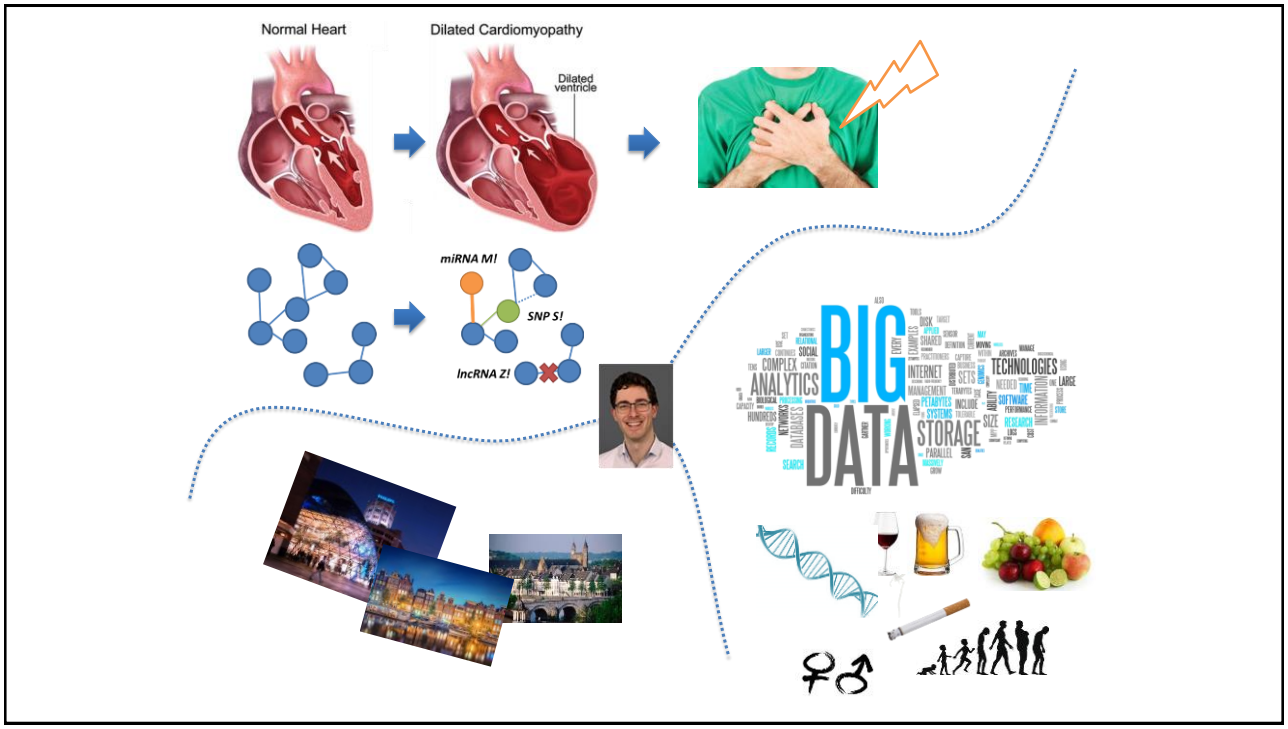
Modeling the genetics and metabolism of heart failure

A multi-omics story in two parts

Michiel Adriaens, PhD
Maastricht Centre for Systems Biology – MaCSBio
Maastricht University

macsbio
MAASTRICHT CENTRE FOR SYSTEMS BIOLOGY
Maastricht University

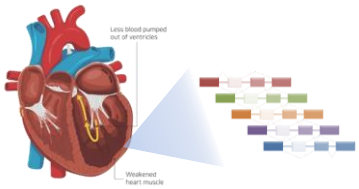
1



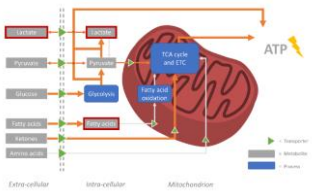
2

Outline

1. Genetic control of gene transcription in heart failure



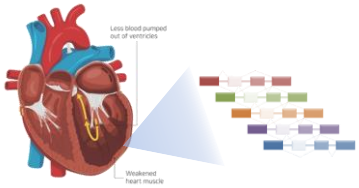
2. Modeling the metabolism of the failing heart



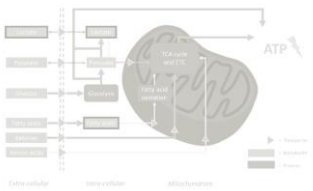
3

Outline

1. Genetic control of gene transcription in heart failure

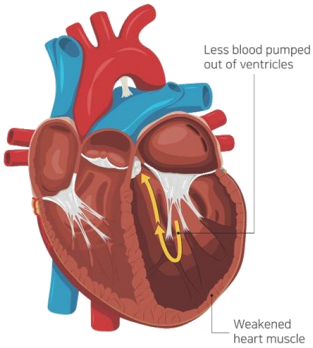


2. Modeling the metabolism of the failing heart

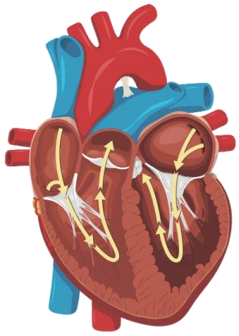


4

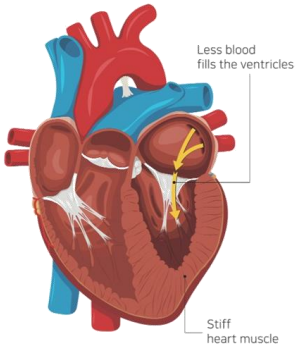
What is heart failure?



Systolic Dysfunction



Normal

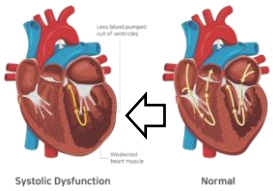


Diastolic Dysfunction

5

Heart failure ↔ complications

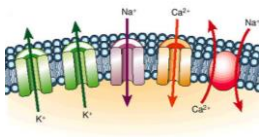
Structural remodeling



Heart Failure

Arrhythmias
Sudden cardiac death
Metabolic impairments

Electrical remodeling



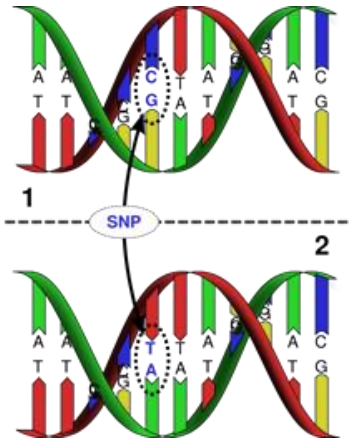
- *Complications are patient specific*
- *What explains this difference?*

Genetics?

6

Genetic association studies

- SNP (single nucleotide polymorphism):
  - A variation in a single nucleotide that occurs at a specific position in the genome
- Example SNP:
  - Base C may appear in most individuals
  - Base T occurs in some individuals
  - C an T are called the “alleles” of the SNP
- We all have two copies of every chromosome (and every gene!)



Allele C

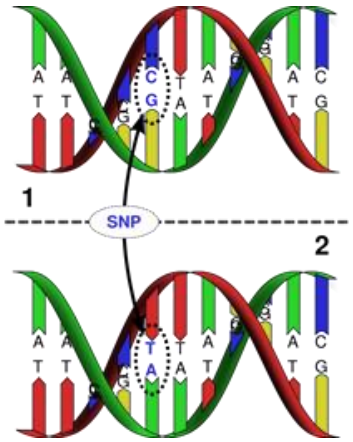
Allele T

Genotype = CT

7

Genetic association studies

- Variations in the DNA affect
  - Disease development
  - Response to pathogens, chemicals, drugs
- How to find these variations?
  - Genotyping of individuals
  - Comparing e.g. cases versus controls



Allele C

Allele T

Genotype = CT

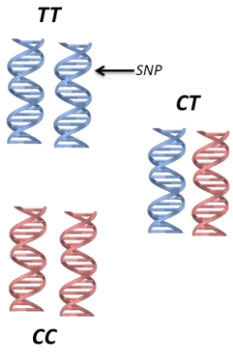
8

Genetic association studies

Example: *E-cadherin* gene SNP and prostate cancer

	Cases	*Controls*
TT or CT	61	84
CC	21	104
Total	82	188

$OR_{TT/CT\ vs.\ CC} = 3.6$



Conclusion: the 'T' allele is associated with prostate cancer (3.6-fold increased risk)

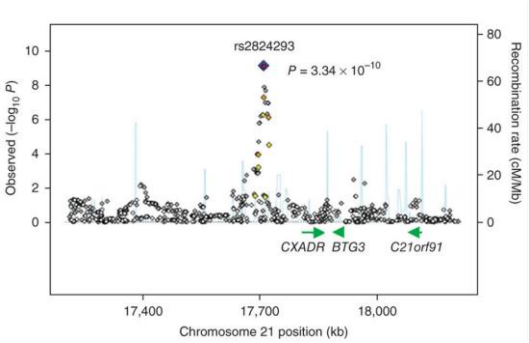
Source: Verhage *et al.* Int J Cancer 2002;100:683-5 (adapted)

Source: Verhage et al. (2002)

9

Genome-wide association studies

- **GWAS =**
  - Genotype thousands of variants in a population of cases and controls
  - Genetic association for each variant
- **GWAS** have identified many genetic variants associated with complex traits and diseases
  - *Example below: susceptibility to arrhythmias after MI*

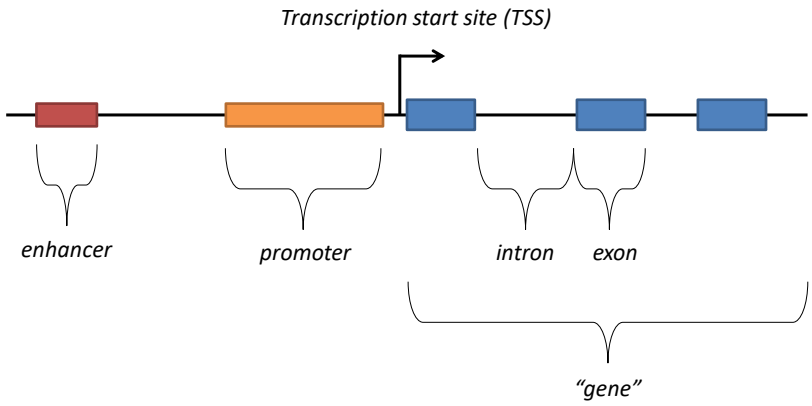


Genes?
Mechanism?

Bezzina, Pazoki, et al., Nat Gen (2010)

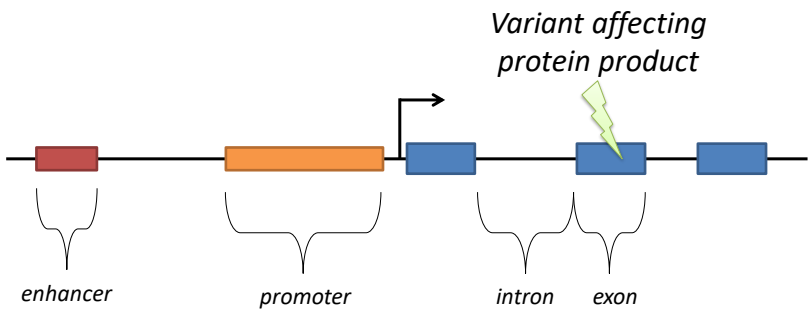
10

Genetic control of gene transcription



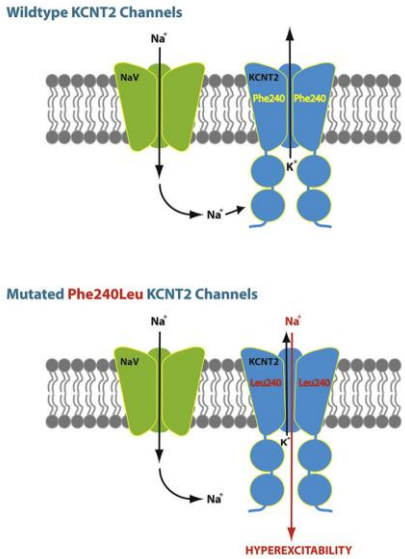
11

Genetic control of gene transcription



12

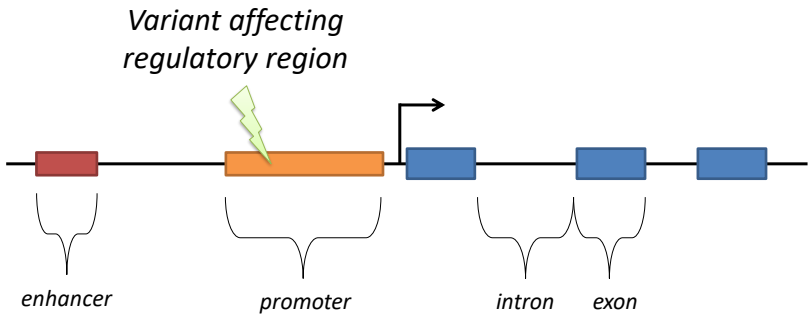
Genetic variants in exons can influence protein structure



Gururaj et al. (2017)

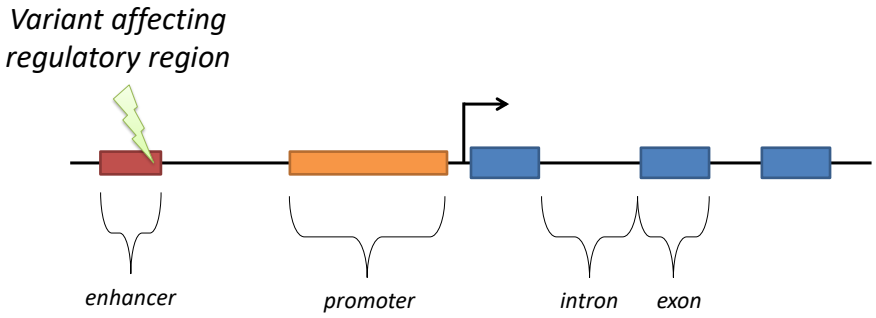
13

Genetic control of gene transcription



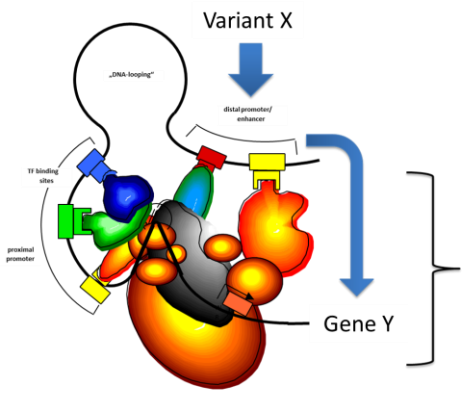
14

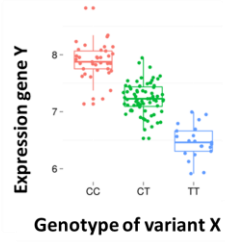
Genetic control of gene transcription



15

Genetic variant modulating expression levels





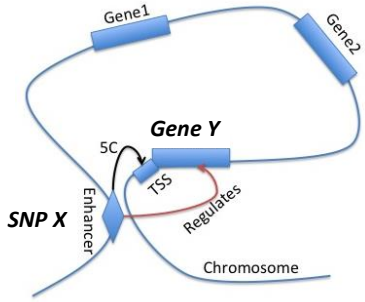
Expression quantitative trait locus
(eQTL)
=
in silico association between genotype and gene expression level within a specific population

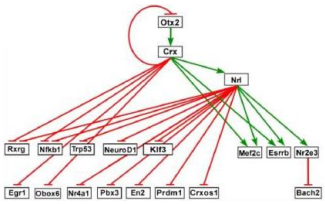
- Method: Linear regression (*GenABEL, MATRIXEQTL*)
- *cis* (= local) effects focused (sample size)

16

What are *cis* & *trans* eQTLs

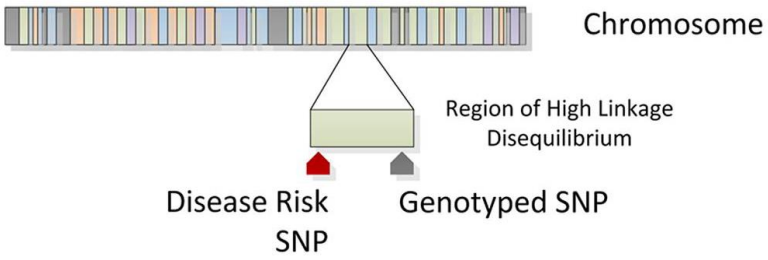
- *trans* eQTL: **SNP X** with **Gene Y**
  - **SNP X** not within 1 megabase of **Gene Y**
  - **SNP X** and **Gene Y** on different chromosomes
- **Distant interactions**
  - **SNP X** could be in a distant regulatory element (interactions between chromosomes)
  - **SNP X** linked to a transcription factor
- Expect small effect sizes → power issues in all but the largest studies





17

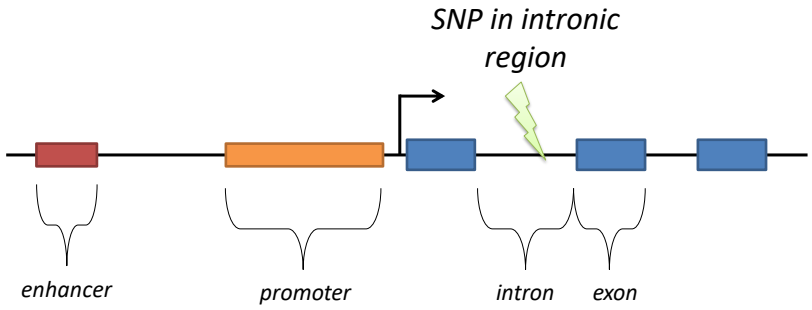
Linkage disequilibrium and eQTLs



- LD = the non-random association of alleles at different loci (i.e. $\rho_{AB} \neq \rho_A \rho_B$)
  - *Often calculated as the square of correlation coefficient: r^2*
  - *Often visualized in GWAS Manhattan plots*
- Indirect association due to LD structure: an eQTL SNP may or may not be the causal SNP

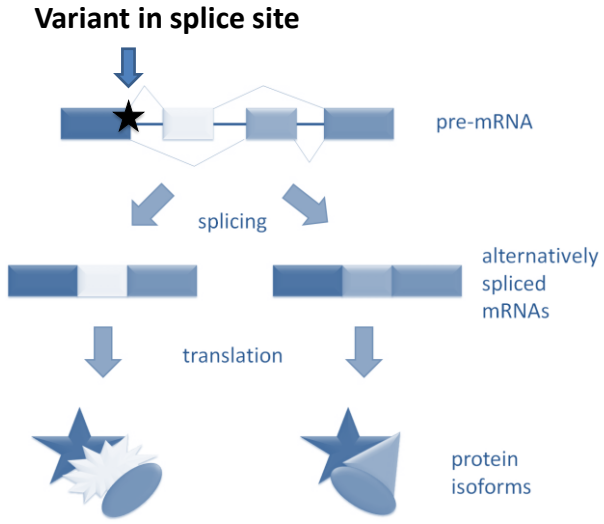
18

Genetic control of gene transcription



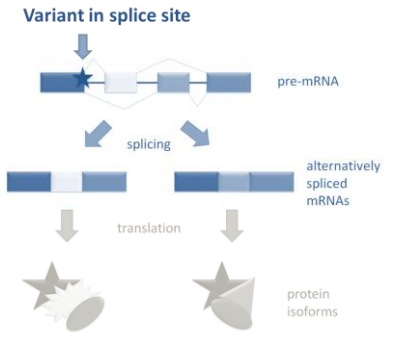
19

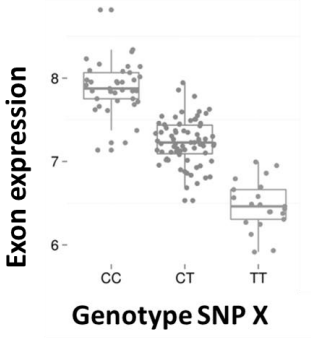
Genetic variants regulate exon usage



20

Genetic variants regulate exon usage





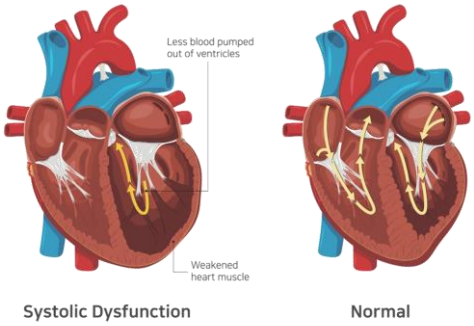
Splicing quantitative trait locus
(sQTL)
=
in silico association between genotype and alternative splicing within a specific population

21

Research: genetics of transcription and splicing in DCM

Samples: Left ventricle
- **108** non-diseased donor hearts
- **97** dilated cardiomyopathy (DCM) hearts

Data:
- RNA-seq: 16,219 unique mRNA levels
- Genotyping: 2 million common variants (SNPs)



Adriaens, Koopmann et al. (2014)
Heinig, Adriaens, Schaefer et al. (2017)

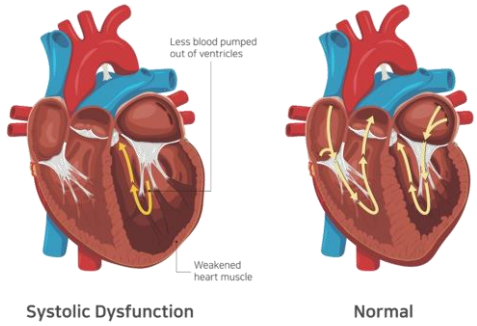
22

Research: genetics of transcription and splicing in DCM

Samples: Left ventricle
- **108** non-diseased donor hearts
- **97** dilated cardiomyopathy (DCM) hearts

Data:
- RNA-seq: 16,219 unique mRNA levels
- Genotyping: 2 million common variants (SNPs)

Research questions:
- **Which variants modulate gene expression? (eQTL)**
- **Which variants modulate splicing? (sQTL)**
- **Do these differ between DCM and controls?**

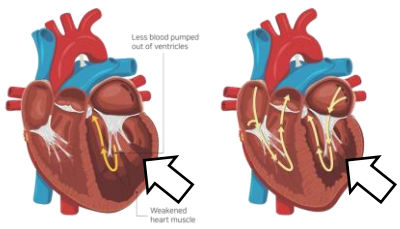


Adriaens, Koopmann et al. (2014)

Heinig, Adriaens, Schaefer et al. (2017)

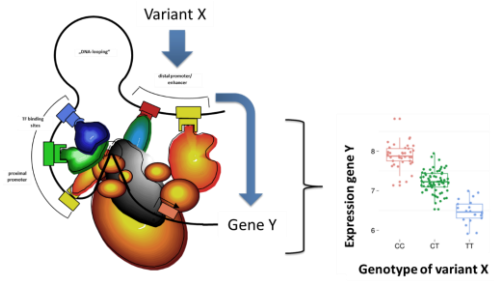
23

Research: genetics of transcription and splicing in DCM



Transcriptomics

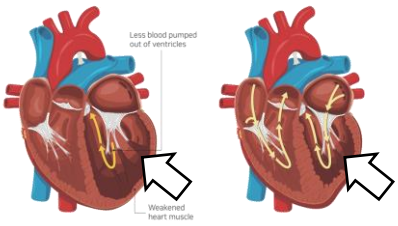
Genomics



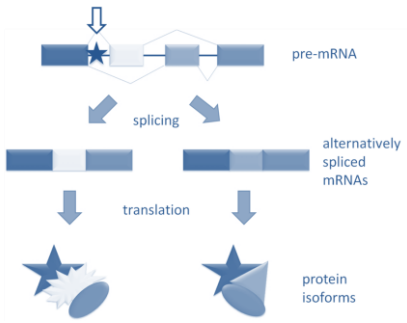
Heinig, Adriaens, Schaefer et al. (2017)

24

Research: genetics of transcription and splicing in DCM



Transcriptomics
Genomics

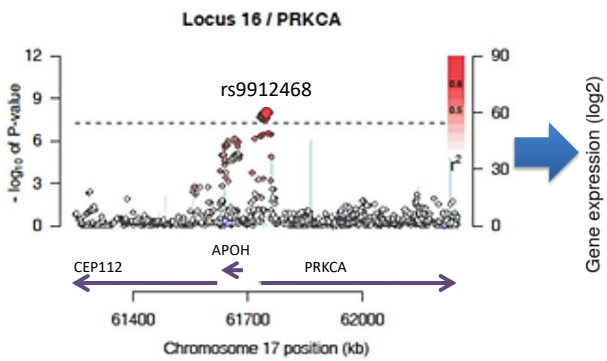


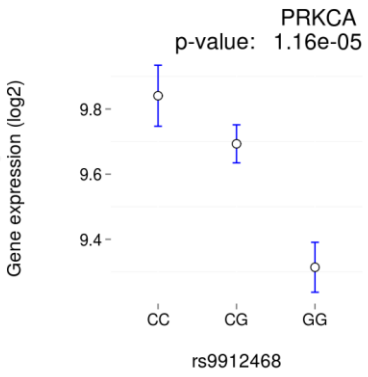
Differential splicing and expression contributes to the DCM phenotype and is genetically controlled for known and novel genes

Heinig, Adriaens, Schaefer et al. (2017)

25

Usage example: eQTLs for known GWAS loci





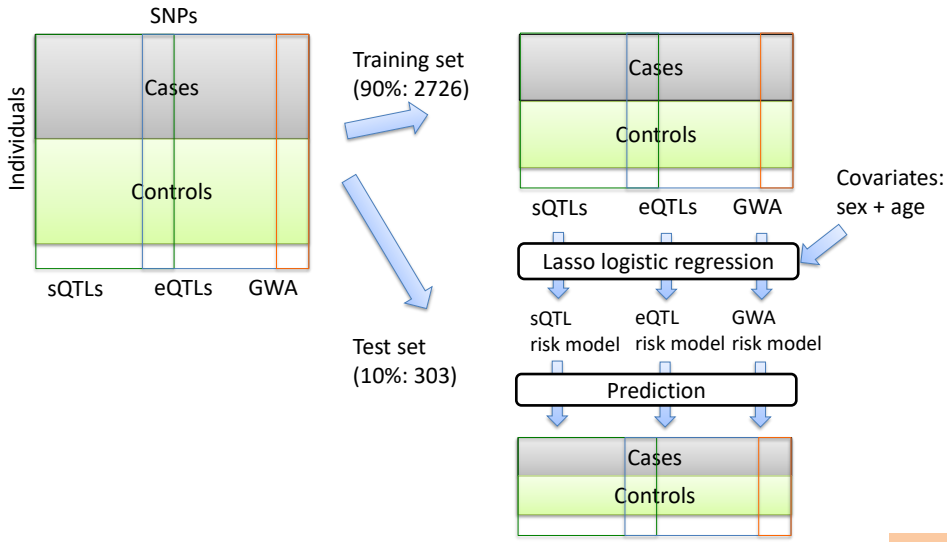
GWAS result from Sotoodehnia et al. (2010)

- ***rs9912468***: associated with QRS prolongation (effect allele = G)
- **Protein kinase C alpha**: regulator of cardiac contractility and Ca^{2+} handling in myocytes

Adriaens, Koopmann et al. (2014)

26

DCM genetic risk prediction



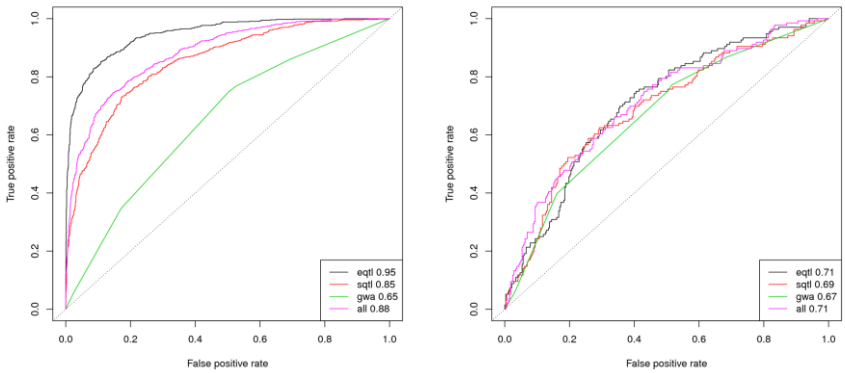
GWAS data from: Meder et al. Eur Heart J 2014

27

DCM genetic risk prediction

Training set (90%)

Test set (10%)



LASSO (glmnet  *)*

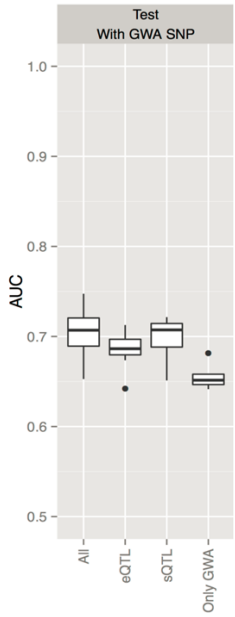
Heinig, Adriaens, Schaefer et al. (2017)

28

DCM genetic risk prediction

Better prediction

Random prediction



Combining
1. Co-variates (age, sex)
2. Genotype of DCM GWA SNP (rs9262636)
3. Genotypes of SNPs modulating expression (eQTLs)
4. Genotypes of SNPs modulating splicing (sQTLs)

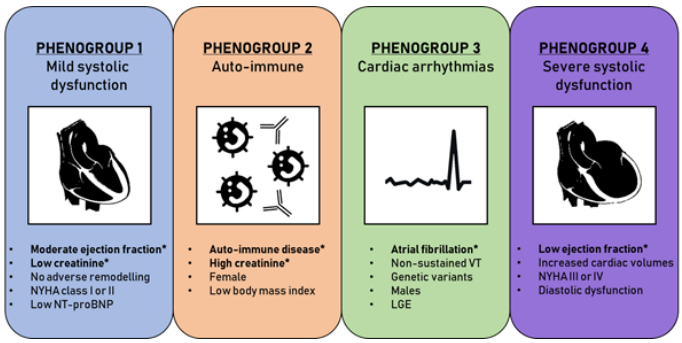
In single predictive model leads to better prediction

Heinig, Adriaens, Schaefer et al. (2017)

29

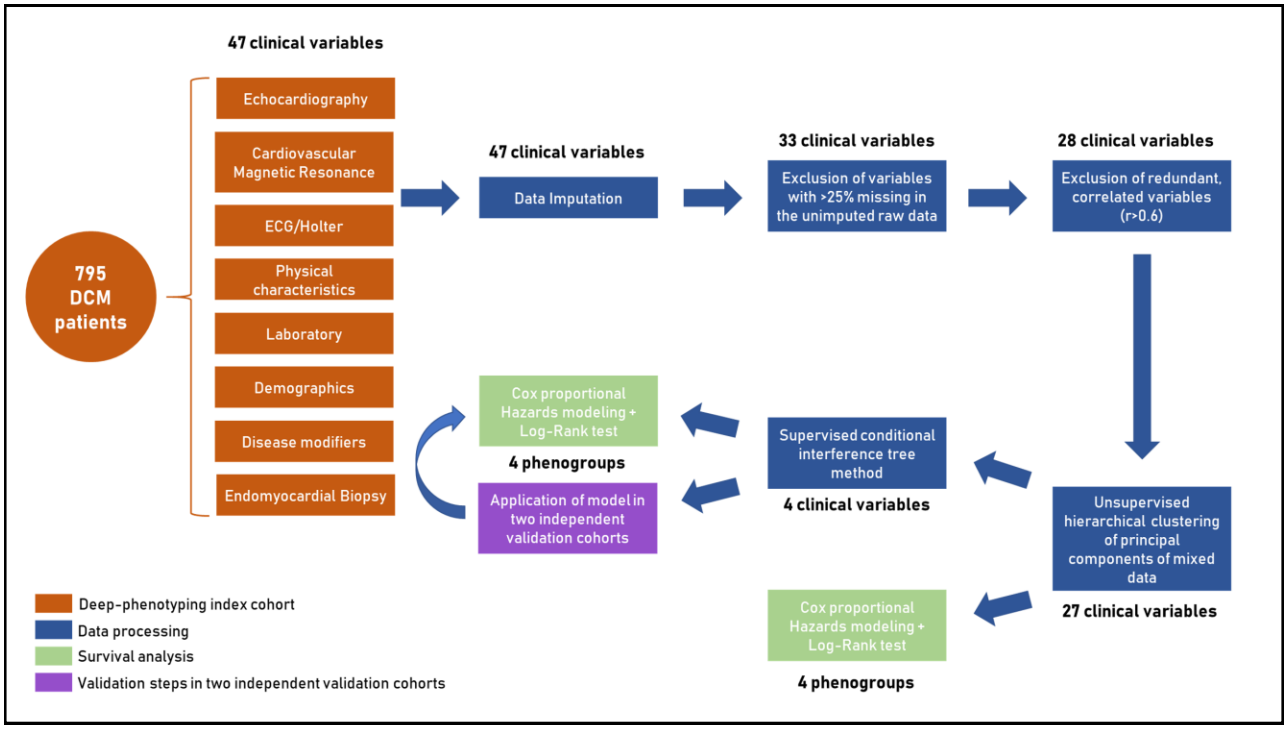
Research: DCM cohort in Maastricht

- Many clinical parameters available for more extensive subtyping:
  - Machine learning resulted in 4 distinct phenotypic clusters ("phenogroups")
- **Questions:**
  - **Which genes show differences in eQTLs and sQTLs between phenogroups?**
  - **In which processes and pathways are the corresponding genes involved?**
- Using RNA-seq of EMBs (n = 76)

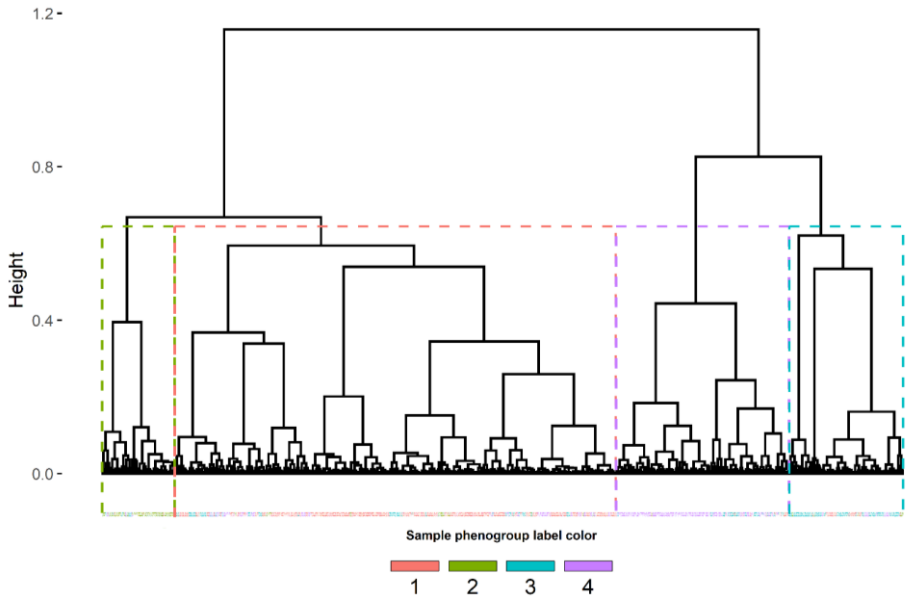


Verdonschot et al. (2020)

30


47 clinical variables
Echocardiography
Cardiovascular Magnetic Resonance
ECG/Holter
Physical characteristics
Laboratory
Demographics
Disease modifiers
Endomyocardial Biopsy
795 DCM patients
47 clinical variables
Data Imputation
33 clinical variables
Exclusion of variables with >25% missing in the unimputed raw data
28 clinical variables
Exclusion of redundant, correlated variables (r>0.6)
Unsupervised hierarchical clustering of principal components of mixed data
27 clinical variables
Supervised conditional interference tree method
4 clinical variables
Cox proportional Hazards modeling + Log-Rank test
4 phenogroups
Application of model in two independent validation cohorts
Cox proportional Hazards modeling + Log-Rank test
4 phenogroups
Deep-phenotyping index cohort
Data processing
Survival analysis
Validation steps in two independent validation cohorts

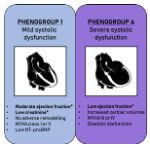
31


1.2
0.8
0.4
0.0
Height
Sample phenogroup label color
1
2
3
4

32

Severe versus mild systolic dysfunction

- 96 unique genes that are significantly differentially imbalanced between phenogroup 4 and 1
- Gene Ontology enrichment analysis:

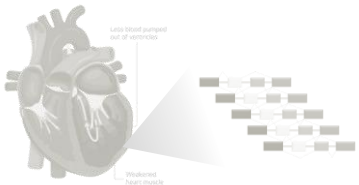


Term	***P-value***
cyclosporin A binding	6.00E-04
muscle structure development	9.60E-04
establishment of protein localization to membrane	1.15E-03
negative regulation of oxidative phosphorylation	1.44E-03
electron transport chain	9.35E-03
fat cell differentiation	1.05E-02
regulation of actin filament-based movement	1.50E-02
cellular response to stress	1.68E-02
response to calcium ion	1.70E-02
mitochondrial respiratory chain complex assembly	1.76E-02

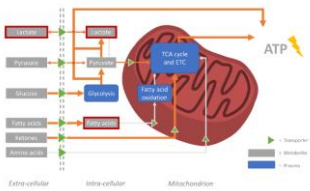
33

Outline

1. Genetic control of gene transcription in heart failure

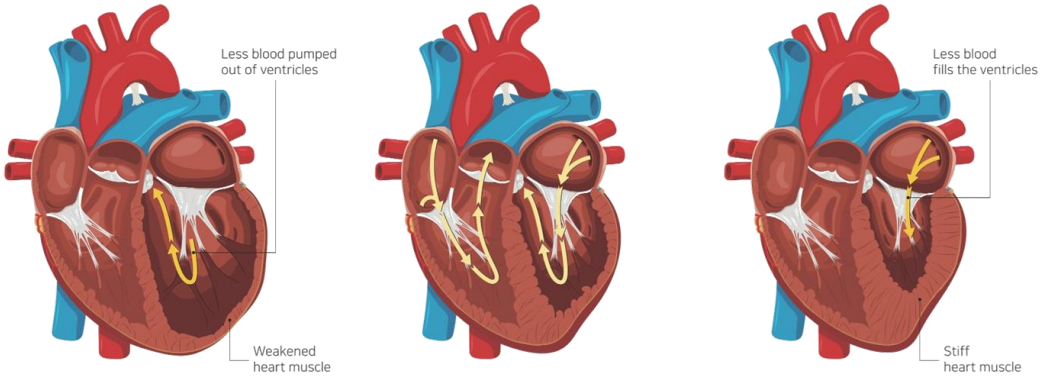


2. Modeling the metabolism of the failing heart



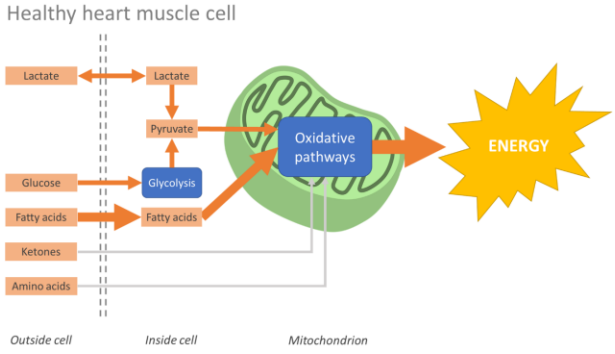
34

Recap: what is heart failure?



35

Loss of metabolic flexibility in DCM



Healthy heart

36

Restoring metabolic flexibility?

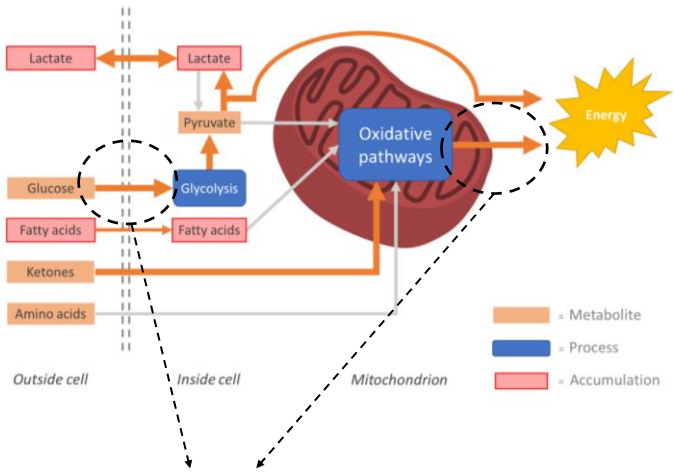
- Clinical trials aimed at restoring metabolic flexibility have so far led to mixed results
- Patient-to-patient differences are currently poorly understood

  → Targeted metabolic therapies have therefore not seen clinical implementation yet

37

Diagnosing loss of metabolic flexibility

- To diagnose, we need to determine **metabolic fluxes**
  - **Fluxomics: reaction fluxes of all known metabolic reactions**
  - Identify which pathways differ between patients
- Ideally: *in vivo* tracer studies to measure metabolic fluxes:
  - Problem 1: expensive and low sensitivity
  - Problem 2: some impairments only appear under stress

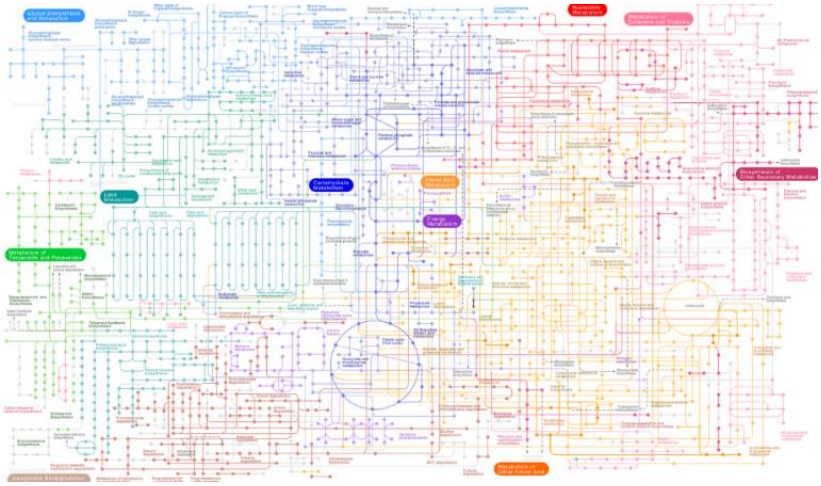


How to quantify? How to place in context?

38

Genome-scale metabolic model (GEM)

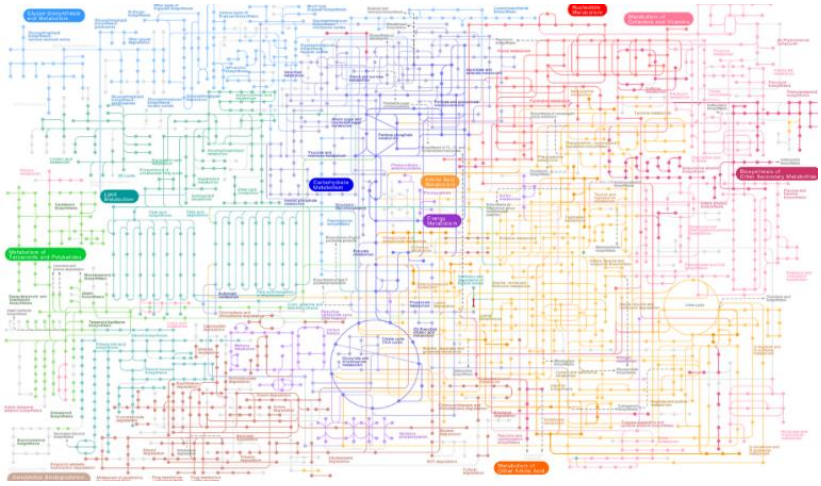
- Contains all know metabolic reactions including:
  - Transport reactions
  - Enzymatic reactions
- Derived from existing knowledge:
  - Pathway databases
  - Literature
- Creating and curating such a network is a lot of work:
  - Only a few dedicated groups world-wide



39

Genome-scale metabolic model (GEM)

- Metabolites
- Reactions
- Stoichiometry
- Directionality of reactions (thermodynamic considerations)
- Enzymes (catalysing a reaction)
- Gene-Protein-Reaction rules
- Mass- and charge-balance of reactions
- Compartmentalisation



40

Genome-scale metabolic model (GEM)

Metabolites

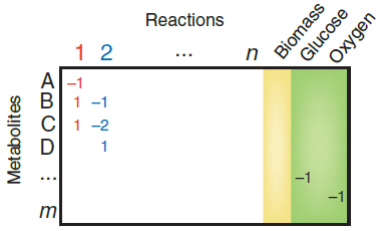
Reactions

Stoichiometry

Reaction: HEX1: hexokinase

ATP	D-glucose	proton	ADP	D-glucose 6-phosphate
atp[c]	+ glc_D[c]	→ h[c]	+ adp[c]	+ g6p[c]
-1	-1	1	1	1

Stoichiometric matrix S:



41

Genome-scale metabolic model (GEM)

Metabolites

Reactions

Stoichiometry

Reaction: HEX1: hexokinase

ATP	D-glucose	proton	ADP	D-glucose 6-phosphate
atp[c]	+ glc_D[c]	→ h[c]	+ adp[c]	+ g6p[c]
-1	-1	1	1	1

c = cytoplasm	*e = extracellular space*
g = Golgi apparatus	*l = lysosome*
m = mitochondrion	*n = nucleus*
r = endoplasmic reticulum	*x = peroxisome*

42

Genome-scale metabolic model (GEM)

- Metabolites
- Reactions
- Stoichiometry

Reaction: HEX1: hexokinase

ATP	D-glucose	proton	ADP	D-glucose 6-phosphate
atp[c]	+ glc_D[c]	→ h[c]	+ adp[c]	+ g6p[c]
-1	-1	1	1	1

Flux of reaction
- *has upper and lower bound*
- *often expressed in mmol/gDW/s*
- *gDW = gram dry weight*

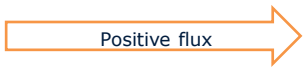
43

Genome-scale metabolic model (GEM)

- Metabolites
- Reactions
- Stoichiometry

Reaction: HEX1: hexokinase

ATP	D-glucose	proton	ADP	D-glucose 6-phosphate
atp[c]	+ glc_D[c]	→ h[c]	+ adp[c]	+ g6p[c]
-1	-1	1	1	1



Flux of reaction
- *has upper and lower bound*
- *often expressed in mmol/gDW/s*
- *gDW = gram dry weight*

44

Genome-scale metabolic model (GEM)

- Metabolites
- Reactions
- Stoichiometry

Reaction: HEX1: hexokinase

ATP	D-glucose	proton	ADP	D-glucose 6-phosphate
atp[c] +	glc_D[c] →	h[c] +	adp[c] +	g6p[c]
-1	-1	1	1	1

Negative flux

Flux of reaction
- *has upper and lower bound*
- *often expressed in mmol/gDW/s*
- *gDW = gram dry weight*

45

Genome-scale metabolic model (GEM)

- Metabolites
- Reactions
- Stoichiometry
- Directionality of reactions (thermodynamic considerations)
- Enzymes (catalysing a reaction)
- Gene-Protein-Reaction rules
- Mass- and charge-balance of reactions
- Compartmentalisation

Reaction: HEX1: hexokinase

ATP	D-glucose	proton	ADP	D-glucose 6-phosphate
atp[c] +	glc_D[c] →	h[c] +	adp[c] +	g6p[c]
-1	-1	1	1	1

→ vs. ↔, irreversible vs. reversible

Hexokinase 1, 2, 3, or 4 (glucokinase) catalyze the reaction

(3098) or (3099) or (3101) or (2645)…

Gene number for hexokinase 1

glc_D[e] <=> glc_D[c]

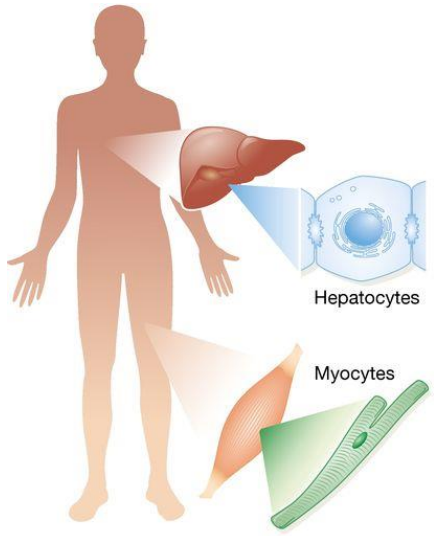
Glucose transport from extracellular space to cytosol

46

The aim of a model is context specific

- GEMs are often organism-specific, but not tissue/cell type specific
- Tissue-specific models include only reactions that are active in the respective tissue
- Two "static" omics types can inform this modeling process:
  - Transcriptomics
  - Metabolomics

- *Transcriptomics: reaction is inactive if catalyzing enzyme is not expressed*
- *Metabolomics: reaction is inactive if the product is not present*



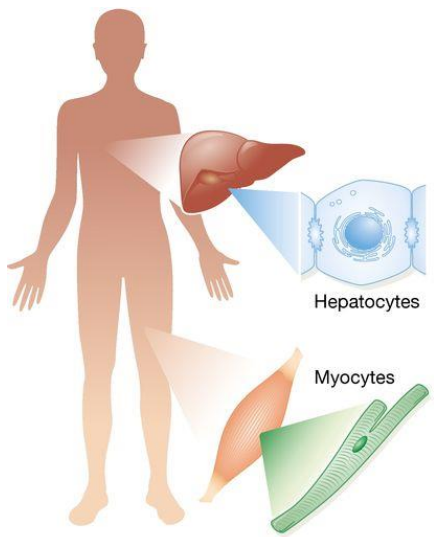
Uhlen, M et al. Mol Sys Bio, 12:862 (2016)

47

The aim of a model is context specific

- GEMs are often organism-specific, but not tissue/cell type specific
- Tissue-specific models include only reactions that are active in the respective tissue
- Two "static" omics types can inform this modeling process:
  - Transcriptomics
  - Metabolomics

- *Transcriptomics: reaction is inactive if catalyzing enzyme is not expressed*
- *Metabolomics: reaction is inactive if the product is not present*

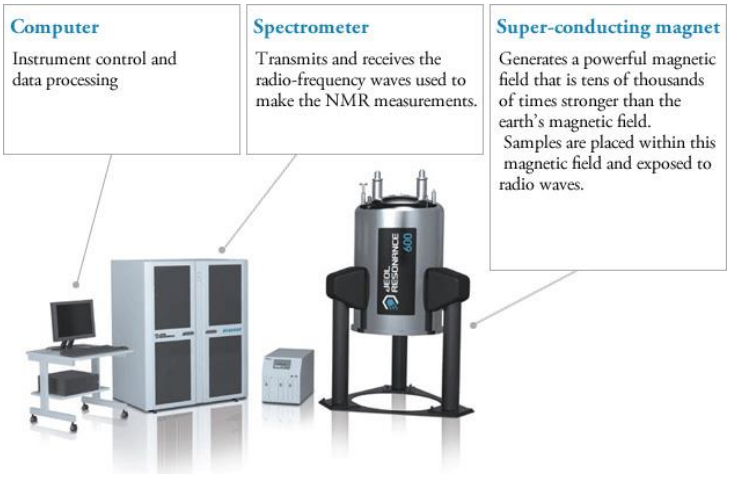


Uhlen, M et al. Mol Sys Bio, 12:862 (2016)

48

NMR metabolomics

- NMR is an abbreviation for Nuclear Magnetic Resonance:
  - allows the molecular structure of a sample to be analyzed by observing and measuring the interaction of nuclear spins when placed in a powerful magnetic field

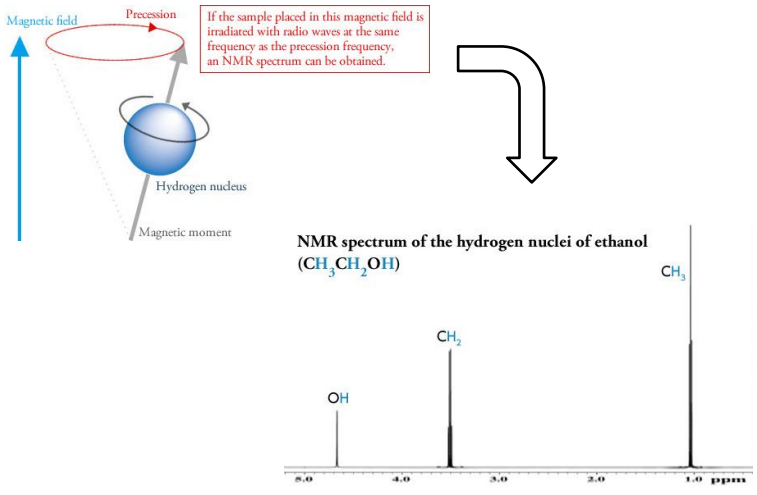


https://www.jeol.co.jp/en/products/nmr/basics.html

49

NMR metabolomics

- NMR is an abbreviation for Nuclear Magnetic Resonance:
  - allows the molecular structure of a sample to be analyzed by observing and measuring the interaction of nuclear spins when placed in a powerful magnetic field
- When a nucleus that possesses a magnetic moment is placed in a strong magnetic field, it will begin to precess, like a spinning top

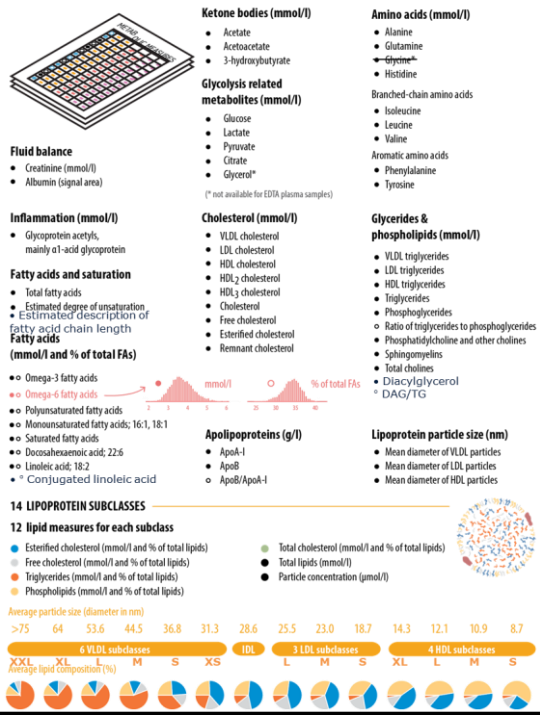


Known "spectral fingerprint" → identify molecule(s)

50

NMR metabolomics

- Nightingale's technology utilizes NMR and proprietary software to provide metabolome profiles
  - Consists of 231 metabolites
  - Quite low compared to total number of known metabolites: 220945
    - We measure only 0.1%!

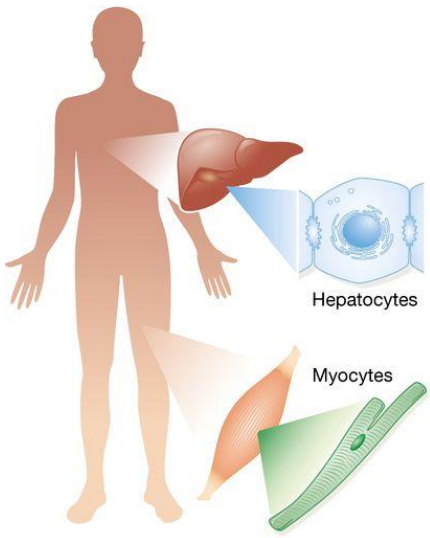


51

The aim of a model is context specific

- GEMs are often organism-specific, but not tissue/cell type specific
- Tissue-specific models include only reactions that are active in the respective tissue
- Two "static" omics types can inform this modeling process:
  - Transcriptomics
  - Metabolomics

- *Transcriptomics: reaction is inactive if catalyzing enzyme is not expressed*
- *Metabolomics: reaction is inactive if the product is not present*
  - *But we can only measure 0.1% of the metabolome*
  - *So mostly useful for validating excretion to the circulation*

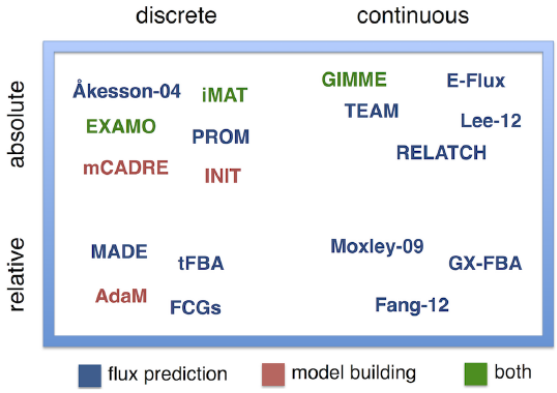


Uhlen, M et al. Mol Sys Bio, 12:862 (2016)

52

Model extraction methods (MEMs)

- Many algorithms have been proposed for building tissue-specific models based on generic models
- Simplest approach: delete reactions of genes that are not expressed
  - Typically based on tissue-specific transcriptomics data
- Problems:
  - Cutoff for being not expressed
  - Orphan reactions & dead-ends
  - Need to check metabolic functions

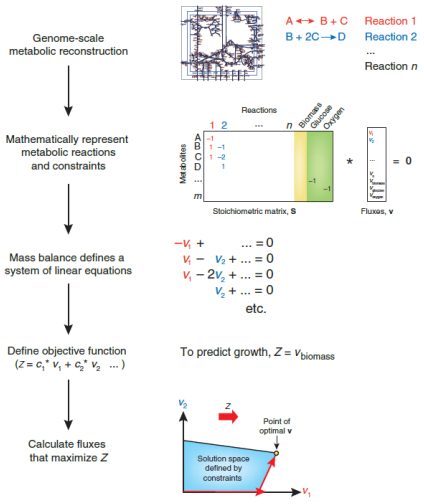


Machado, D et al. PLoS Comp Bio, 10(4):e1003580 (2014)

53

Flux balance analysis

- Used to calculate flow of metabolites through metabolic network
- Predict growth rate of organism or rate of production of given metabolite
- Assumes steady state
- Optimizes a given objective function

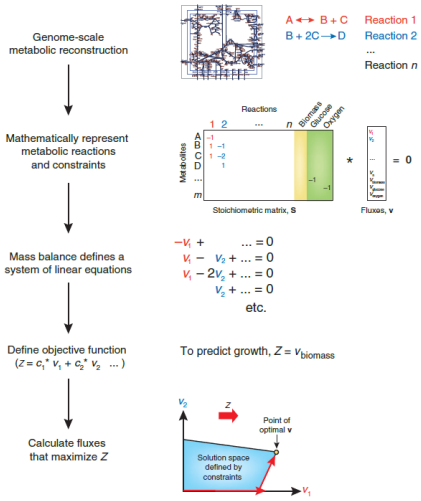


Orth JD, et al. Nat Biotech, 28(3):245-8 (2010)

54

Flux balance analysis

- Used to calculate flow of metabolites through metabolic network
- Predict growth rate of organism or rate of production of given metabolite
- Assumes steady state
- Optimizes a given objective function



Orth JD, et al. Nat Biotech, 28(3):245-8 (2010)

55

Objective function - Examples

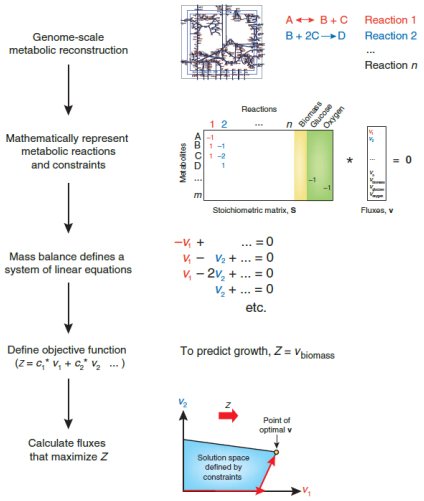
The objective function ≈ the aim of the model

- Biomass reaction (e.g. plants for consumption)
- ATP production (ATP demand reaction)
- Maximize a product of interest (e.g. lysine production)
- …

56

Flux balance analysis

- Used to calculate flow of metabolites through metabolic network
- Predict growth rate of organism or rate of production of given metabolite
- Assumes steady state
- Optimizes a given objective function

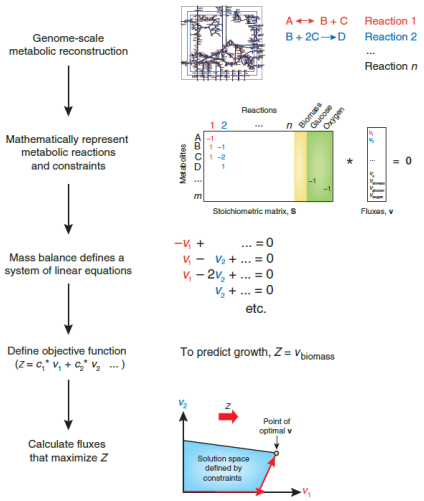


Orth JD, et al. Nat Biotech, 28(3):245-8 (2010)

57

Flux balance analysis

- Used to calculate flow of metabolites through metabolic network
- Predict growth rate of organism or rate of production of given metabolite
- Assumes steady state
- Optimizes a given objective function



Orth JD, et al. Nat Biotech, 28(3):245-8 (2010)

58

Steady-state assumption

Assumption to reduce model complexity:

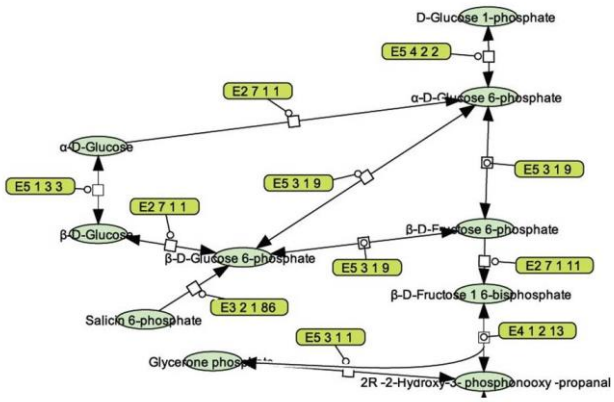
Metabolite concentrations and reaction rates stay constant over time (steady-state)

Benefit:

1. We have to estimate only one value (reaction rate/flux) per reaction instead of a function over time
2. We do not have to care about different metabolite concentrations
3. Introduces a direct dependence between reactions: Production and consumption of each metabolite cancel out

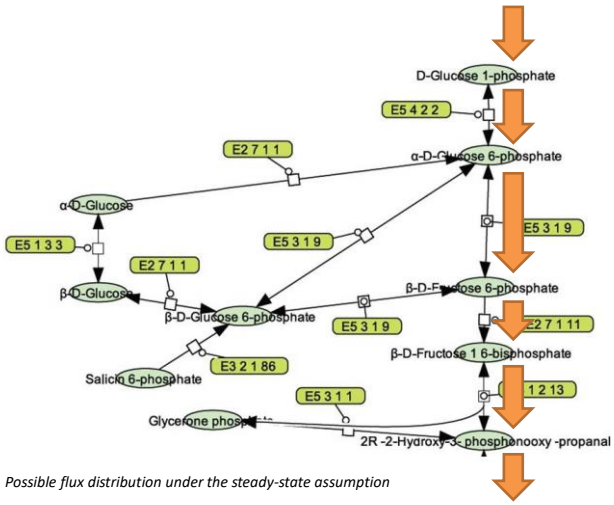
59

Steady-state assumption visualized



60

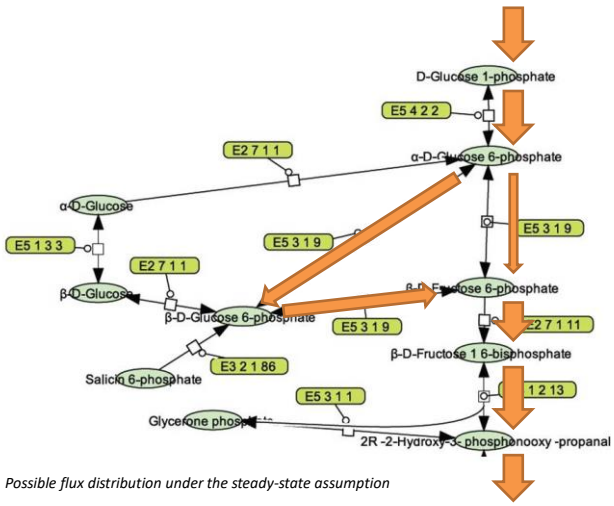
Steady-state assumption visualized



Possible flux distribution under the steady-state assumption

61

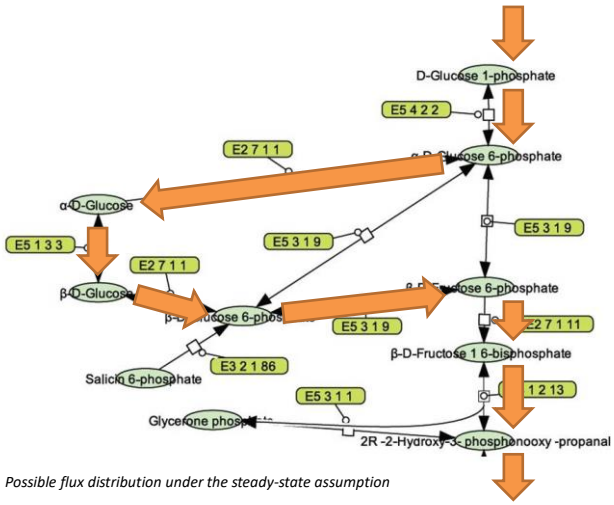
Steady-state assumption visualized



Possible flux distribution under the steady-state assumption

62

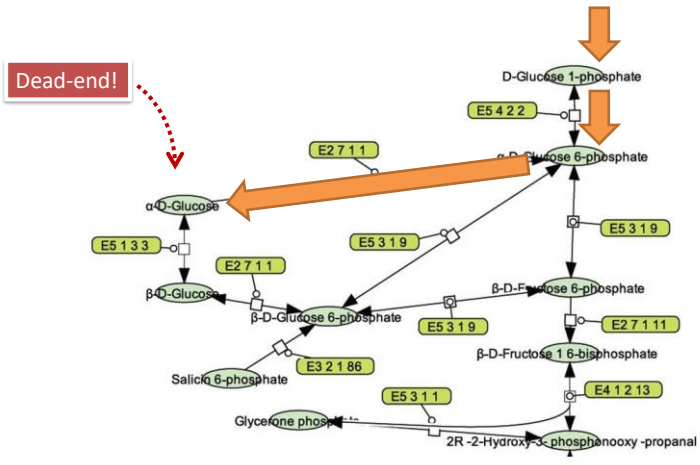
Steady-state assumption visualized



Possible flux distribution under the steady-state assumption

63

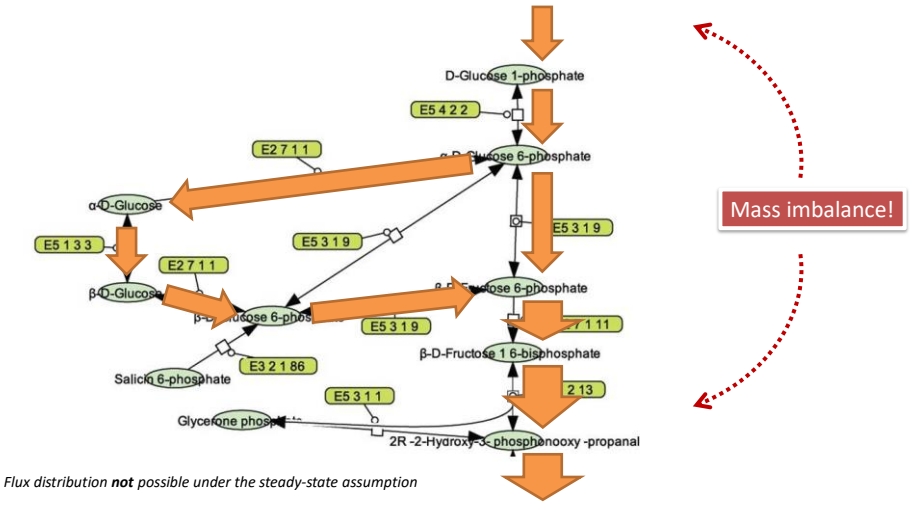
Steady-state assumption visualized



*Flux distribution **not** possible under the steady-state assumption*

64

Steady-state assumption visualized



*Flux distribution **not** possible under the steady-state assumption*

65

Steady-state assumption visualized

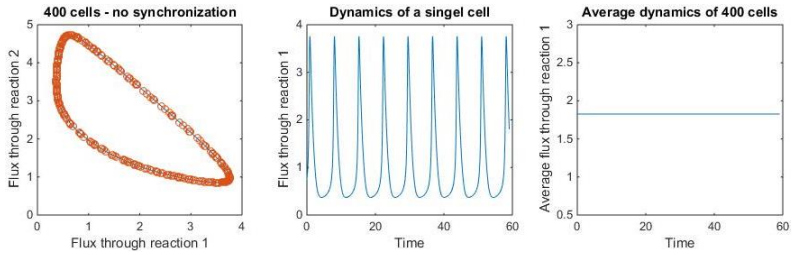
Like a water supply network!

- *Some places might use more than others*
- *But what goes in, must come out*

66

Validity of the steady state assumption

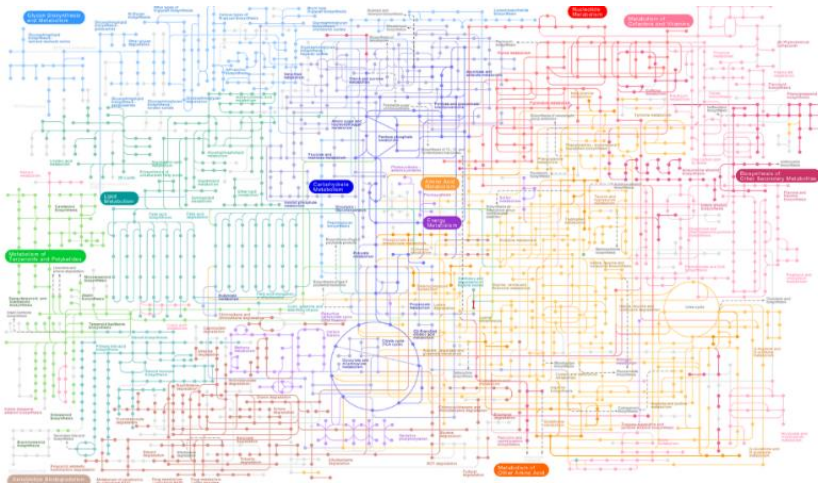
- Cyclic behavior (e.g. limit cycles/periodic fixed points)
- No steady state for single cells
- Consider average of many cells (no synchronization) → steady state reasonable



67

Modeling loss of metabolic flexibility

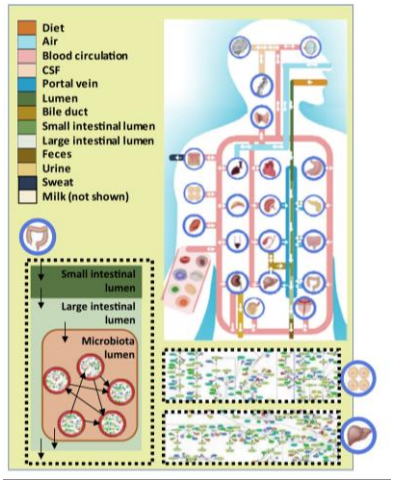
- Genome-scale metabolic models



68

Metabolic model and patient data

- Model: heart model from Harvey and Harvetta
  - Sex specific whole-body metabolic reconstructions
  - 26 organs and 6 blood cell types
  - Organ resolved → isolated organ models available as well
- Data: left ventricular RNA-seq data from MAGNet consortium
  - DCM, HCM, PPCM and healthy controls
- Constrain influx of substrates:
  - "Fed" a standard western-European diet

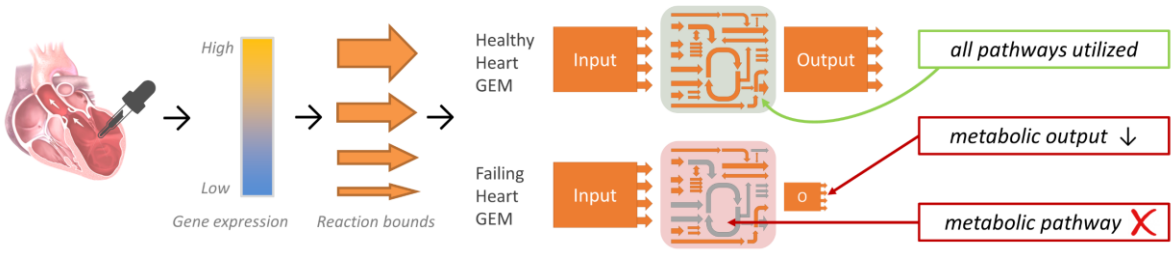


Transcriptomics: MAGNet (GSE141910)

Figure 1B, Thiele et al. 2020

69

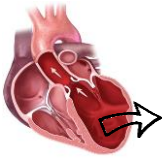
Methodological setup



70

Modeling loss of metabolic flexibility

- Genome-scale metabolic models
- Activate and deactivate reactions based on **gene activity**

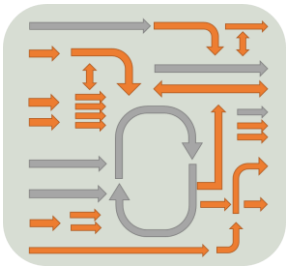


Measure gene activity as proxy for metabolic enzyme activity

Active gene =

Inactive gene =

Personalized model

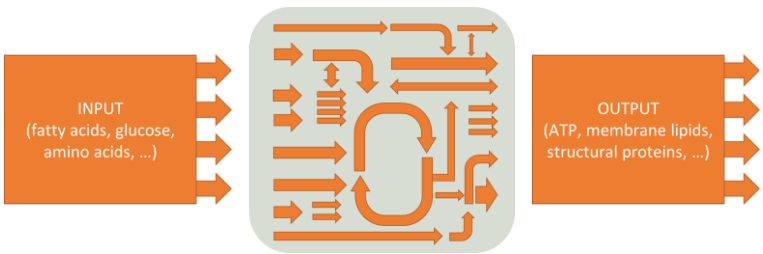


71

Modeling loss of metabolic flexibility

- Genome-scale metabolic models
- Activate and deactivate reactions based on **gene activity**
- Simulate metabolism for individual
  - Choose objective
  - Find optimal combination of fluxes to maximize objective
  - High flux pathway = active
  - Low flux pathway = less active

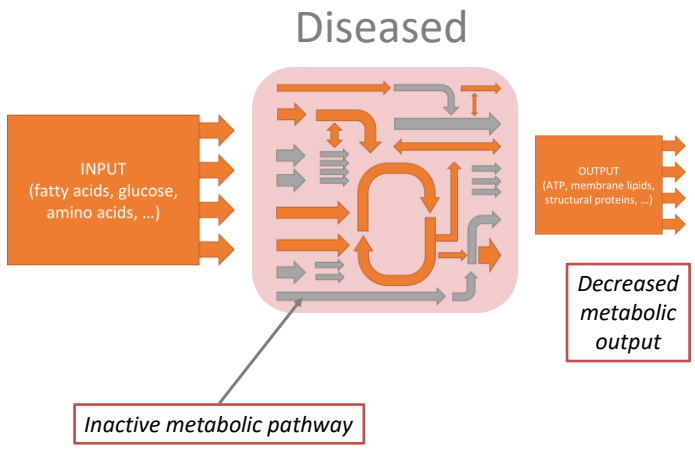
Healthy



72

Modeling loss of metabolic flexibility

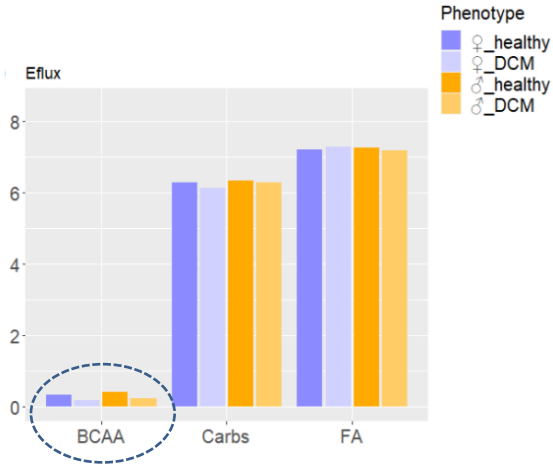
- Genome-scale metabolic models
- Activate and deactivate reactions based on **gene activity**
- Simulate metabolism for individual
  - Choose objective
  - Find optimal combination of fluxes to maximize objective
  - High flux pathway = active
  - Low flux pathway = less active



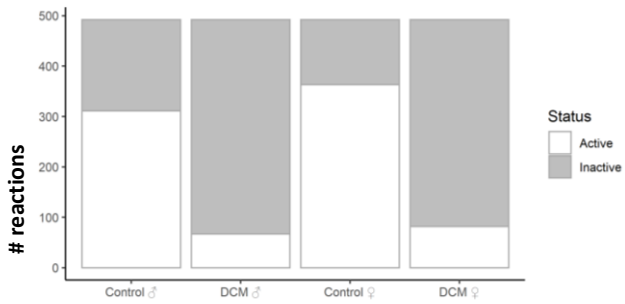
73

Pilot results: models capture broad trends

Branched chain amino-acids contribute less to mitochondrial energy production in DCM



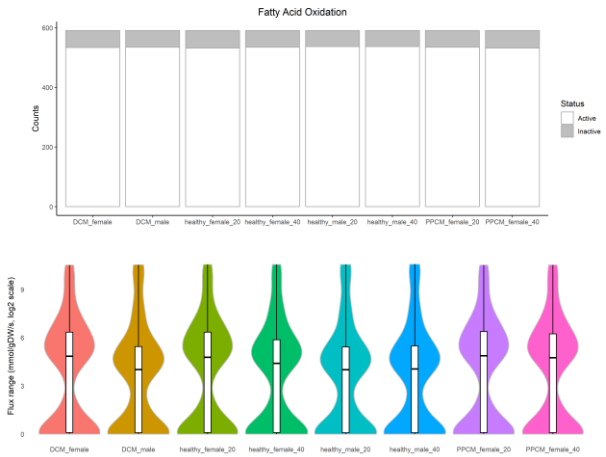
The fatty acid oxidation pathways have many inactive reactions in DCM



74

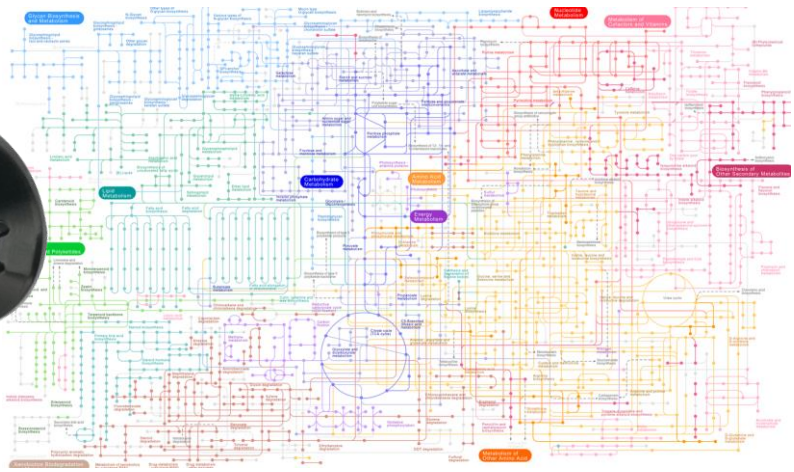
Pilot results: models capture sex differences

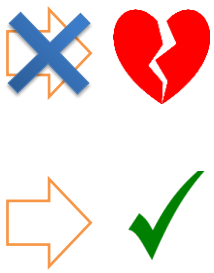
- Higher activity of fatty acid oxidation pathways in female heart compared to male heart
  - Also in DCM: cardioprotective?



75

Work in progress: test the effect of different diets

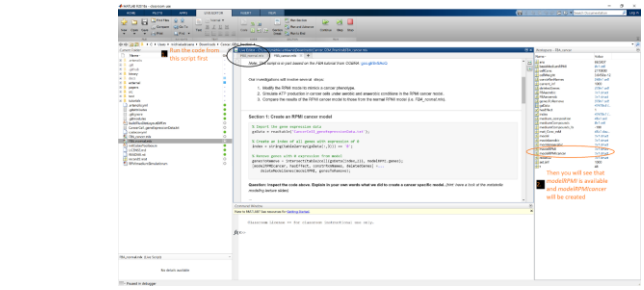


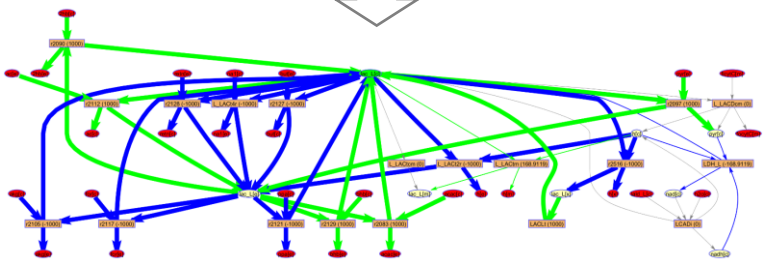


76

Software & tools

- Matlab
  - Python can be an alternative open-source solution for GEM analysis
- CobraToolbox
  - https://opencobra.github.io/cobratoolbox/stable/
  - Model extraction methods
    - Transcriptomics data integration
  - Flux balance analysis





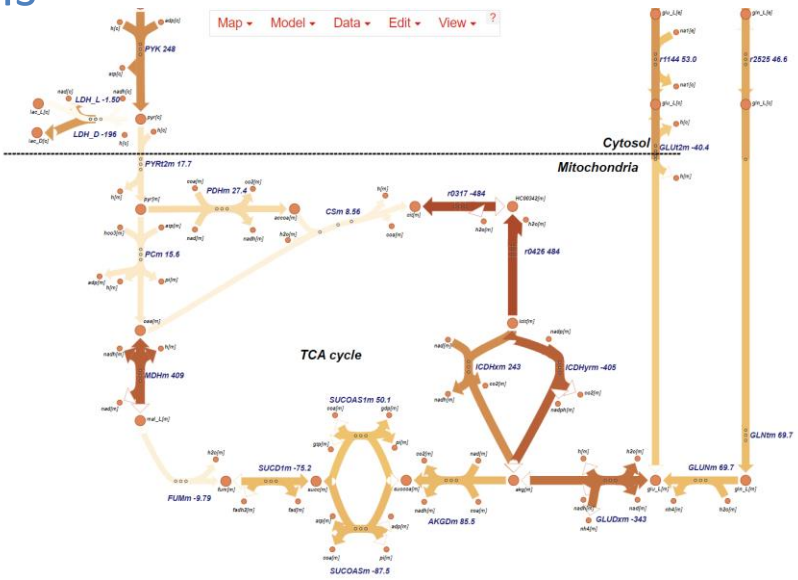
Aerobic glycolysis in cancer

77

Software & tools

- Escher maps:
  - Demo:

https://sbrg.github.io/escher-fba



Male DCM cardiomyocyte metabolism simulation

78

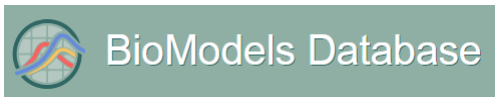
Advantages & limitations of GEMs

+ Relatively little information needed
+ Applicable to large networks
+ Quantitative flux estimations

- Only steady state estimation
- Often no unique solutions (large solution space)
- Optimization assumptions (FBA) critical

79

Model databases



https://www.ebi.ac.uk/biomodels-main/

https://vmh.uni.lu/

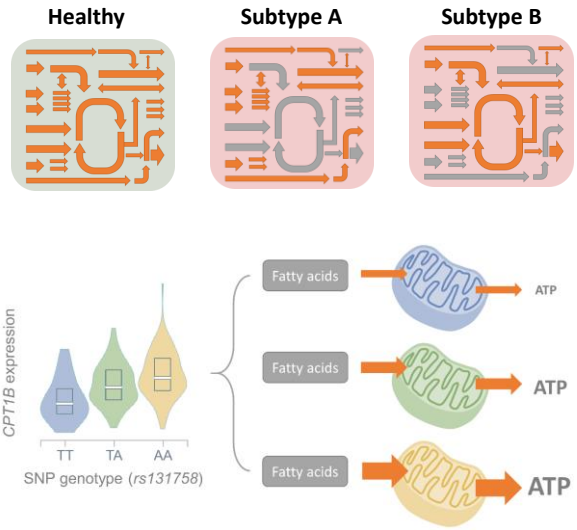
http://www.metabolicatlas.org/

http://bigg.ucsd.edu/

80

What's next?

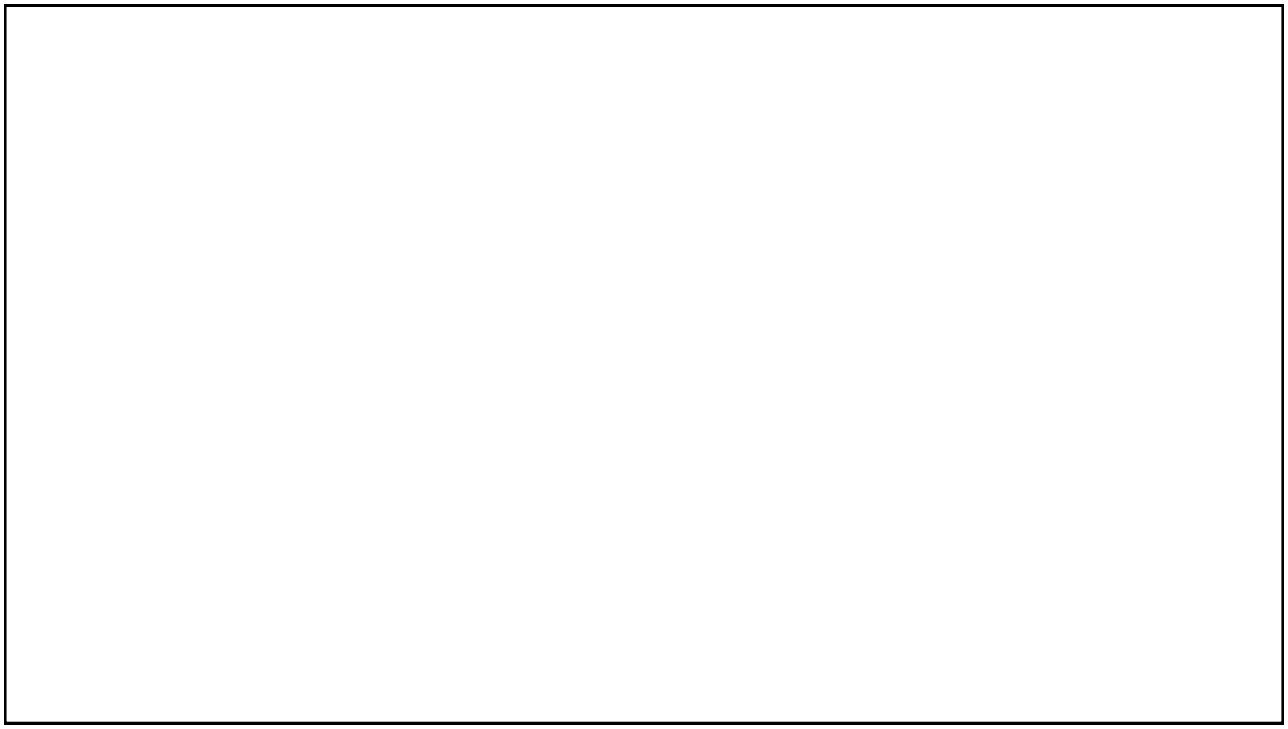
- Patient subtyping
  - Metabolic tasks (Dr. Marian Breuer)
    - Knowledge driven: which known metabolic tasks are active/inactive (qualitative & quantitative)
  - COMMET pipeline (Chaitra Sarathy)
    - Data-driven: which paths / reactions differ between individuals in entire network (quantitative)
- Integration of genetic regulation into models
  - Model effect of eQTLs
  - Genetic association to EFM / Principal components



81

michiel.adriaens@maastrichtuniversity.nl

82



83