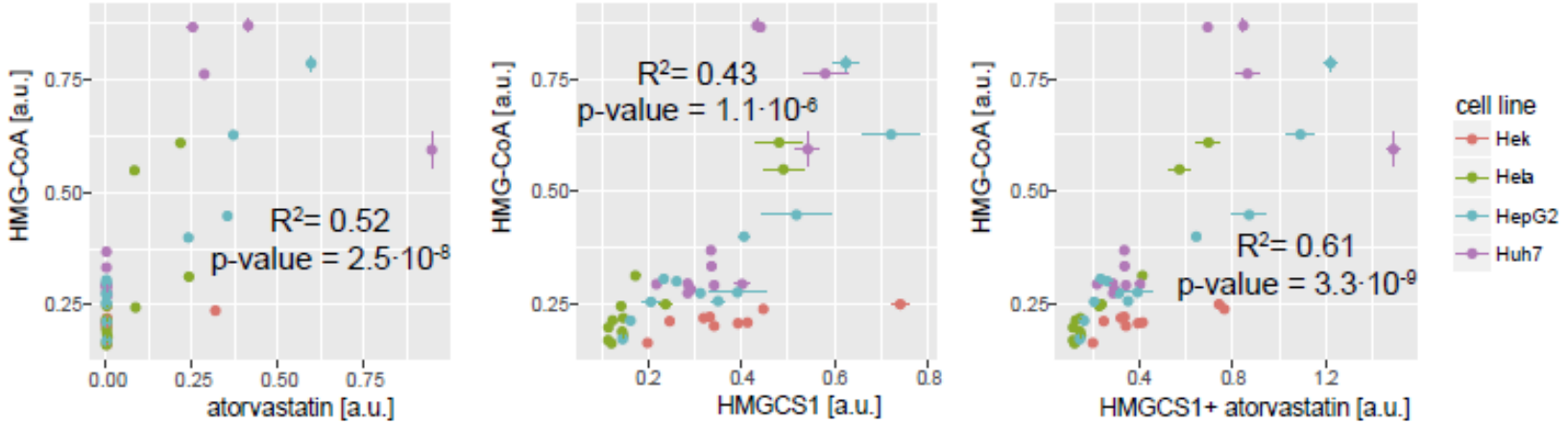
| Metabolite | HMDB id | m/z | controls | | | | GW3965 | | | | T090137 | | | | 25HC | | | | atorvastatin | | | | LPDS | | | | LPDS+ atorvastatin | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | HEK | Hela | Huh7 | HepG2 | HEK | Hela | Huh7 | HepG2 | HEK | Hela | Huh7 | HepG2 | HEK | Hela | Huh7 | HepG2 | HEK | Hela | Huh7 | HepG2 | HEK | Hela | Huh7 | HepG2 | HEK | Hela | Huh7 | HepG2 |
| Citric acid | HMDB00094 | 191.020 | | | | | | | | ↓ | | | ↓ | ↓ | | | | | | | | | | | | | | | | |
| Acetic acid | HMDB00042 | 59.013 | | | | | | | | | | | | | | | | | | | | | ↑ | | | | | | | |
| Acetyl-CoA | HMDB01206 | 808.118 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Acetoacetyl-CoA | HMDB01484 | 850.121 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| HMG-CoA | HMDB01375 | 910.149 | | | | | | | | | | | | | | | | | | | ↑ | ↑ | | | | | | ↑ | ↑ | ↑ |
| Mevalonate | HMDB00227 | 147.066 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Mevalonate-5P | HMDB01343 | 227.032 | | | | | | | | | | | | | | | | | | | | | ↑ | | | | | | | |
| Squalene | HMDB00256 | 409.387 | | | | | | | | | | | | | ↑ | | | | | | | | | | | | | | | |
| 4a-Methylzymosterol | HMDB01217 | 397.350 | | | | | | | | | | | | | ↑ | | | | | | | | | | | | | | | |
| 7-Dehydrocholesterol | HMDB00032 | 383.334 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Cholesterol | HMDB00067 | 385.347 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Coenzyme Q10 | HMDB01072 | 861.676 | | | | | | | | | | | | | | | | | | | ↓ | | | | | | | ↓ | | |
| 3b,12a-Dihydroxy-5a-cholanoic acid | HMDB00348 | 391.286 | | | | | | | | | | | | | | | | ↓ | | | | | | | | | | | | |
| 7-Ketocholesterol | HMDB00501 | 399.327 | | | | | | | | | | | | | | | | | | | | | | | ↑ | | | | | ↑ |
| Cholesterol sulfate | HMDB00653 | 465.304 | | | | | | | ↓ | ↓ | | | ↓ | ↓ | | | | | | | | | ↓ | ↓ | ↓ | | | ↓ | ↓ | ↓ |



# Relationship between abundances are not linear

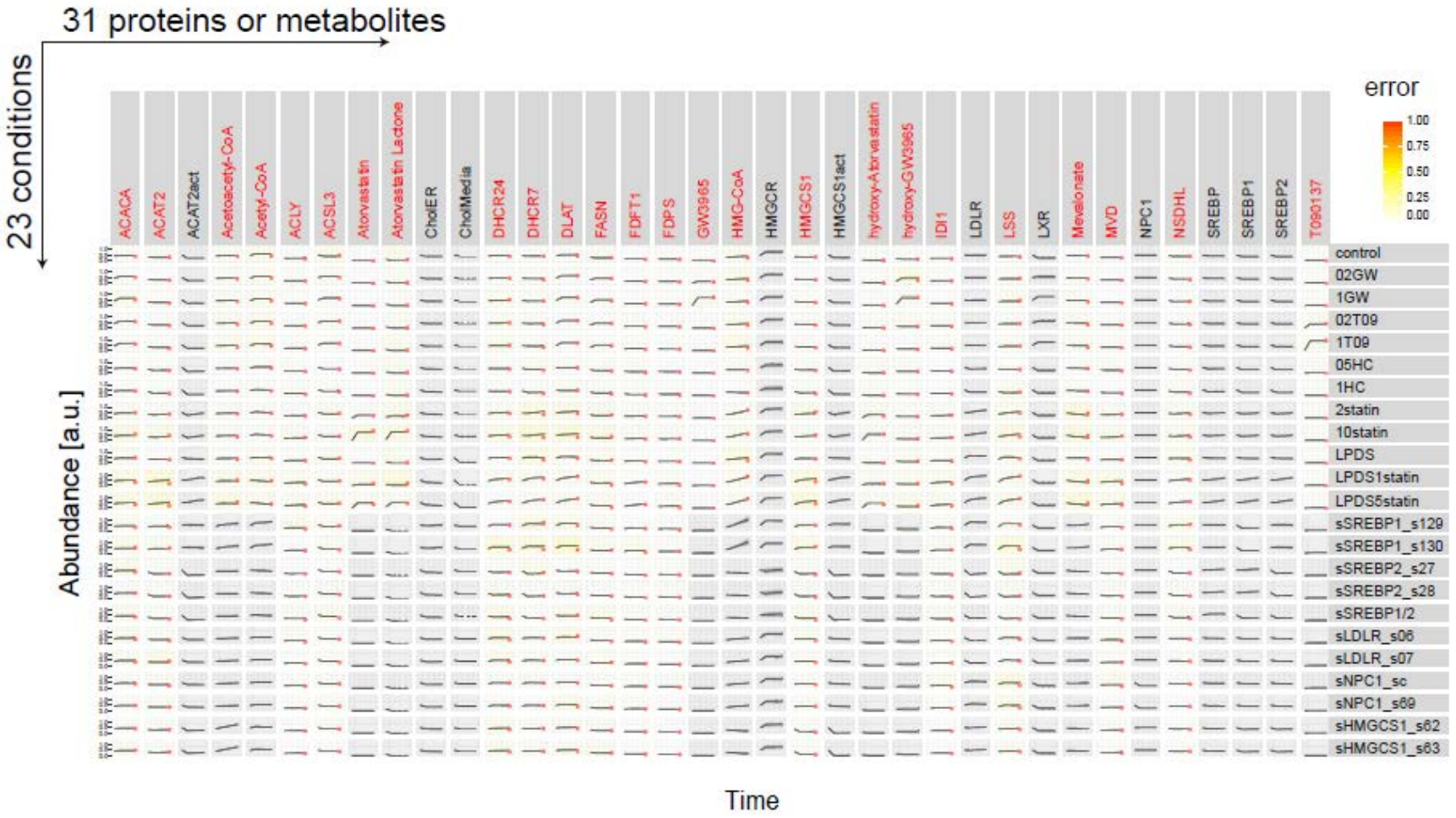
- There is a significant correlation, but the relationship are not simply linear
- More sophisticated models are required to explain the variability in drug response



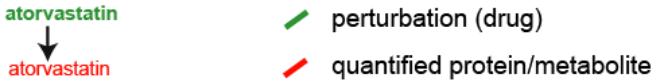
# Integrate data into a network model

- Logic-based ODE model



# Transfer function describing functional edges

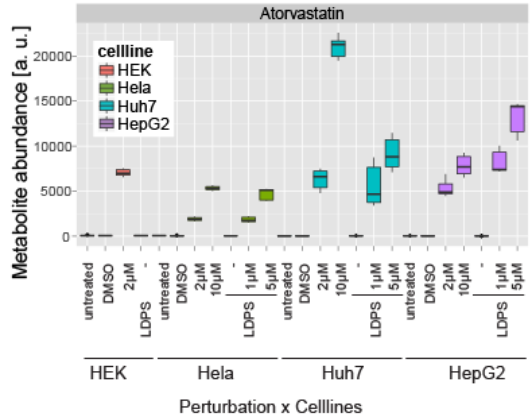
A



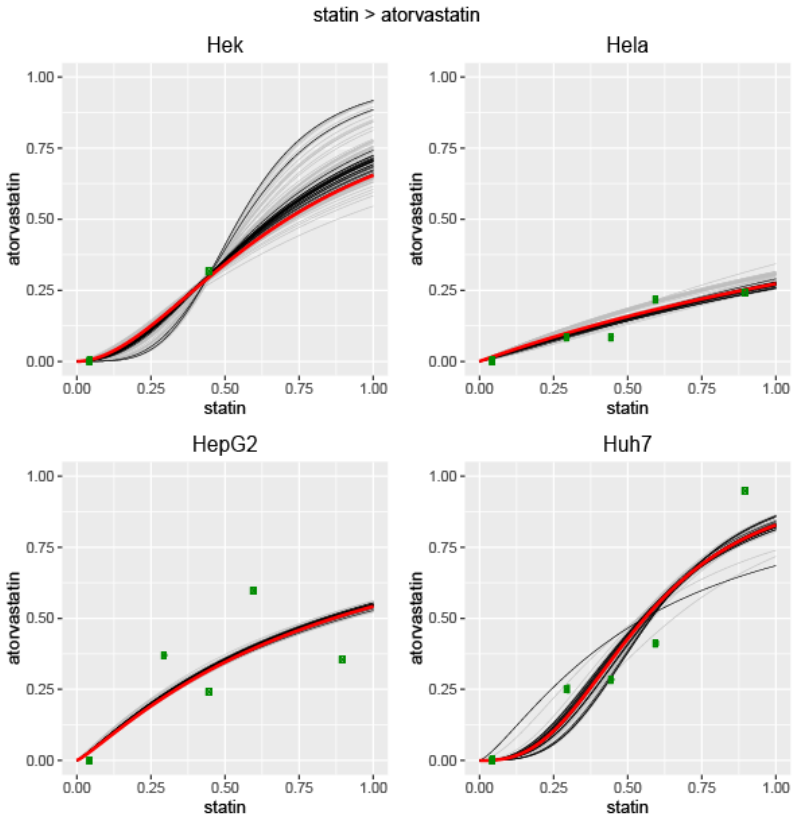

B

| Perturbation | Value |
|---|---|
| control | 0.05 |
| 2µM statin | 0.45 |
| 10µM statin | 0.9 |
| 1µM atorvastatin + LDPS | 0.3 |
| 5µM atorvastatin + LDPS | 0.6 |

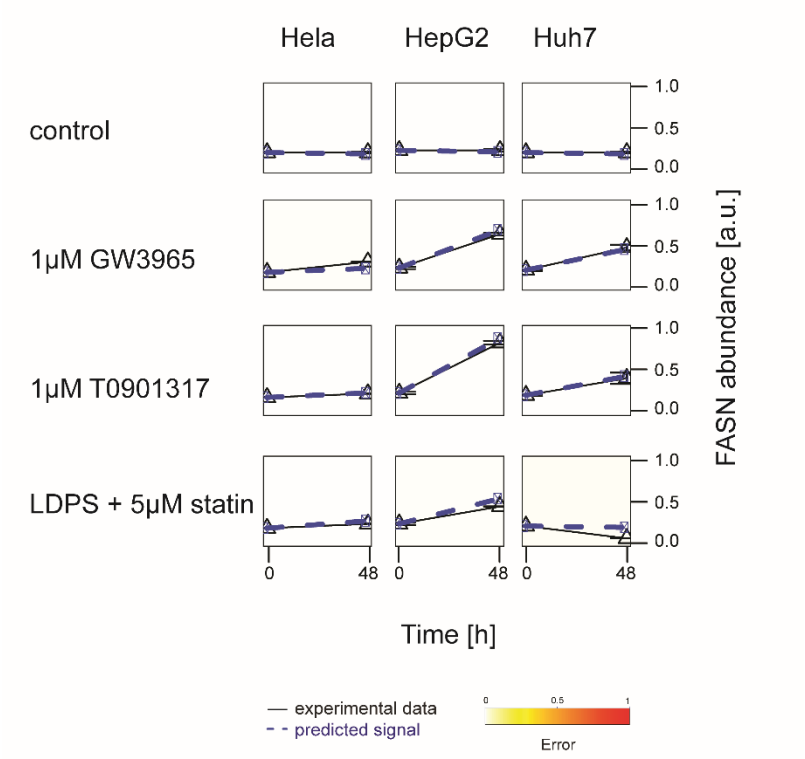
C

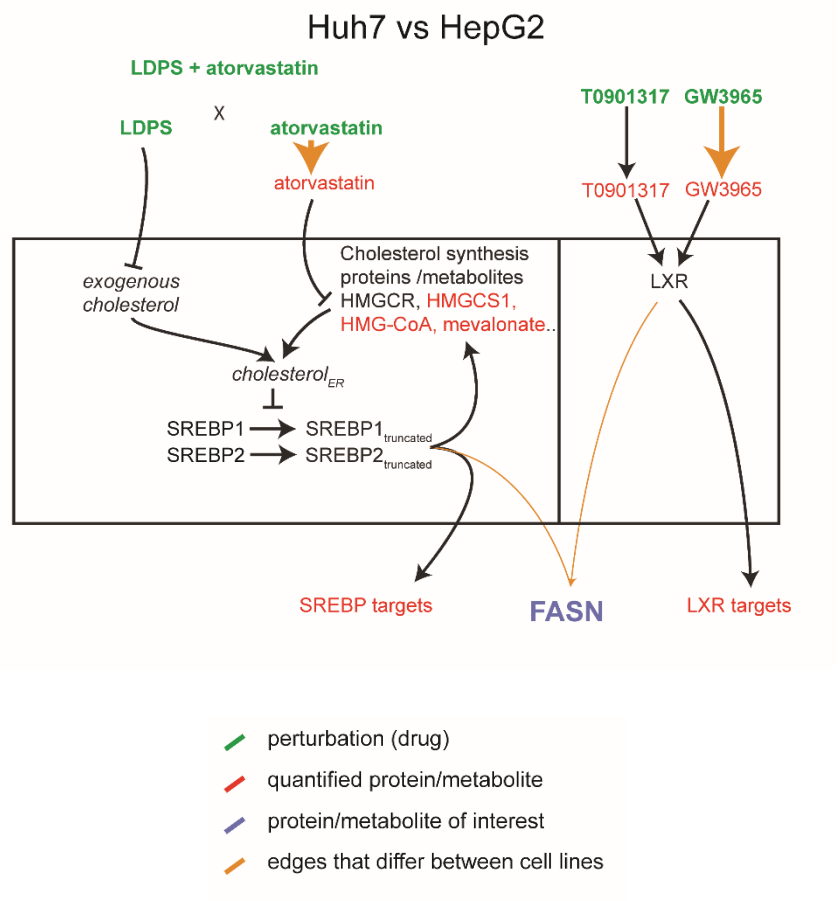



D



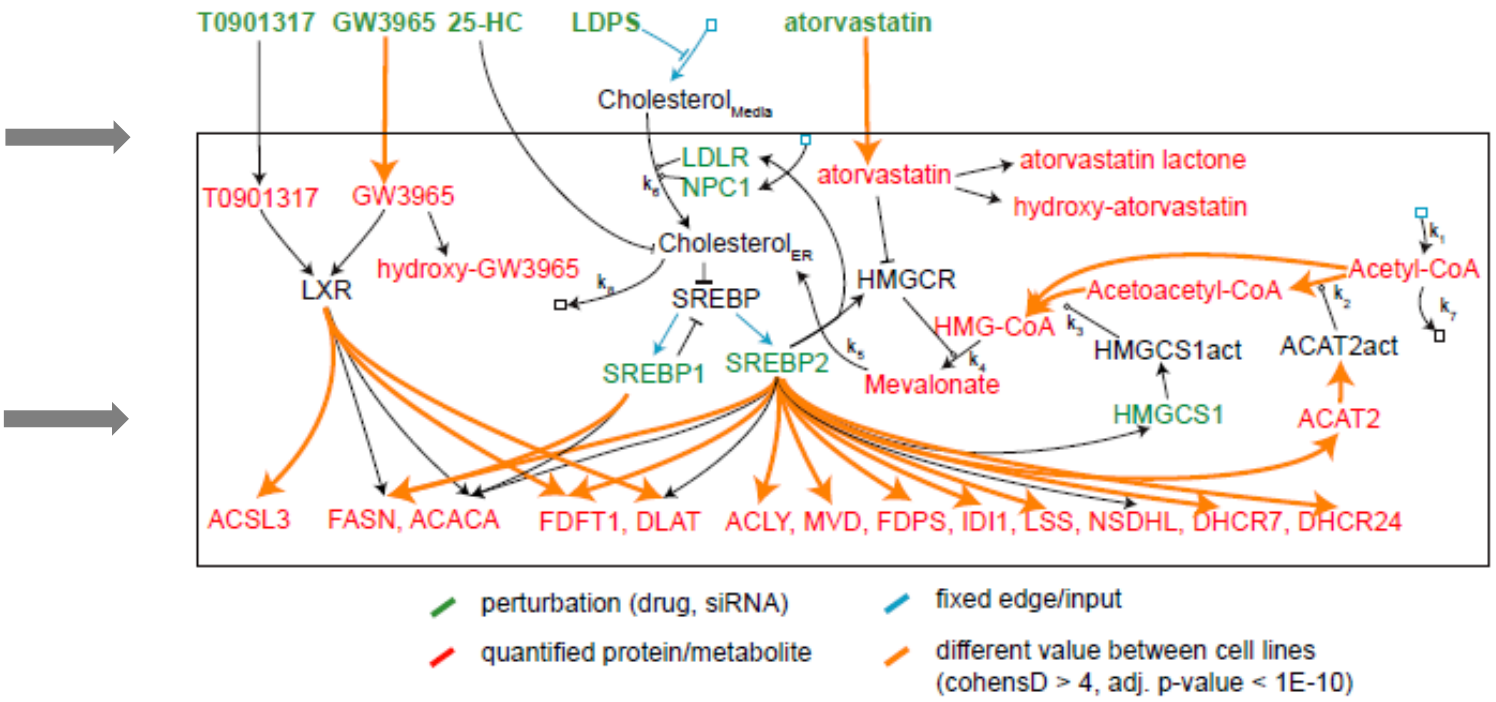
# Comparison between cell lines



# Comparison between cell lines

- Differences in drug uptake and transcription factor effects underlie the heterogeneity in drug response
- The variability in response phenotypes cannot be explained by a few different factors but only with the network as a whole.



# What are the next steps?

- What could one explain with this model?
- Could one predict an outcome with this model?
- How could one validate the quality of such a model?

- What future (modeling) experiments could one do?

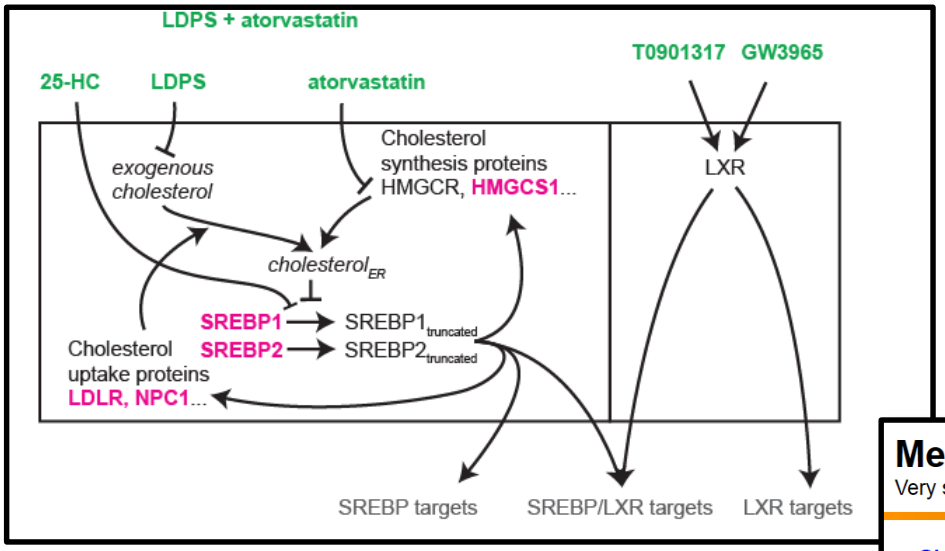
# Systems Biology Project

Most Systems Biology project follow this rationale:

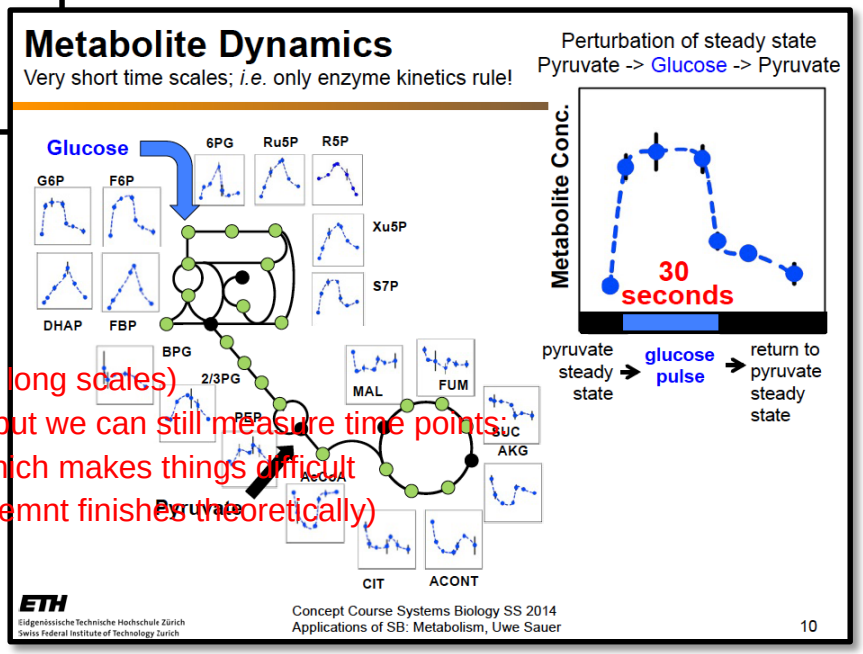
1. Define the network (nodes and edges)
2. Perturb the network while measuring its response
3. Integrate the measured data into the network
4. Validate the model

(Ideker et al. 2001)

# What was the second challenge about and how is it different from the first?




## Challenge II




in challenge 1, we are interest in short time scales (but also in long scales)
in chall 2, we generally wanna know how the systems works, but we can still measure time points; problem is in chall 2, there are bigger time scale variations, which makes things difficult
metabolites: seconds. proteins: hours (take weeks until experiemnt finishes theoretically)

## Challenge I

challenge 1 networks are more detailed, we know the stoichiometry already and have information on K-values which are not known in challenge 2

# What are the main challenges?

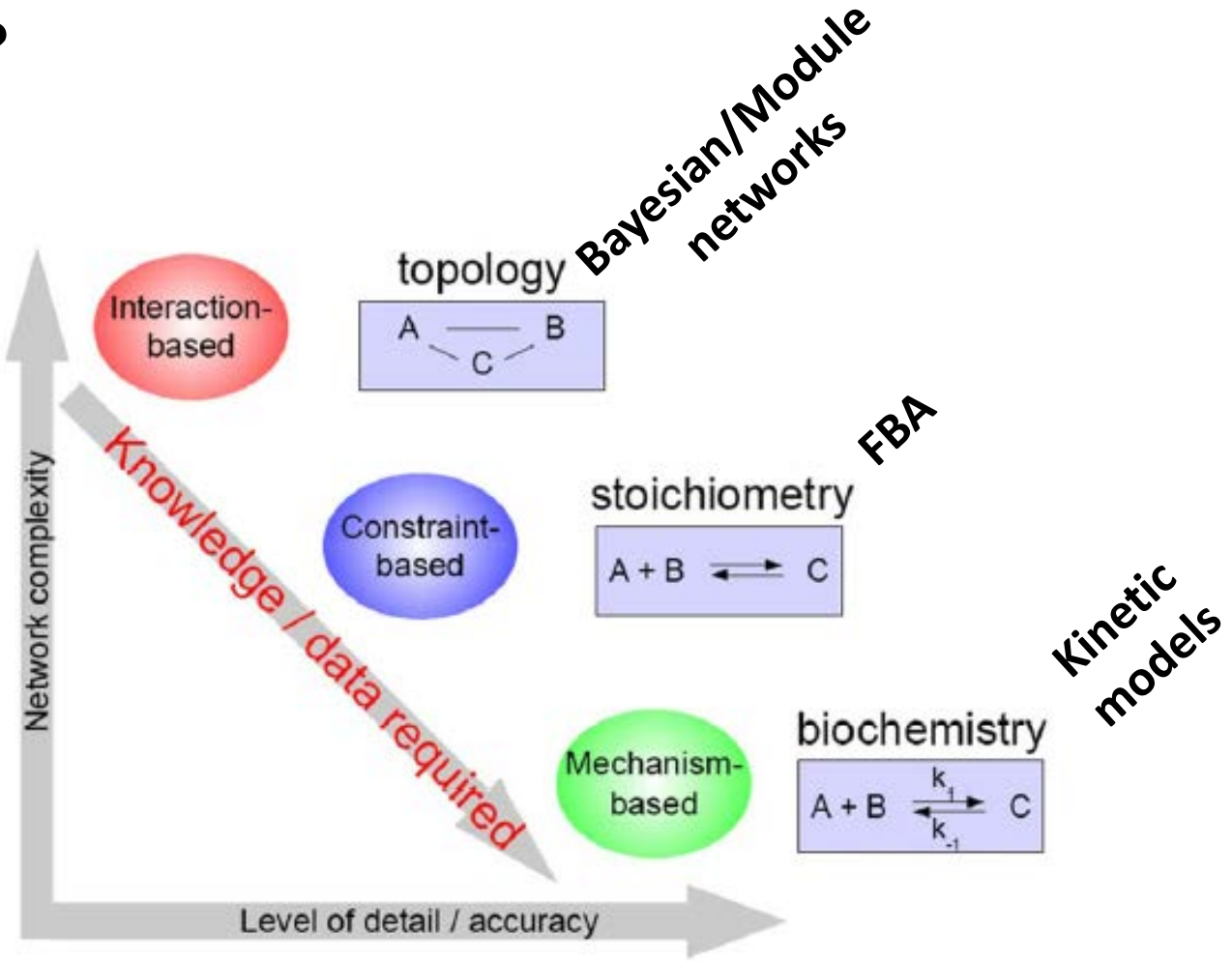
- Regulatory network reconstruction
- Selection of informative targets to measure
- Technology to measure targets

# Was challenge 2 a prototypical challenge?

- To what other biological processes do you think one could apply a similar approach?
- Which prerequisites need to be fullfilled?

partial knowledge of basics of network needed (approach requires prior knowledge, which has to be quite accurate)
experimentally: perutbration of proteins possible, measurement of proteins possible?

# What different models did we encounter in the course?

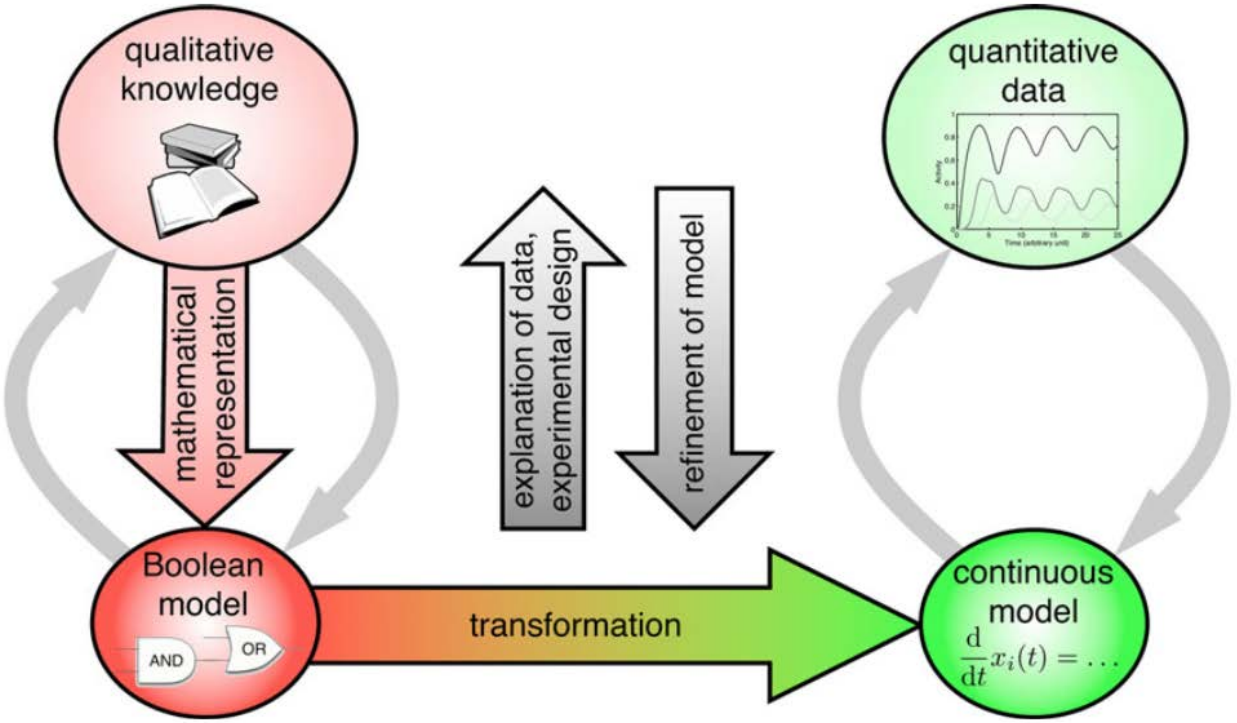



→ Where do Boolean networks fit?

hybrid between interaction based and mechanism based model

# Logic-based ODE modeling

- Transformation of boolean models into ordinary differential equations (ODE) allows kinetic modeling despite rather scarce knowledge of mechanism




Wittmann et al. 2009

# Modeling

- What were the aims in Challenge 2?
- What kind of data did we use for modeling?

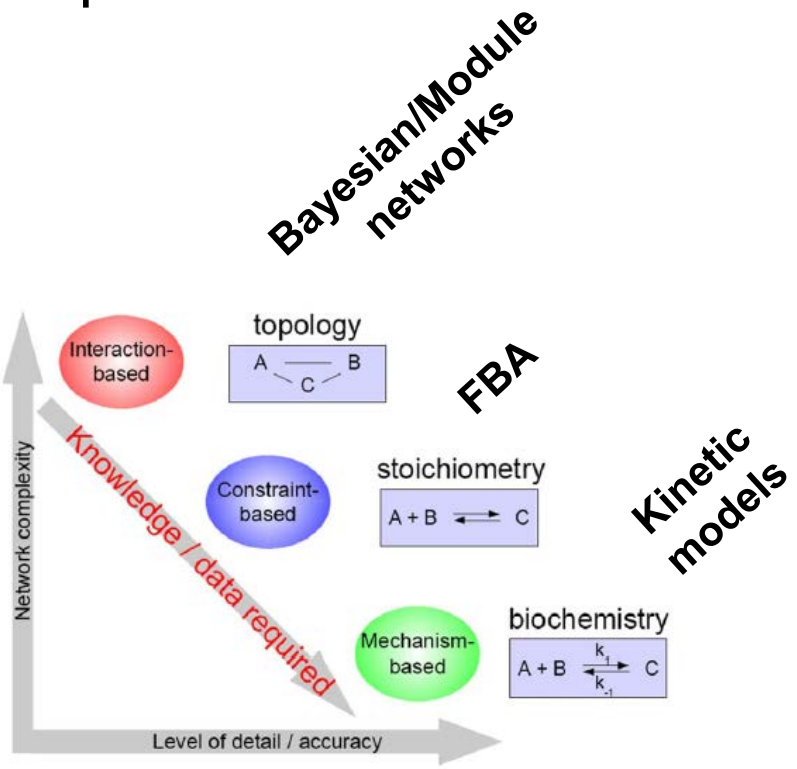
- What are the modeling challenges with regards to model/data situation?

- What would be the ideal data to support the modeling further?
- Is it feasible to generate this kind of data?

# Different modeling approaches

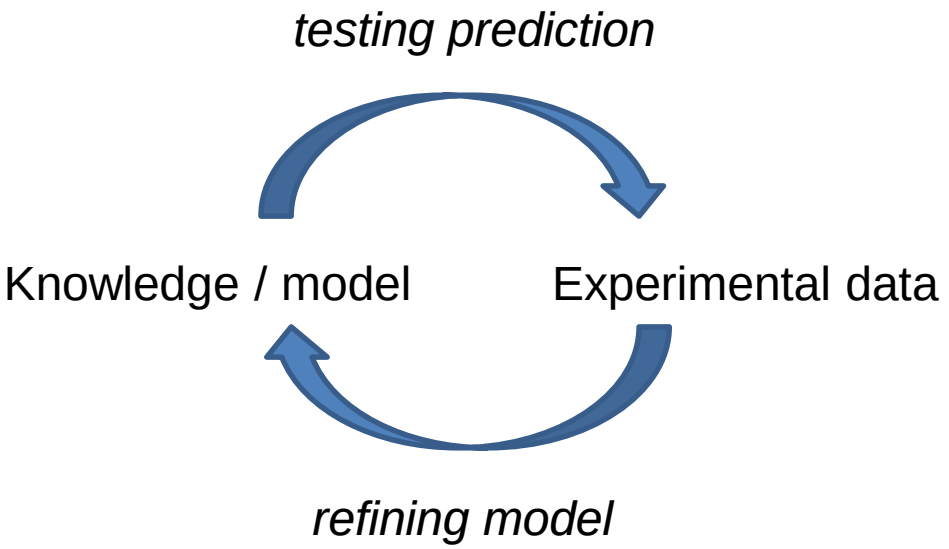
answer this yourself

- Which model would you use for what type of question/problem?



# Validation of a model

- Aggreement of prediction and experimental data
  1. Explain the data that was used to build the model
  2. Predict independent data



# “ useful” models

- *What is a “useful” model?*

*“Essentially, all models are wrong, but some are useful.”* *(George Box)*

  - *Able to accurately predict a certain behaviour*
  - *Generating hypothesis*
  - *…?*

# Personalized/precision medicine

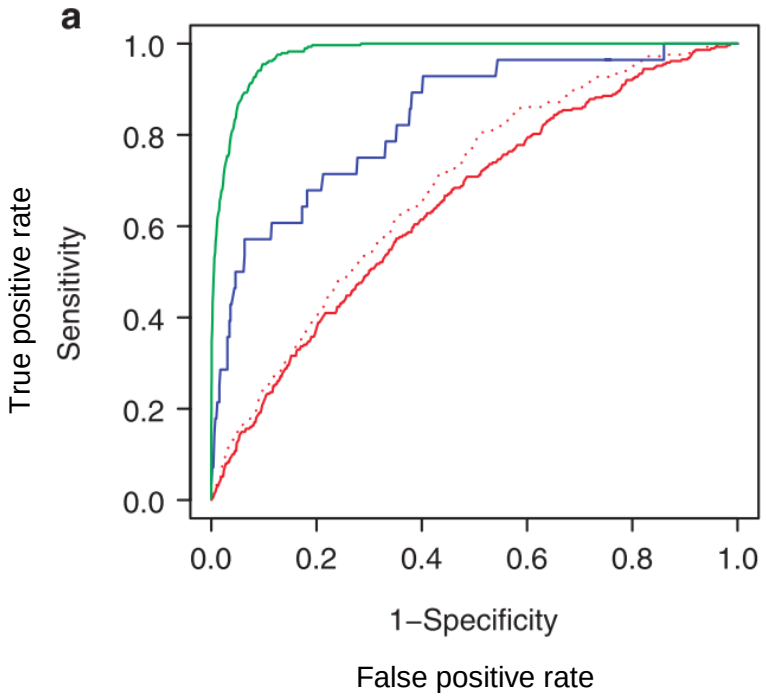
- Individuals are different (genetically and environment)
- Individuals have different disease predispositions
- Individuals react differently to different drugs

→ Use the right medicine for the right person

4P Medicine: Predictive, Personalized, Preventive, Participatory
(by Lee Hood)

# Predictability of model or tests

- Area under the Receiver operating curve (ROC)
- Example predicting 5% tallest people :
  green: 80% of genetic variance known (simulation) (AUC: 0.97),
  blue: Victorian method (AUC=0.83); red: from 54 GWAS loci (0.65-0.68)




random if line is straight in 45°

(Aulchenko et al. 2009)

**Other examples:**

**Prostate cancer** (Gleason score > 7) AUC:
PSA alone 0.56-0.78,
STHLM3 0.74
(Grönberg et al. 2015, Thompson et al. 2005)

**Breast cancer** (by age 70) RR:
Normal incidence ~12%
With BRCA1 mutation ~65%
With BRCA2 mutation ~45%
(Antoniou et al. 2003)

prediction better if curve is like the green one, bad if like red one

# Predictive medicine

- From the 4Ps probably the «predictive» is the biggest challenge today

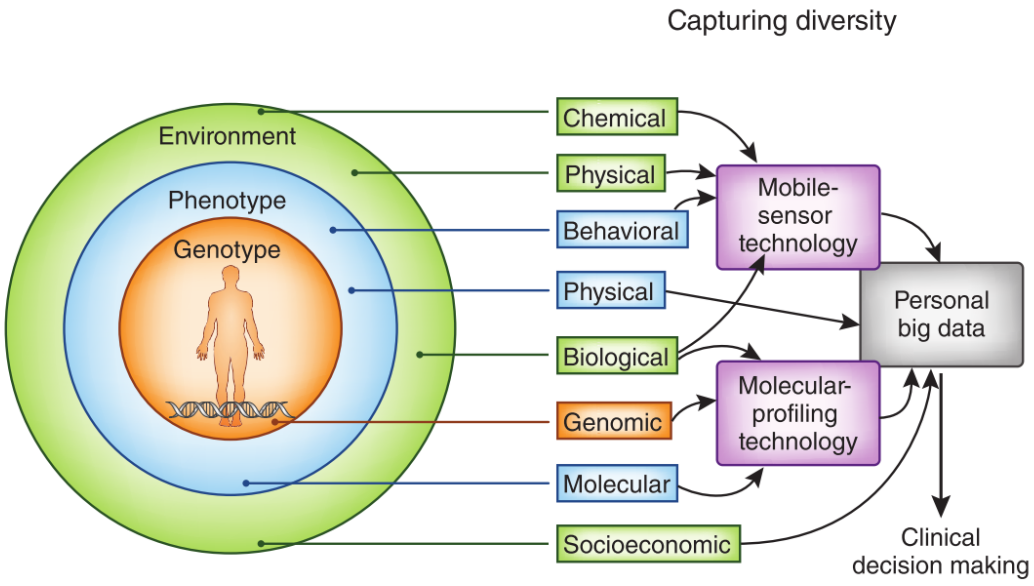
- Systems Biology approaches may be helpful in developing more predictive models for complex diseases

- Big data and big funding initatives

Criticial voices:

- “low-input, high-throughput, no-output biology” (Sydney Brenner)

# Promises and challenges




Chaussabel & Panendran Nature Immunology 2015