Causally-cohesive genotype-phenotype (cGP) models

- systems biology meets genetics

Arne B. Gjuvsland

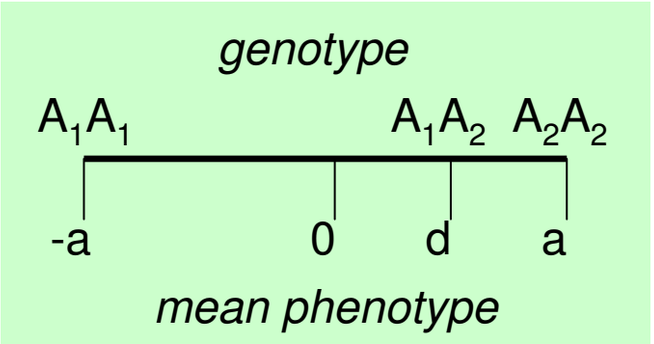
”Bioinformatics for molecular biology”
16.09.2009

Overview of talk

- Motivating and defining cGP models
- cGP models for gene regulatory networks
- Quantitative genetic analysis
  - Statistical analysis
  - Functional description
- Exploring the link systems biology and genetics
  - Simple gene regulatory network models
  - Systemic properties: feedback structure, gene regulation function
- Towards more complex cGP models
  - Preliminary analysis of two cGP models from public repositories

The genotype-phenotype (GP) map – two different views

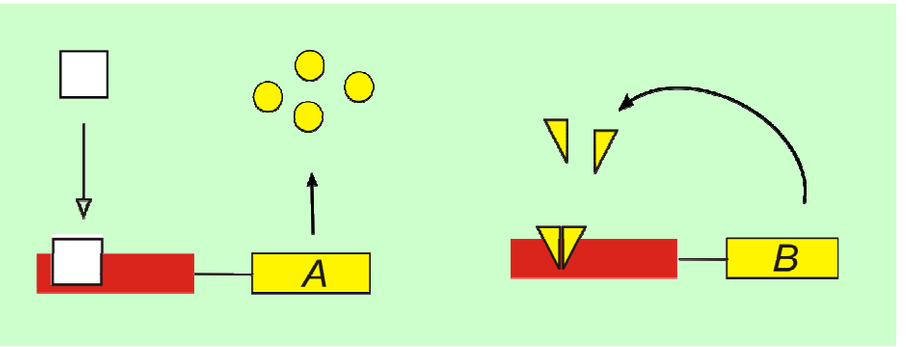
Quantitative genetics:



- d=0 : additive gene action
- |d|<a : partial dominance
- |d|=a : complete dom.
- |d|>a : overdominance

- mathematical GP map
- useful statistical machinery
  - production biology, medicine
  - QTL-methods

Regulatory biology:



- downstream gene (A)
- feedback regulation (B)
- activation or inhibition

- requires molecular insights
- biological GP map
- complex connection with classical gene action

Causally cohesive genotype-phenotype (cGP) models

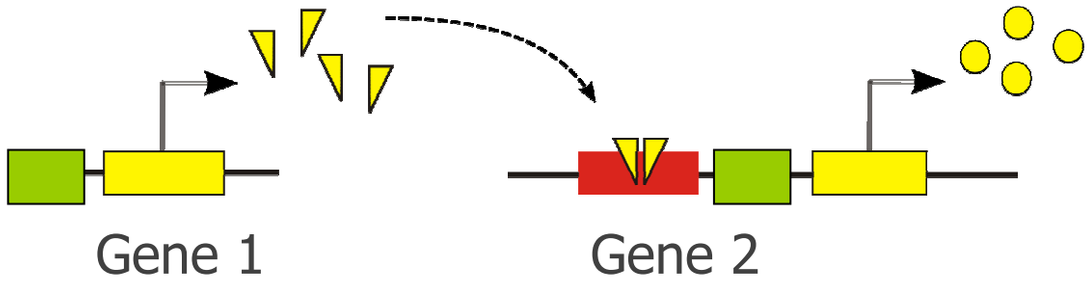
A mathematical model *M* of a biological system is a cGP model if:

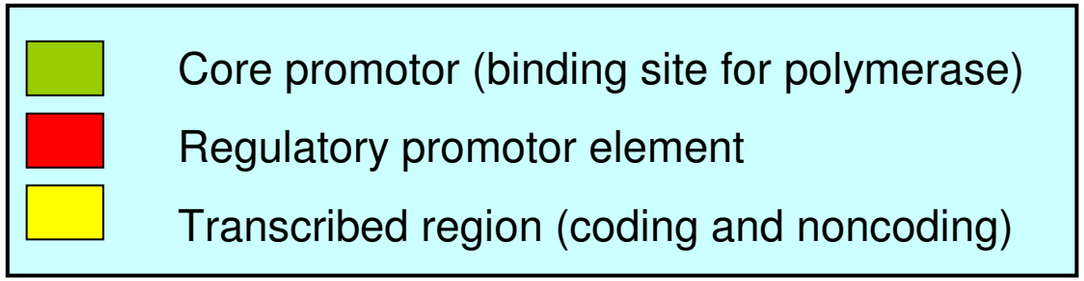
- Model elements (variables or parameters) are associated with genes
- Genotypic variation is represented by variation in a set of parameters
- It describes how phenotypes arise from lower level processes in a causally cohesive way

Defines a GP map $T_M : \mathrm{G} \rightarrow \mathrm{P}$ from a set G of genotype indexes to a set P of real-valued phenotypes.

How to build a cGP model – 1: Biological system

Consider a simple regulatory system of two genes:





Gene 1 is constitutively expressed and activates production of gene 2.

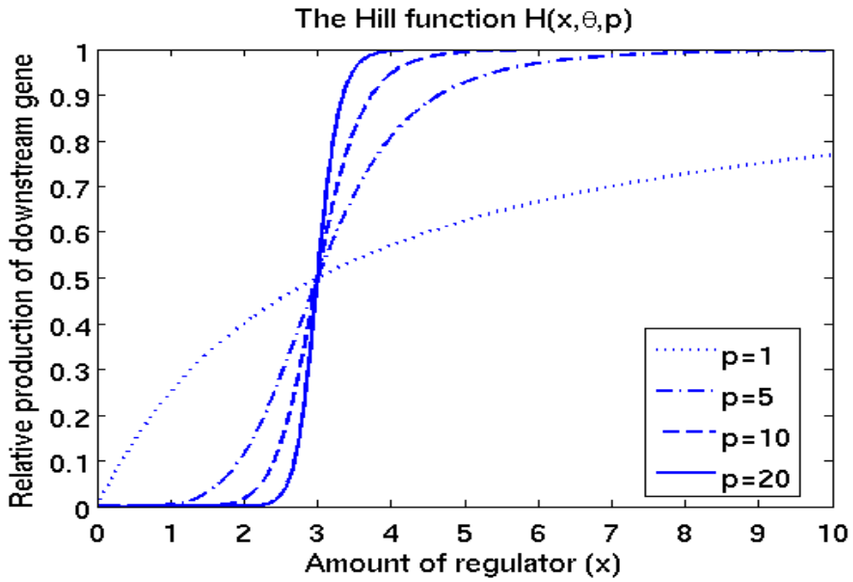
How to build a cGP model – 2: Mathematical model

- x_1 and x_2 denote expression levels of gene 1 and 2
- Time rate of change of x_i determined by two processes: production and decay

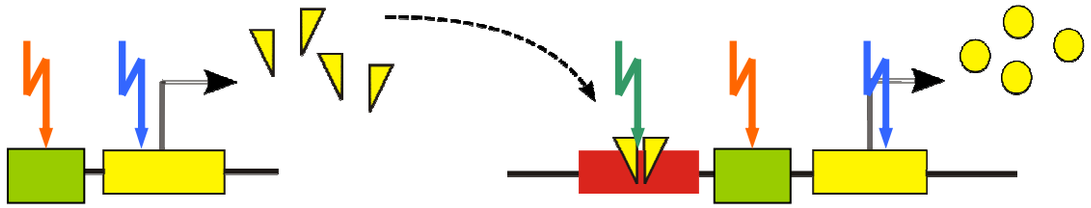
$$\frac{dx_1}{dt} = \alpha_1 - \gamma_1 x_1$$

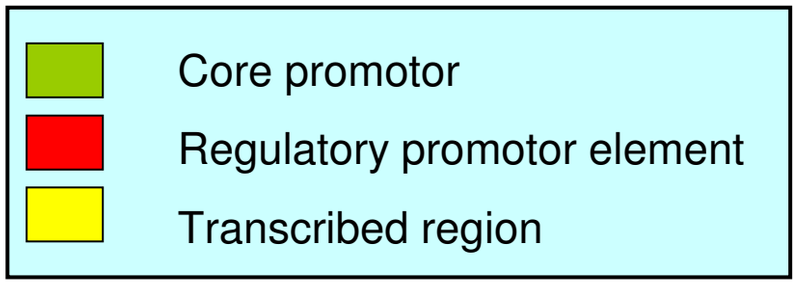
$$\frac{dx_2}{dt} = \alpha_2 H(x_1, \theta_2, p_2) - \gamma_2 x_2$$

  - α :maximal production rate
  - H : gene regulation function (GRF)
  - θ_2: threshold, p_2: steepness
  - γ : relative decay rate



Building a cGP model – 3: Representing genetic variation





Mutations in core promotor (or general regulatory elements) can change the rate of initiation of transcription -> maximal production rates (α)

Mutations in transcribed region (introns and synonymous mutations) can change mRNA stability or RNA prosessing rates -> decay rates (γ)

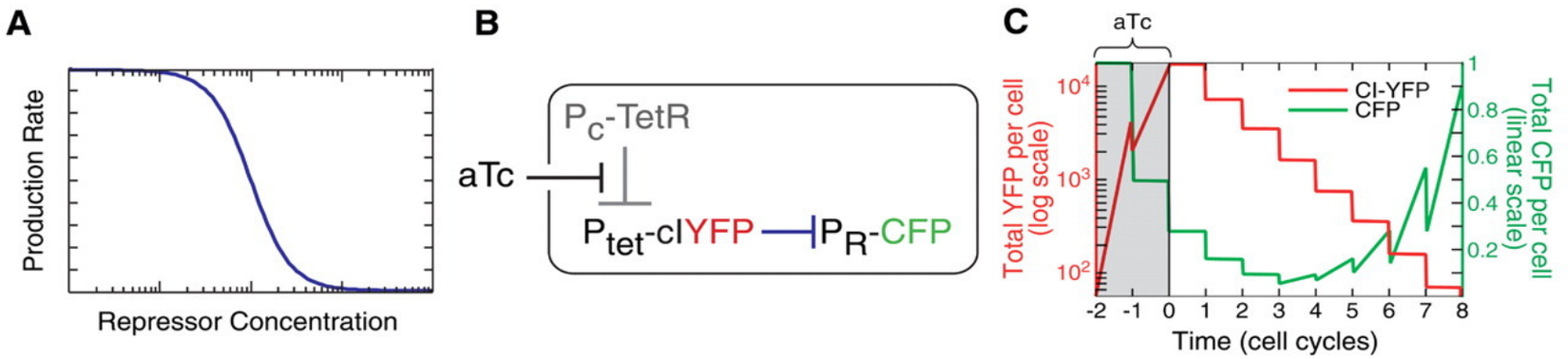
Mutations in specific regulatory elements can change the shape of the gene regulation function -> θ and p

$$\frac{dx_1}{dt} = \tilde{\alpha}_1 - \tilde{\gamma}_1 x_1$$

$$\frac{dx_2}{dt} = \tilde{\alpha}_2 H(x_1, \tilde{\theta}_2, \tilde{p}_2) - \tilde{\gamma}_2 x_2$$

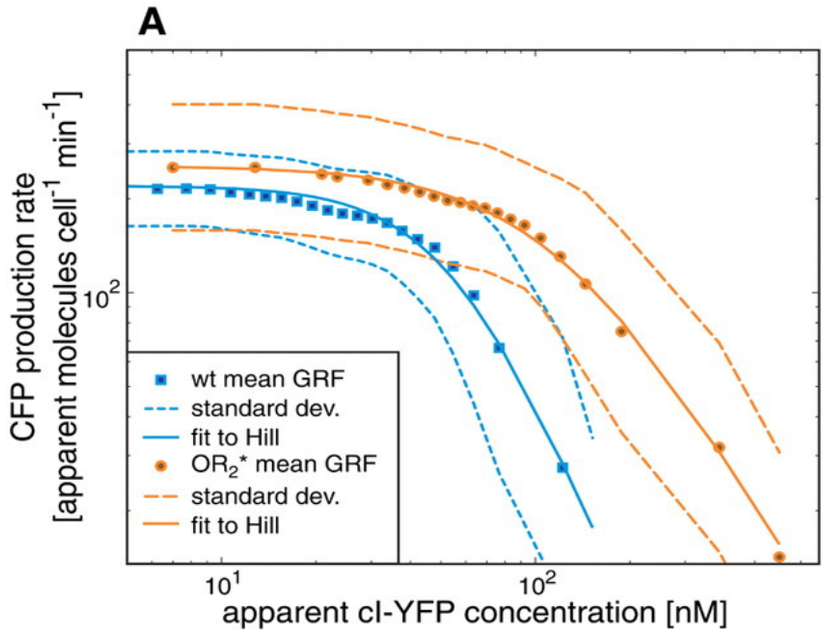
Empirical evidence for genetic variation changing the shape of a gene regulation function

Rosenfeld *et al* (2005) measured the shape of the GRF of the lambda promoter P_R for both a wild-type and a mutant (point mutation in O_R2)



Figures from Rosenfeld et al, Science (2005) , *doi: 10.1126/science.1106914*

- The Hill function describes the shape of the GRF very well
- A point mutation changes both the steepness (*p*) and the threshold ()



How to build a cGP model – 4: Define phenotypes

- The solution of the differential equations describes the gene expression level as a funtion of time
- Any characteristic aspect (qualitative and quantitative) of the solution can be used as a phenotype
- The steady state level is a simple and relevant phenotype

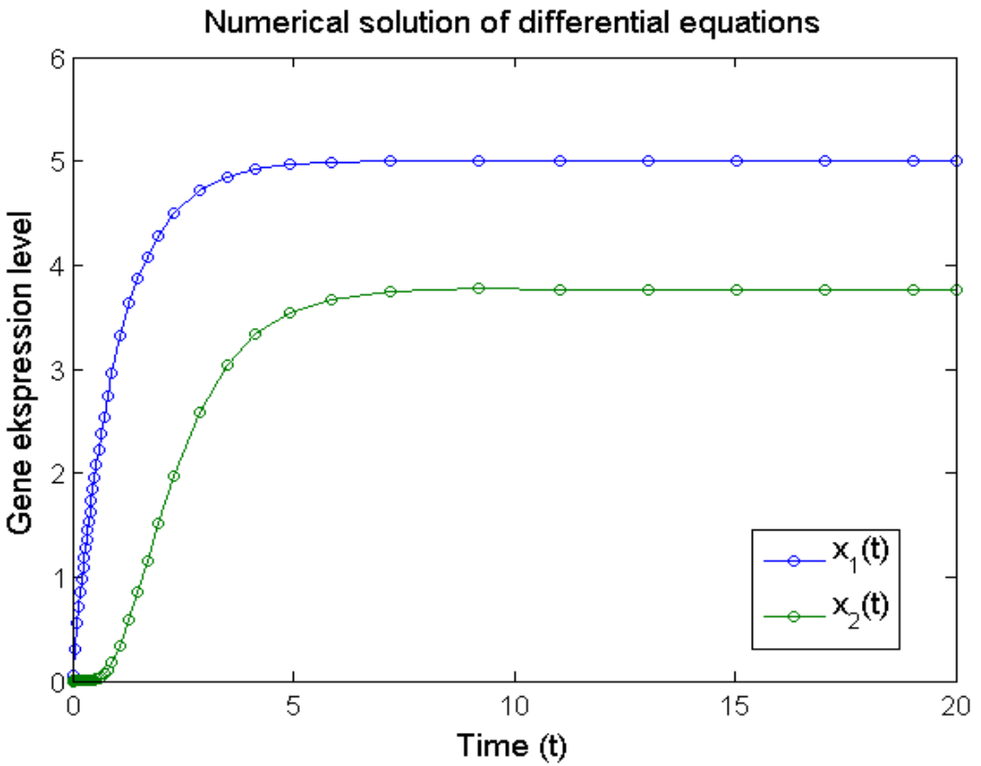
$$\frac{dx_1}{dt} = \alpha_1 - \gamma_1 x_1$$

$$\frac{dx_2}{dt} = \alpha_2 H(x_1, \theta_2, p_2) - \gamma_2 x_2$$

$$\alpha_1, \alpha_2 = 5$$

$$\theta_2 = 4, \quad p_2 = 5$$

$$\lambda_1, \lambda_2 = 1$$



Causally cohesive genotype-phenotype (cGP) models

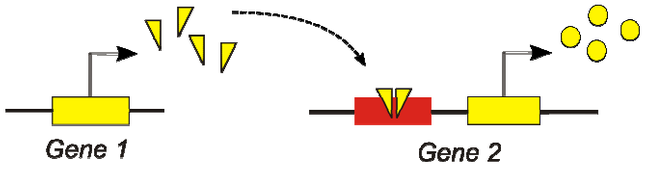
A mathematical model *M* of a biological system is a cGP model if:

- Model elements (variables or parameters) are associated with genes
- Genotypic variation is represented by variation in a set of parameters
- It describes how phenotypes arise from lower level processes in a causally cohesive way

Defines a GP map $T_M : \mathrm{G} \rightarrow \mathrm{P}$ from a set G of genotype indexes to a set P of real-valued phenotypes.

G

- Production rates
- Decay rates
- Shape of gene regulation function



$$\frac{dx_1}{dt} = \alpha_1 - \gamma_1 x_1$$

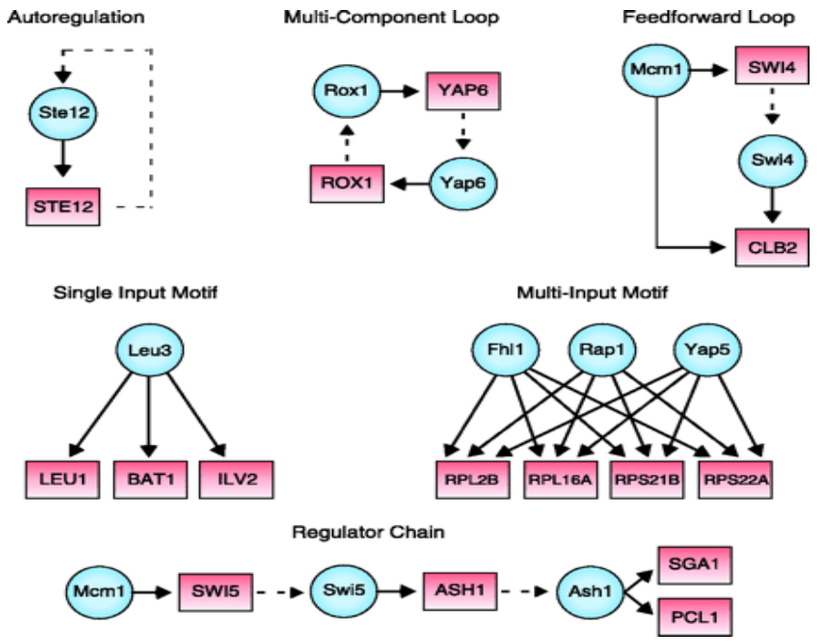
$$\frac{dx_2}{dt} = \alpha_2 H(x_1, \theta_2, p_2) - \gamma_2 x_2$$

P

- Steady state expression levels

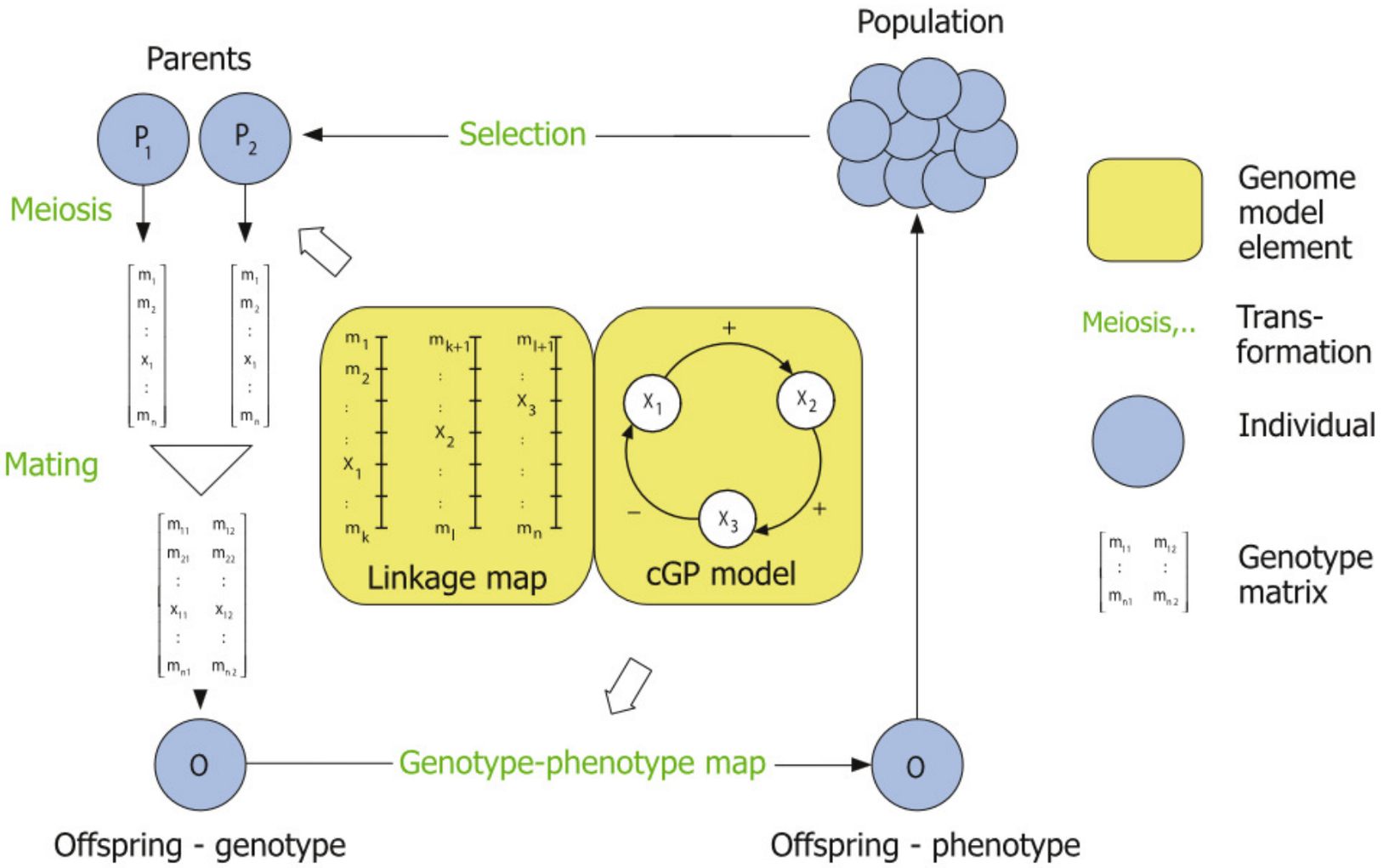
cGP models – what can we do with them

- Explore the link between regulatory principles and genetic descriptors
  - feedback structure, dose-response relationships
  - dominance, epistasis, genetic variance components
- Gene expression phenotypes
  - expression levels are complex genetic traits
  - networks built up by smaller motifs
  - 35 years of modelling experience



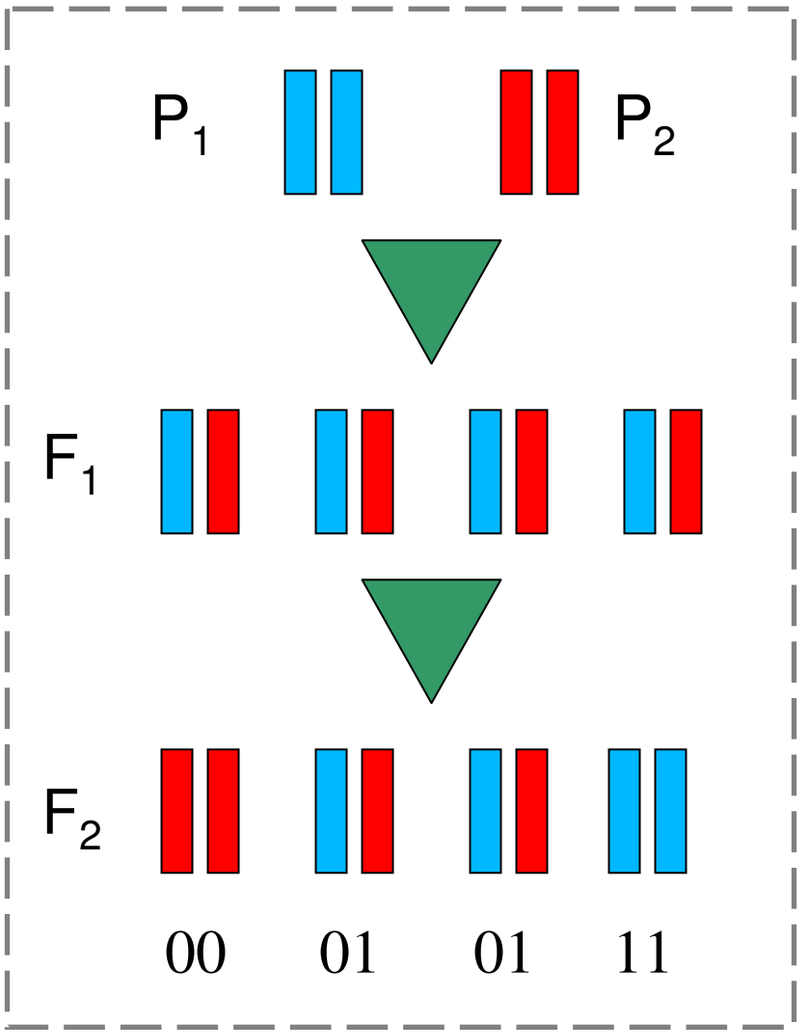
Lee *et al.* (2002), Science

A simulation framework for population studies



- Linkage map : $G \rightarrow G$ creates realistic genotypic variation
  - linkage groups (chromosomes) , linkage disequilibrium (haplotypes)
- cGPmodel : $G \rightarrow P$ associates phenotypes with emergent properties of system

Simulated datasets - F_2 populations



$$F_2 - genotypes: \underset{K\times N}{\begin{bmatrix} 00 & 00 & \cdots & 00 \\ 01 & 00 & \cdots & 00 \\ \vdots & \vdots & \ddots & \vdots \\ 11 & 11 & \cdots & 11 \end{bmatrix}}$$

cGPmodel
N loci
M phenotypes

$$F_2 - phenotypes: \underset{K\times M}{\mathbf{Y}} = [\mathbf{y}_1 \quad \mathbf{y}_2 \quad \cdots \quad \mathbf{y}_M]$$

Quantitative genetic analysis

- F_2 design variables for individual j at gene i

$$w_j^i = \begin{cases} 1 \text{ for genotype } 11 \\ 0 \text{ for genotype } 01, \\ -1 \text{ for genotype } 00 \end{cases} \qquad v_j^i = \begin{cases} -\frac{1}{2} \text{ for genotype } 11 \\ \frac{1}{2} \text{ for genotype } 01. \\ -\frac{1}{2} \text{ for genotype } 00 \end{cases}$$

- Full genetic model N loci:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon,$$

$$\underset{(3^N \times 1)}{\beta} = \left[\mu \quad a_1 \quad d_1 \quad a_2 \quad \cdots \quad d_N \quad aa_{12} \quad \cdots\right]^T,$$

$$\underset{(N \times 3^N)}{\mathbf{X}} = [\mathbf{1} \quad \mathbf{X}_{\text{marg}} \quad \mathbf{X}_{\text{2-way}} \quad \cdots \quad \mathbf{X}_{\text{N-way}}],$$

$$\mathbf{X}_{\text{marg}} = [\mathbf{w}^1 \quad \mathbf{v}^1 \quad \mathbf{w}^2 \quad \cdots \quad \mathbf{v}^N], \quad \mathbf{X}_{\text{2-way}} = [\mathbf{w}^1 \cdot \mathbf{w}^2 \quad \mathbf{w}^1 \cdot \mathbf{v}^2 \quad \mathbf{v}^1 \cdot \mathbf{w}^2 \quad \cdots \quad \mathbf{v}^{N-1} \cdot \mathbf{v}^N], \quad \cdots$$

- Orthogonal regressors in F_2 populations
  - Straightforward to go from regressors to variance components
  - Estimated effects are the same in reduced and full model
  - R package (noia) used for the analysis

Quantitative genetic analysis – continued

- Variance components:

$$V_P = \mathrm{var}(\mathbf{y}), \quad \text{(Phenotypic variance)}$$

$$V_G = \mathrm{var}(\mathbf{X}\beta), \quad \text{(Genetic variance)}$$

$$V_A = \mathrm{var}(\mathbf{X}_A\beta_A), \quad \text{(Additive variance)} \qquad V_D = \mathrm{var}(\mathbf{X}_D\beta_D), \quad \text{(Dominance variance)}$$

$$\mathbf{X}_\mathrm{A} = [\mathbf{w}^1 \quad \mathbf{w}^2 \quad \cdots \quad \mathbf{w}^N], \qquad \mathbf{X}_\mathrm{D} = [\mathbf{v}^1 \quad \mathbf{v}^2 \quad \cdots \quad \mathbf{v}^N],$$

$$\beta_A = [a_1 \quad a_2 \quad \cdots \quad a_N]. \qquad \beta_A = [d_1 \quad d_2 \quad \cdots \quad d_N].$$

$$V_I = V_G - (V_A + V_D), \quad \text{(Epistatic variance)}$$

- Use of variance components:
  - Heritability
  - Breeding
  - QTL mapping
  - V_I statistical epistasis/interaction

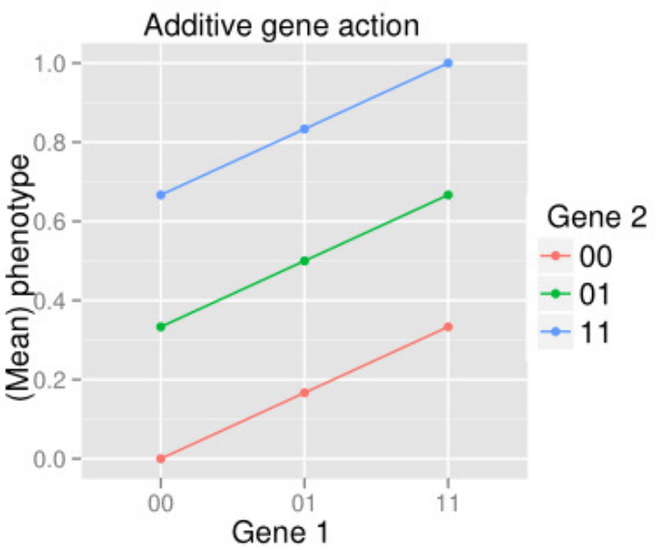
$$H^2 = \frac{V_G}{V_P}, \quad \text{(broad sense heritability)}$$

$$h^2 = \frac{V_A}{V_P}, \quad \text{(narrow sense heritability)}$$

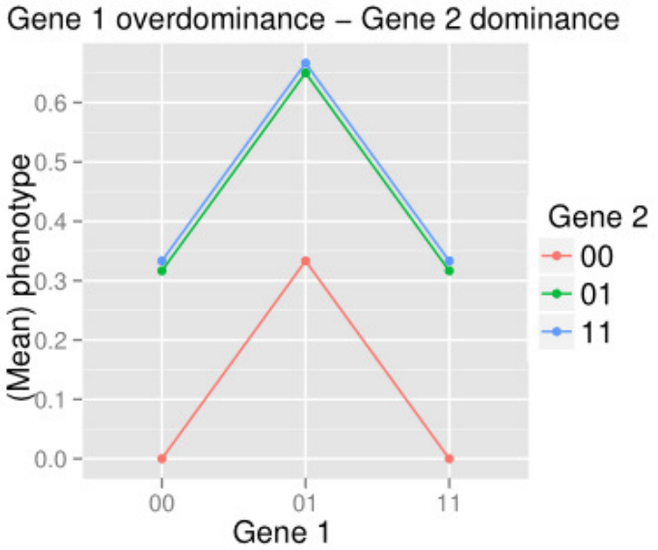
$$R = h^2 \Delta S, \quad \text{(Breeder's equation)}$$

$$\mathrm{b_{OP}} = h^2. \quad \text{(midparent-offspring regression)}$$

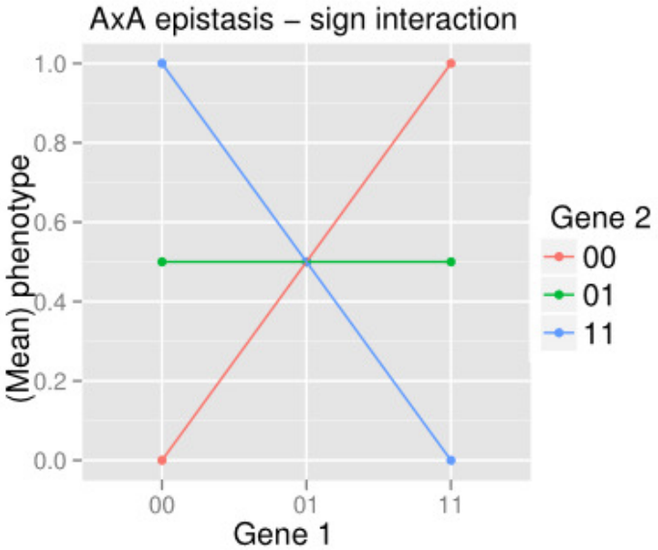
Functional/physiological description of genetic architecture



$$V_A / V_G = 1$$



$$V_A / V_G = 0.3$$
$$V_D / V_G = 0.7$$



$$V_A + V_D / V_G = 0$$
$$V_I / V_G = 1$$

- Functional/physiological vocabulary exists for 1 and 2 loci
  - Additivity, dominance, overdominance, epistasis, sign epistasis, magnitude epistasis
- Based only on the genotype-phenotype map, not on allele frequencies

Connection between feedback and epistasis

GENETICS

HOME | HELP | FEEDBACK | SUBSCRIPTIONS | ARCHIVE | SEARCH | TABLE OF CONTENTS

Institution: Norges Landbrukshoegskoles Bibliotek Sign In as Member

Originally published as *Genetics* Published Articles Ahead of Print on October 8, 2006.

Genetics, Vol. 175, 411-420, January 2007, Copyright © 2007
doi:10.1534/genetics.106.058859

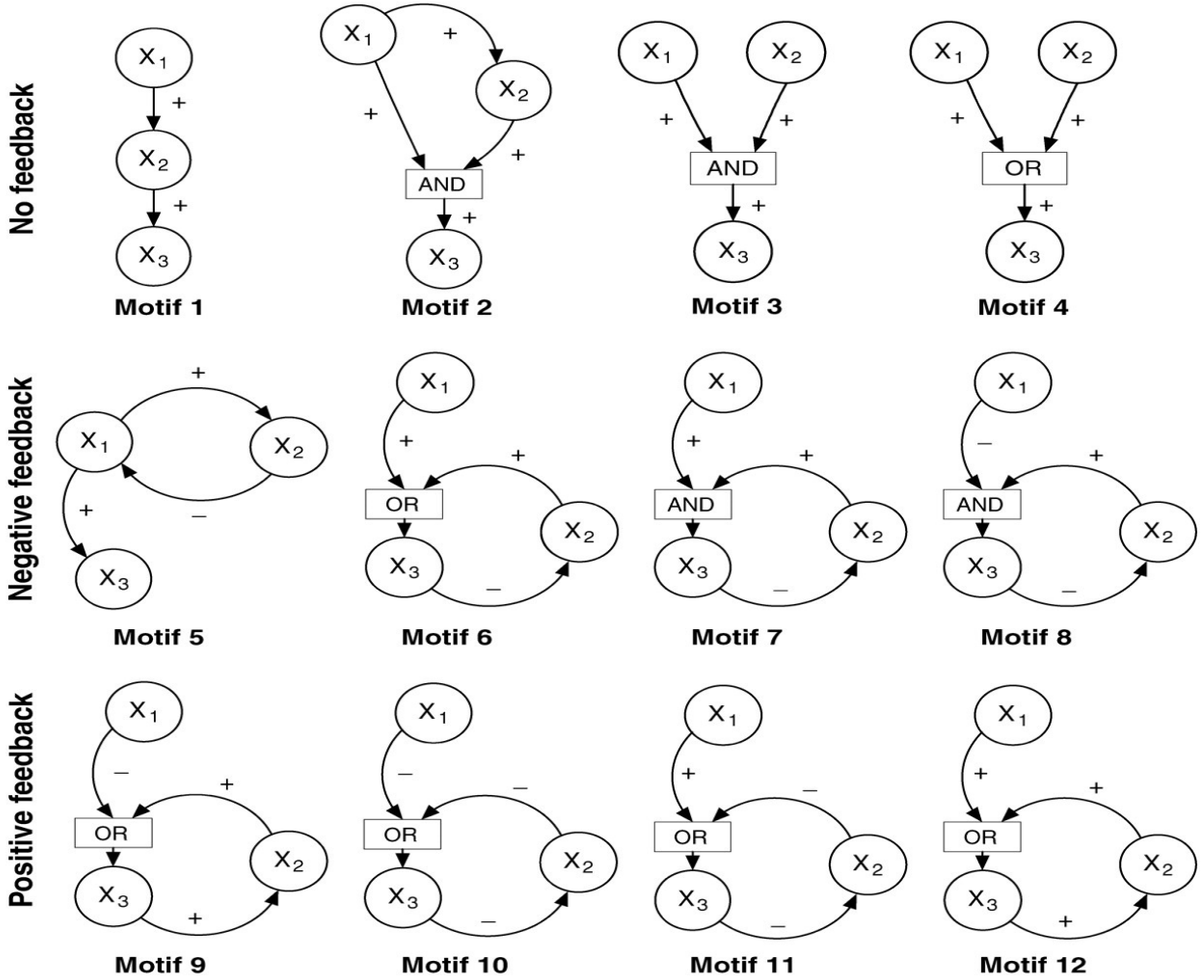
Statistical Epistasis Is a Generic Feature of Gene Regulatory Networks

Arne B. Gjuvsland[*,1], Ben J. Hayes[†], Stig W. Omholt[*] and Örjan Carlborg[‡]

[*] *Centre for Integrative Genetics (CIGENE) and Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, N-1432 Aas, Norway,* [†] *Animal Genetics and Genomics, Department of Primary Industries, Attwood, Victoria, Australia 3049 and* [‡] *Linnaeus Centre for Bioinformatics, Uppsala University, SE-751 24 Uppsala, Sweden*

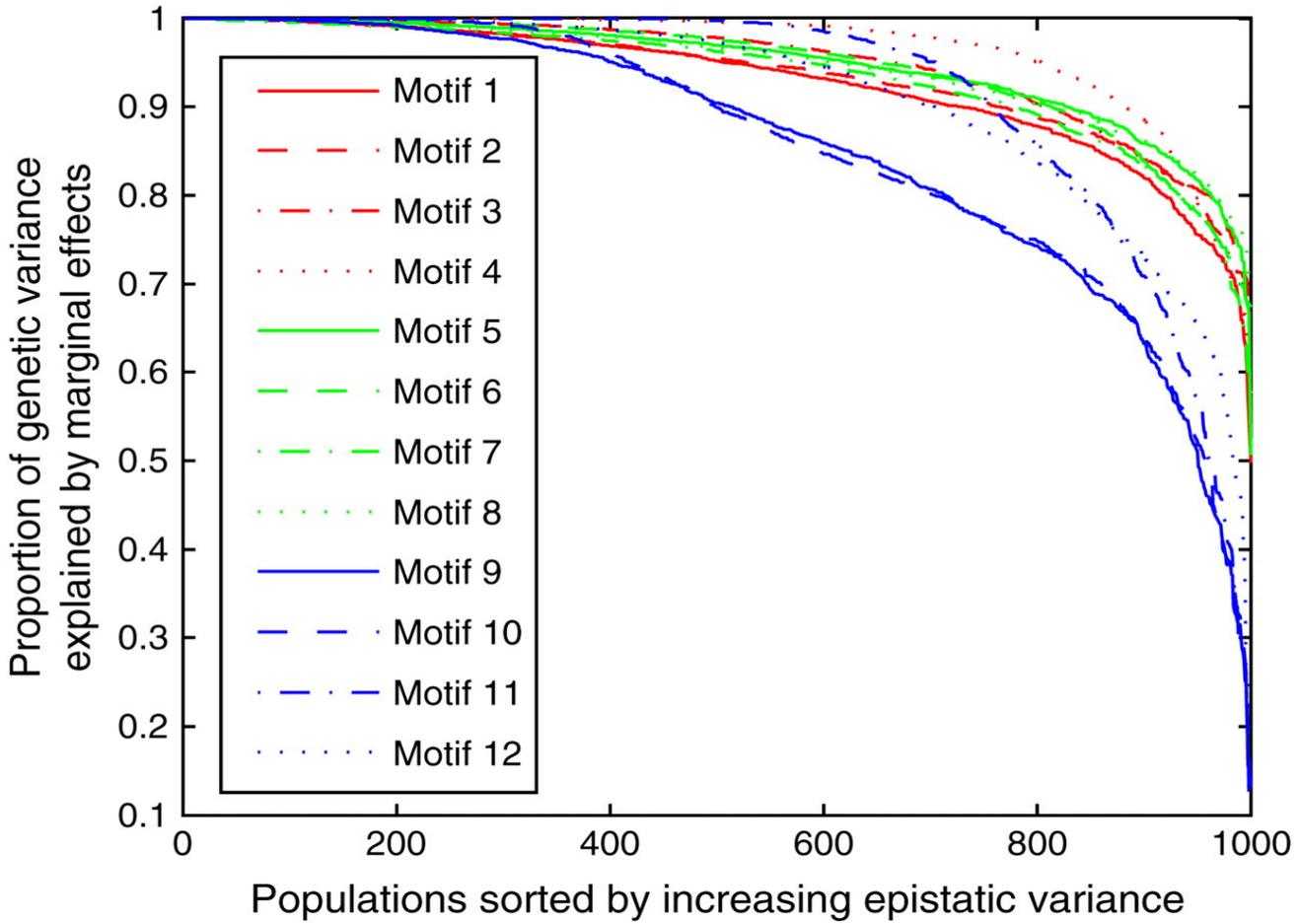
- Feedback loops are ubiquitious in all biological systems (gene regulation, metabolism, signalling)
- How do system level interactions amongs genes map into statistical interactions between the same genes?

Feedback study: Simulations



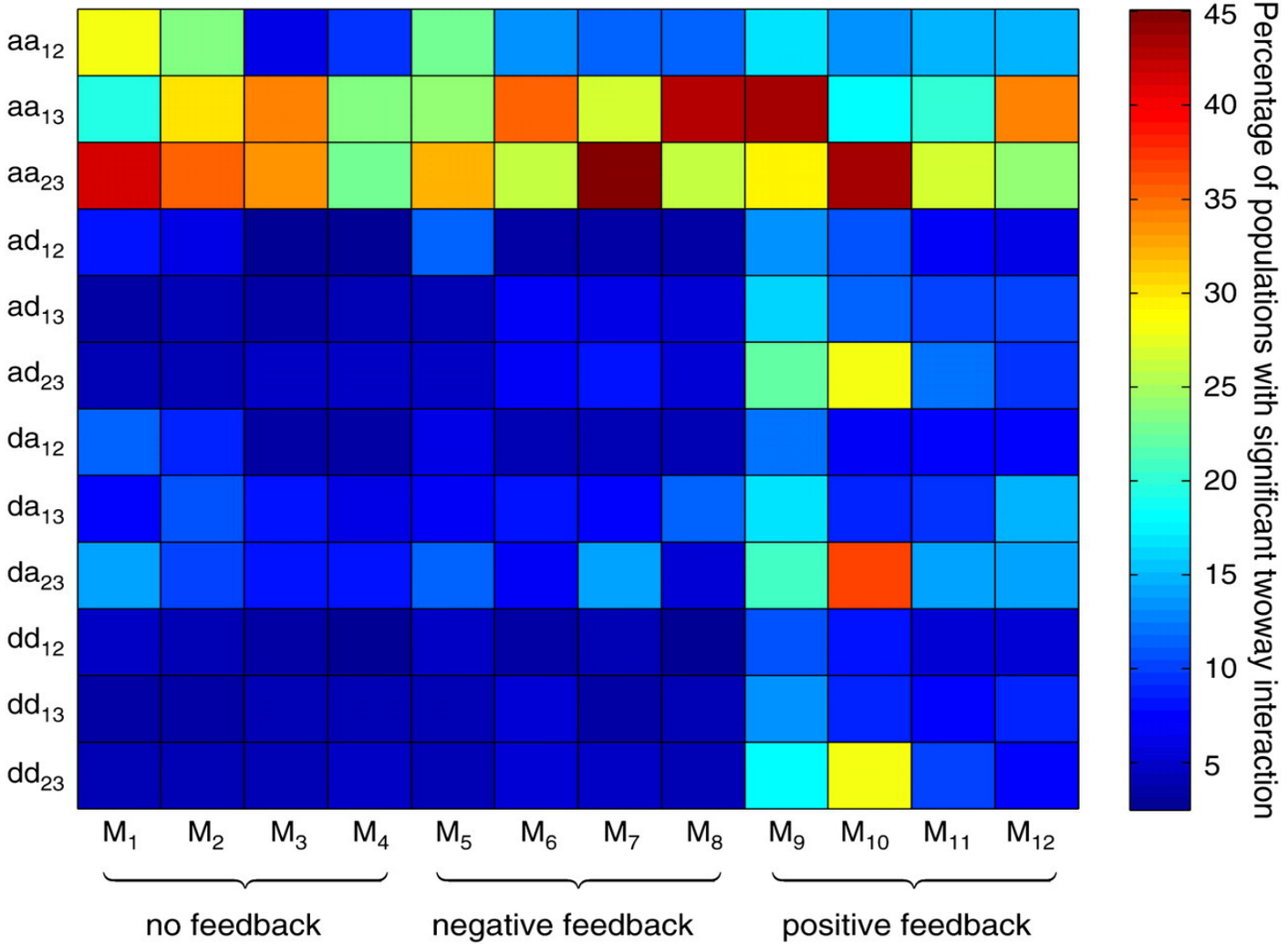
- 3 classes of networks: no, negative and positive feedback
- Simulated 1000 F_2-populations with heritable variation in maximal production rates and gene regulation functions

Positive feedback gives more epistatic variance



- Large span in the statistical genetic architecture
- Additive and dominance variance dominates
- Positive feedback (blue) gives more epistasic variance

Positive feedback give more types of epistasis



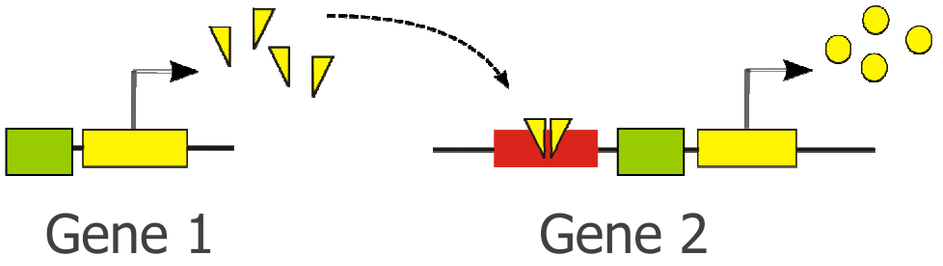
- All motifs give additive-by-additive interactions
- Positive feedback gives richer set of two-way interactions

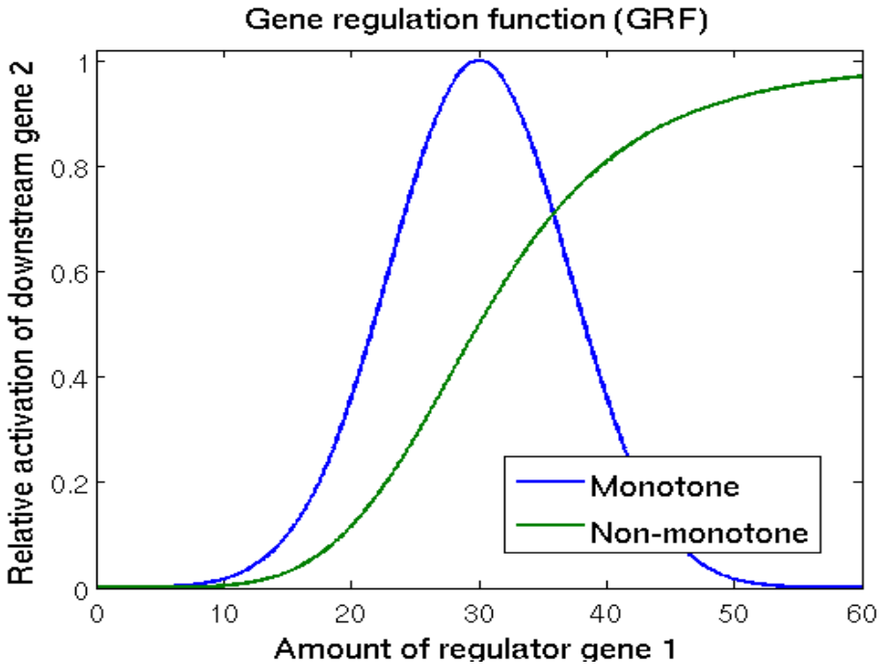
Following up on the feedback study

- How can highly non-linear gene regulatory networks produce mainly additive genetic variance?
  - Focus on the shape of the gene regulation function

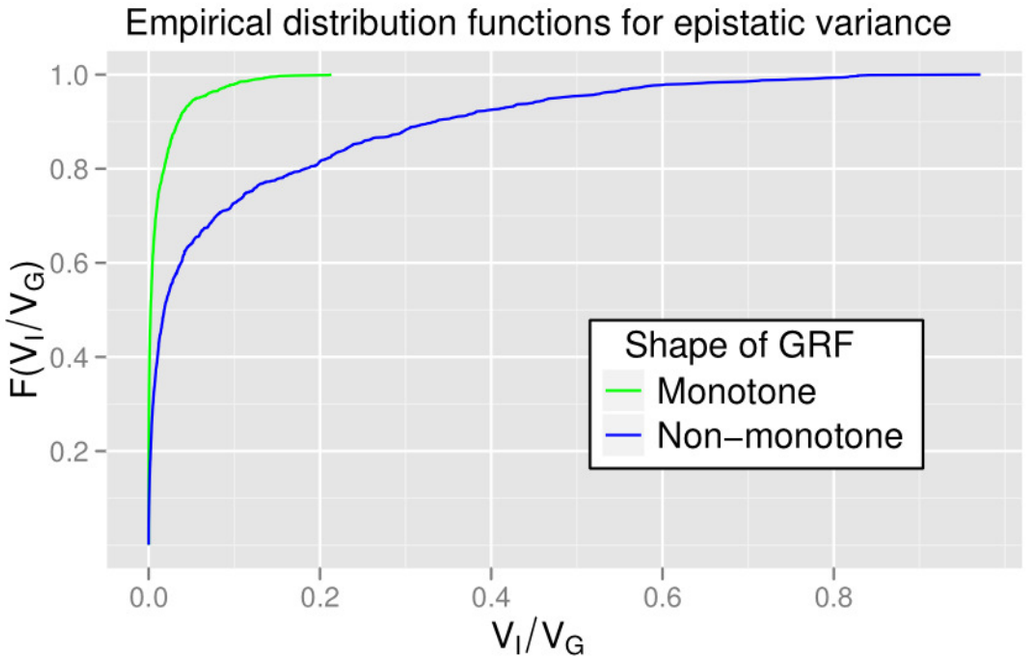
- cGP models for more complex biological systems
  - Utilize publicly available and curated models (SBML and CellML) for cGP-studies of complex biological systems

Connection between gene regulation function and epistasis

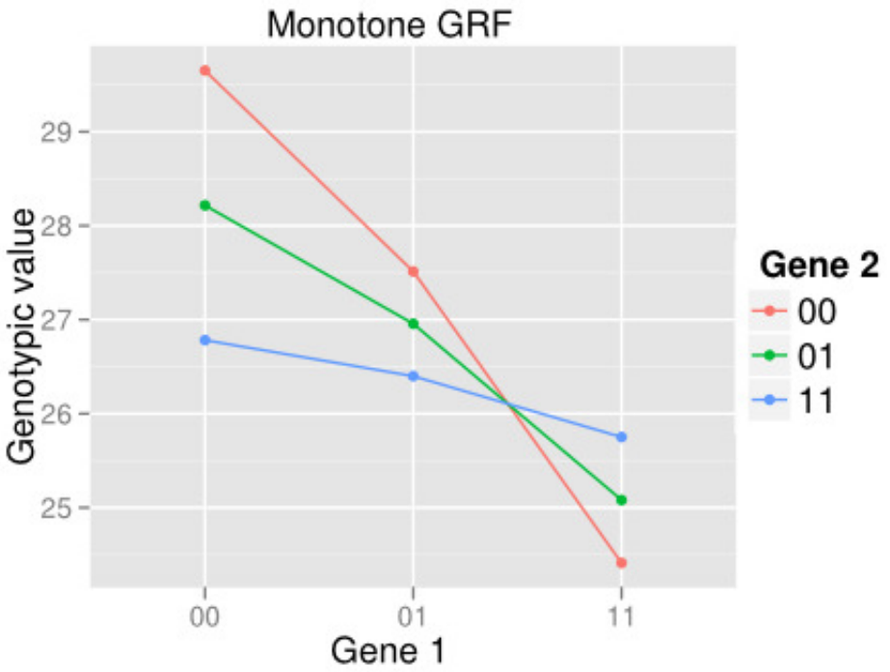


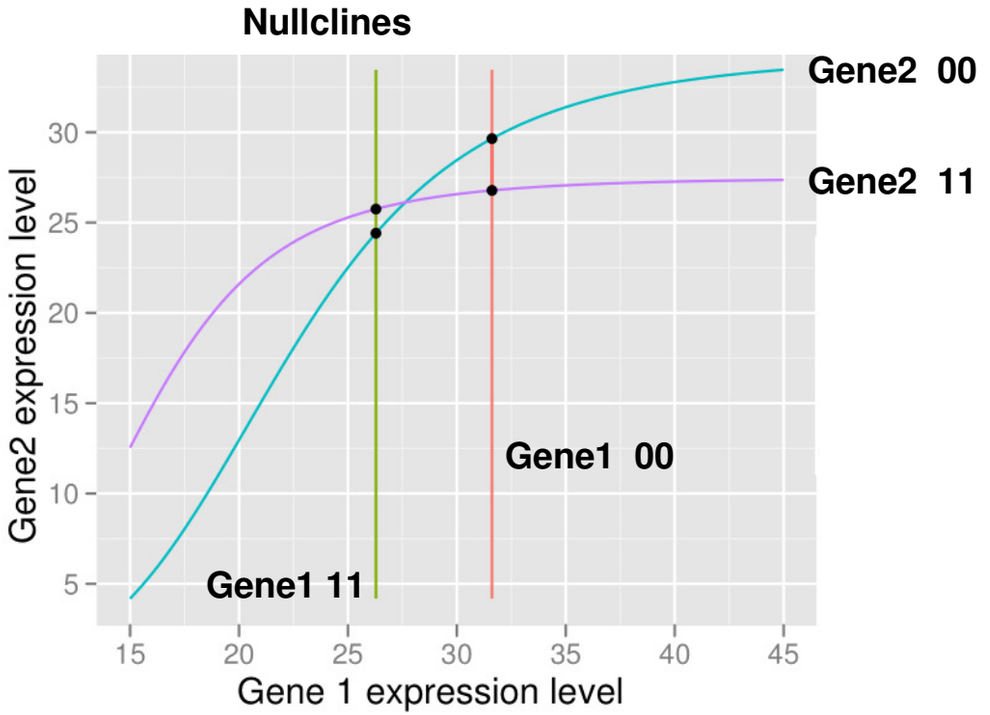


- ODE models with two genes
- Monotone vs. non-monotone gene regulation function
- Introduce genetic variation on production, decay and shape parameters
- Phenotype: steady state expression level of gene 2
- Population setup and analysis as in feedback study



Gene regulation function and functional epistasis





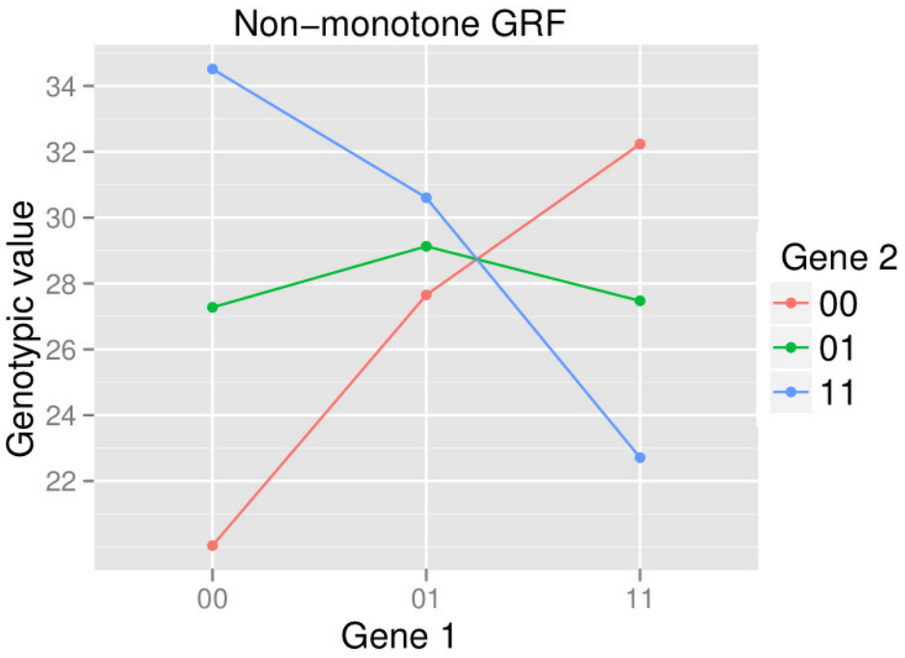
- Highly epistatic datasets show sign epistasis
- Sign epistasis occurs only for gene 2
- The monotonicity of the gene regulation function makes sign epistasis at gene 1 impossible

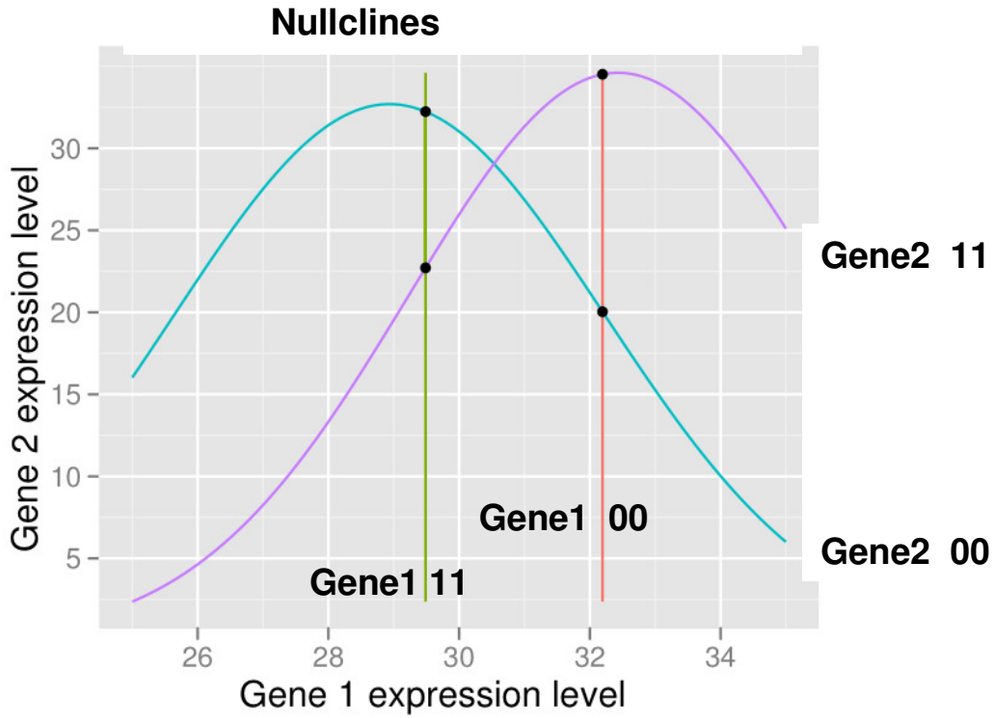
Nullclines:

$$\dot{x}_1 = 0 \quad \Rightarrow \quad x_1 = \frac{\alpha_1}{\gamma_1},$$

$$\dot{x}_2 = 0 \quad \Rightarrow \quad x_2 = \frac{\alpha_2}{\gamma_2} H(x_1, \theta_2, p_2).$$

Gene regulation function and functional epistasis





- Datasets with high epistatic variance show sign epistasis
- Sign epistasis occurs for both genes due to the non-monotonicity of the gene regulation function

Nullclines:

$$\dot{x}_1 = 0 \quad \Rightarrow \quad x_1 = \frac{\alpha_1}{\gamma_1},$$

$$\dot{x}_2 = 0 \quad \Rightarrow \quad x_2 = \frac{\alpha_2}{\gamma_2} \exp\left(-\frac{(x_1 - \mu_1)^2}{2\sigma^2}\right).$$

cGP model of glycolysis

- cGP model from Teusink *et al* , 2000
- 13 enzymes identified as genes
- 3 polymorphic loci drawn at random
- Variation in Vmax for polymorphic loci
  - Uniformly +-30% from original value
  - Additive gene action at parameter level
- System solved using Pysces
  - Stable steady state used as phenotypes
  - Dataset without stable s.s. or with s.s. concentrations > 20-fold higher than default discarded
- 1000 Monte Carlo simulations
- Full noia analysis of each dataset
  - 5 phenotypes, 243 genotypes

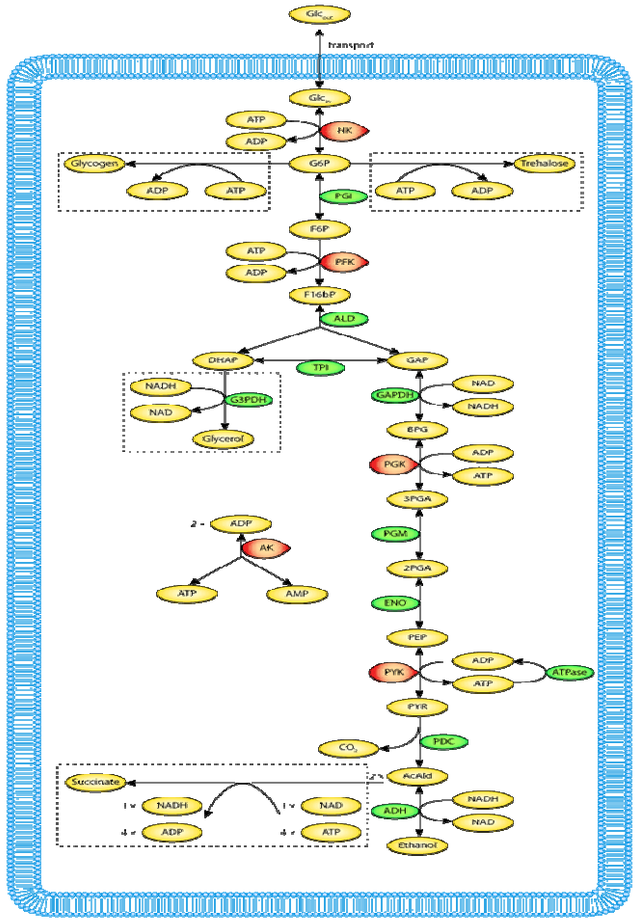
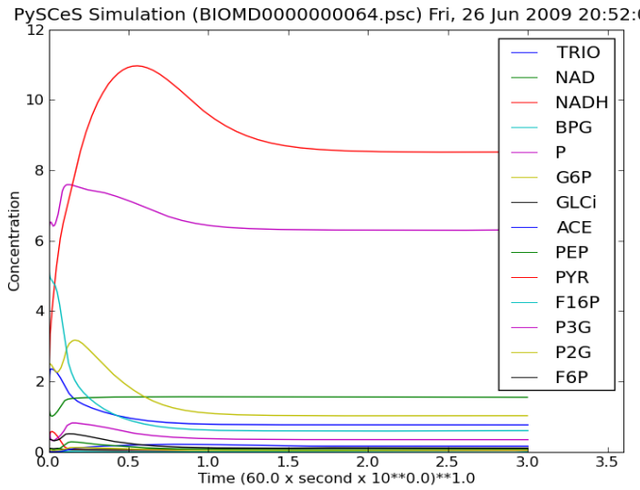


Figure from http://model.cellml.org



cGP model of mammalian circadian clock

- cGP model based on CellML implementation of model by Leloup and Goldbeter, 2004
- 3 genes *Bmal1*, *Per, Cry*
- Variation in mRNA decay rates
  - Uniformly +-30% from original
  - Additive gene action at parameter level
- System solved in PySundials
  - Oscillation period
  - Lowest value and time to peak for 16 state variables
- 1000 repetitions
- Full R\noia analysis of each dataset

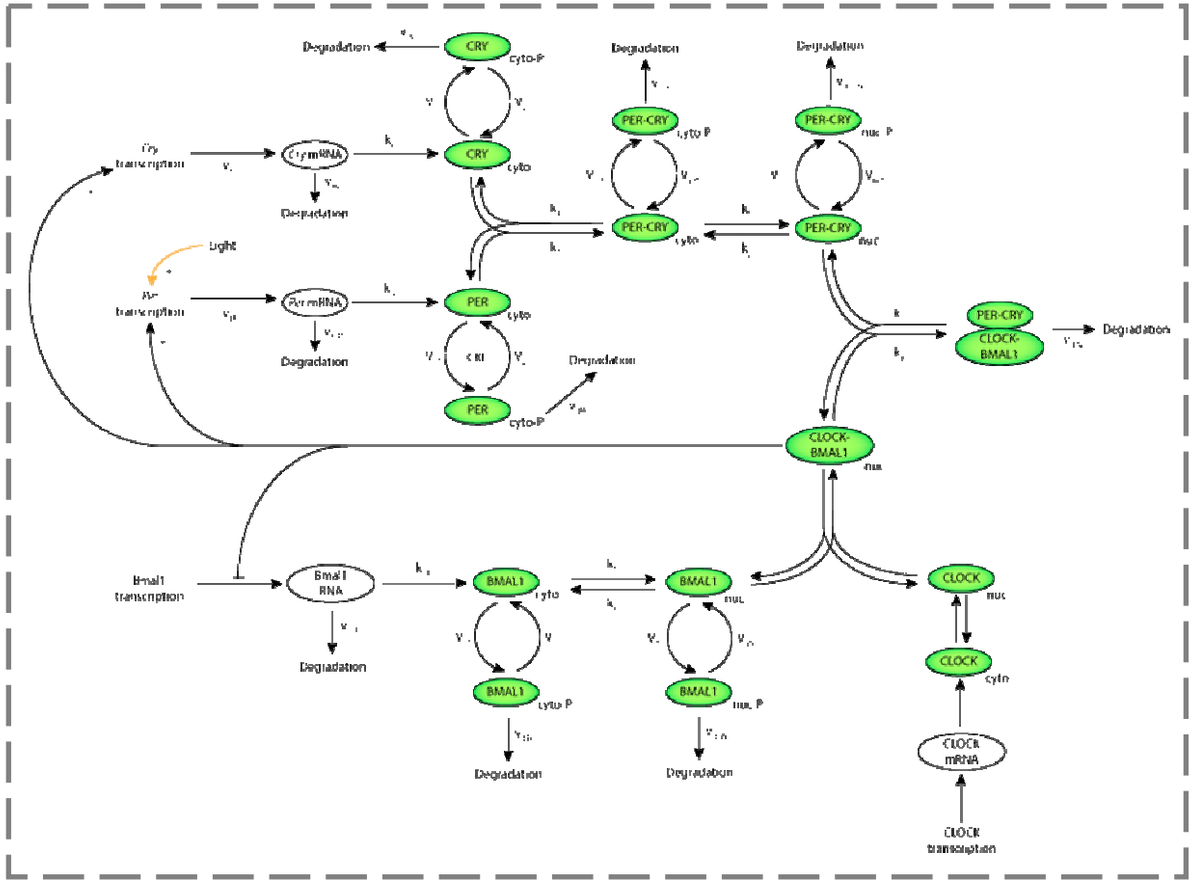
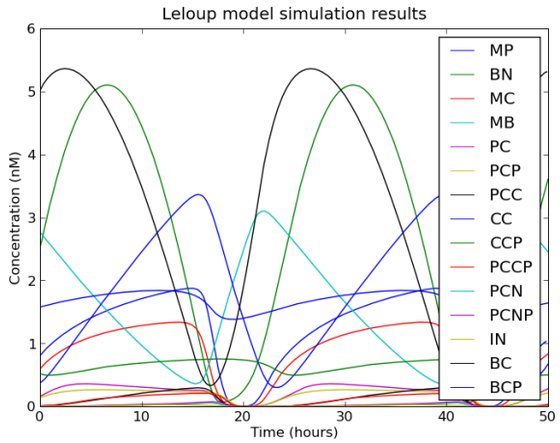
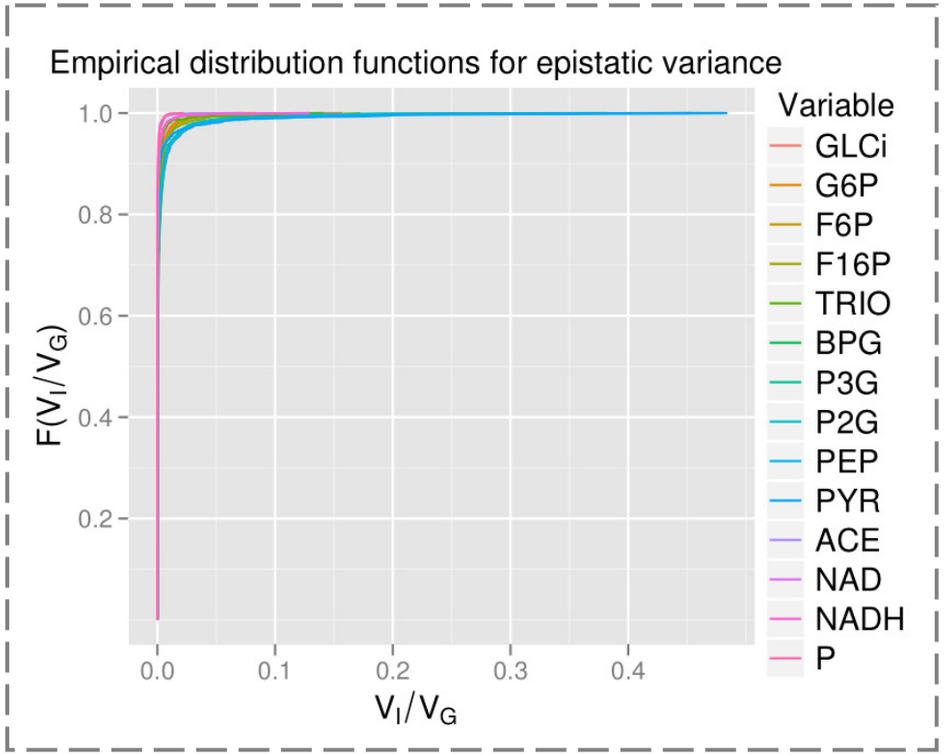
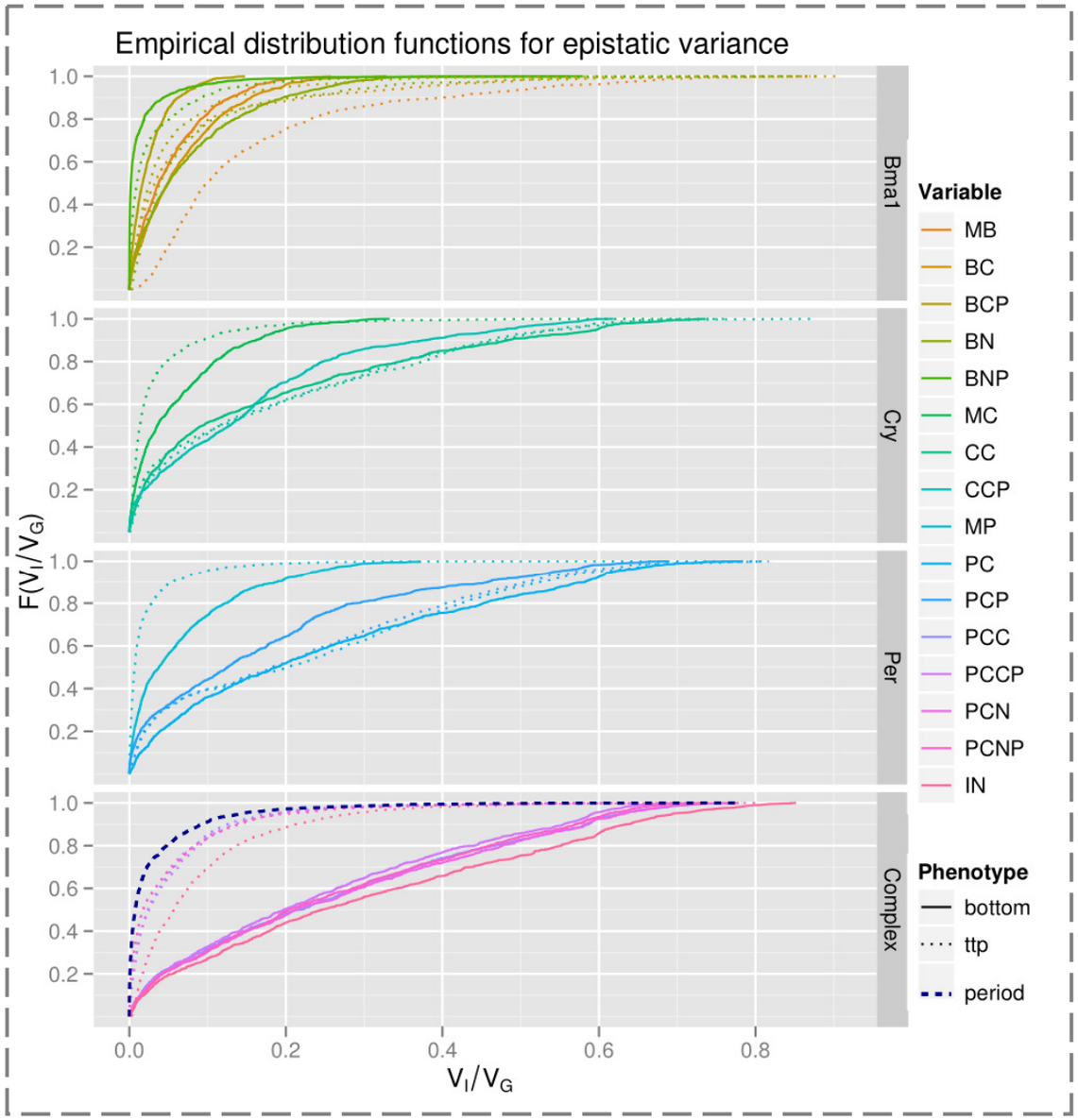


Figure from http://model.cellml.org



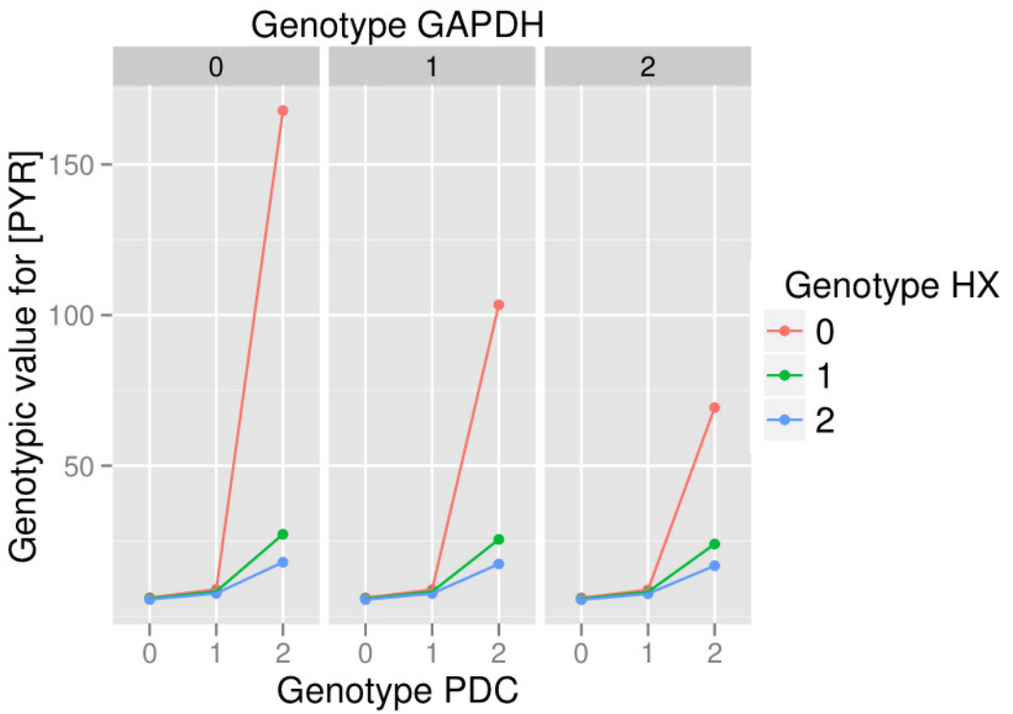
Statistical epistasis comparison glycolysis vs. circadian clock

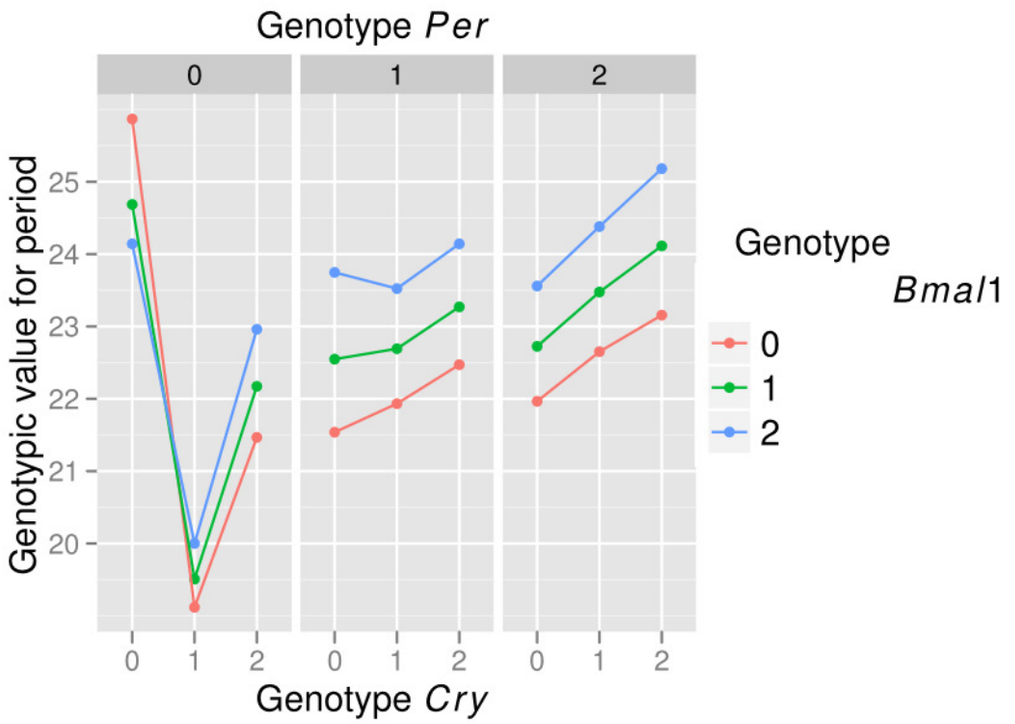




- Glycolysis cGP model has very little room for epistatic variance
- Circadian clock cGP model shows much epistatic variance for all phenotypes
  - For all protein complexes bottom concentration level has shows considerably more epistasis than time to peak

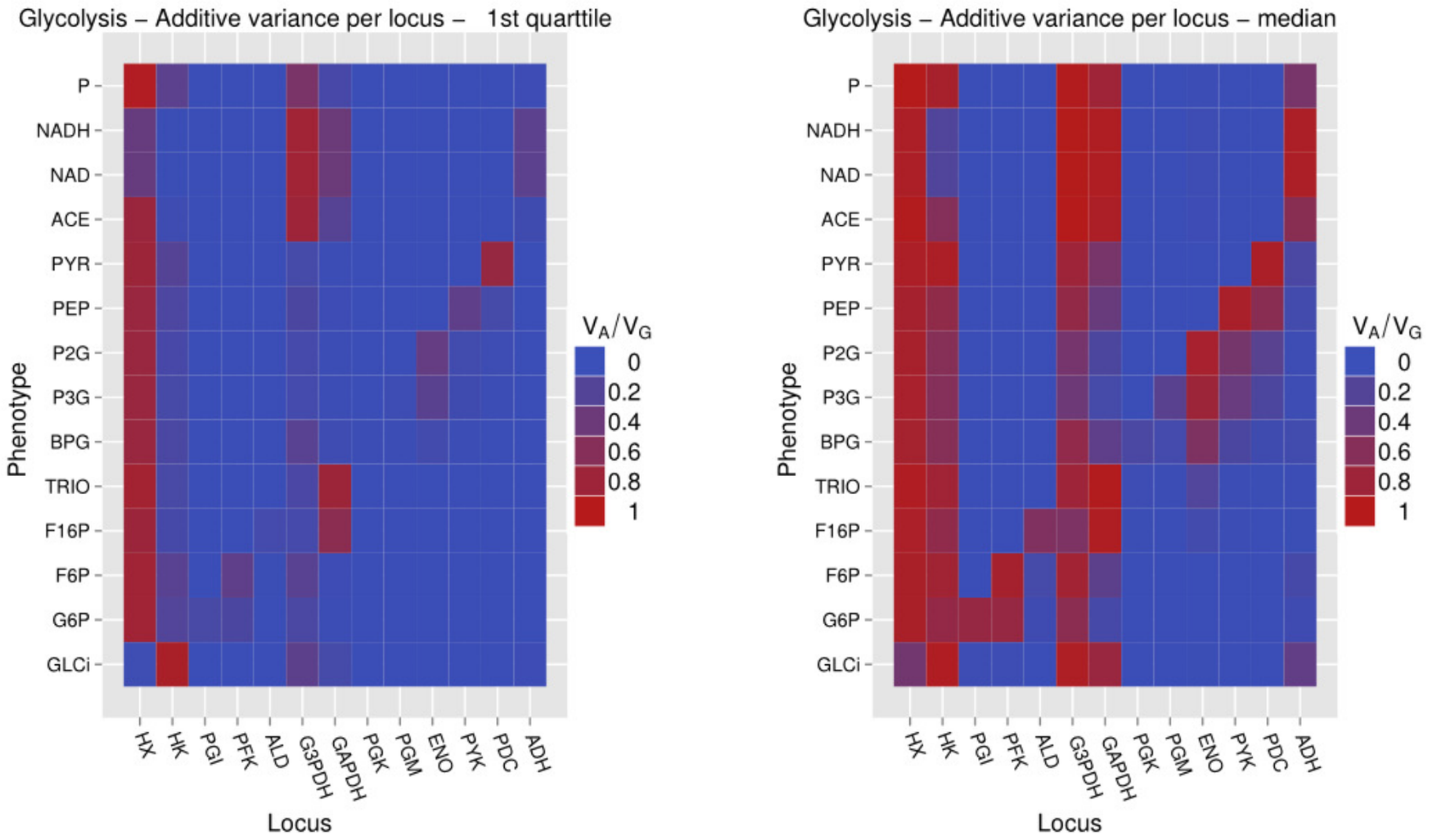
Functional epistasis comparison glycolysis vs. circadian clock





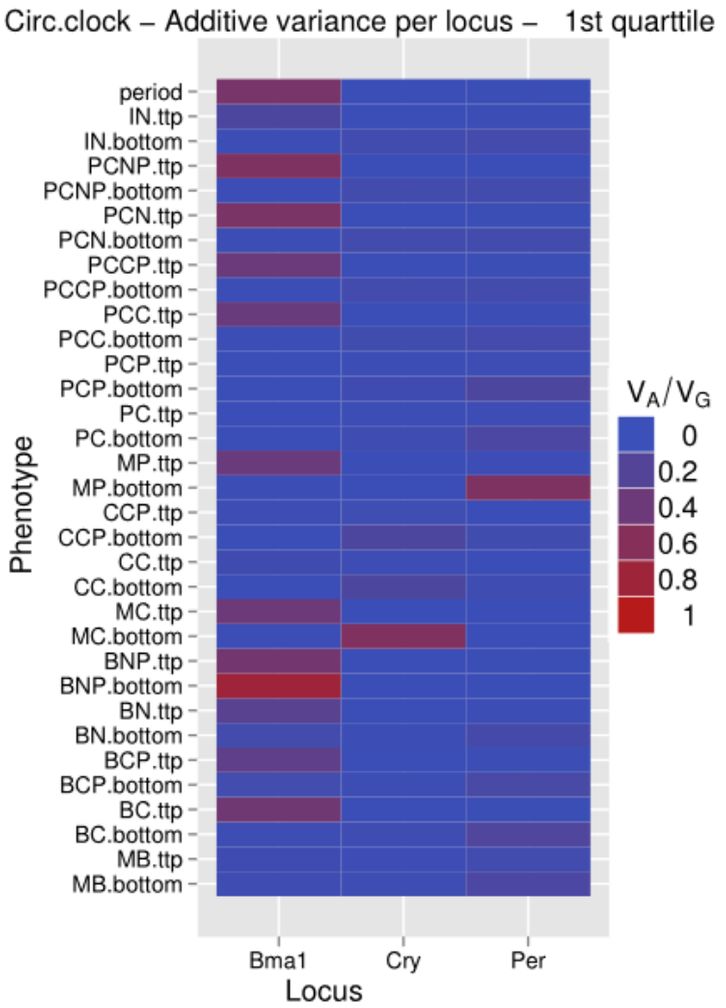
- For glycolysis model datasets the statistically most epistatic datasets show strong magnitude epistasis, but no overdominance or sign epistasis
- For circadian clock model the statistically most epistatic datasets show both overdominance and sign epistasis

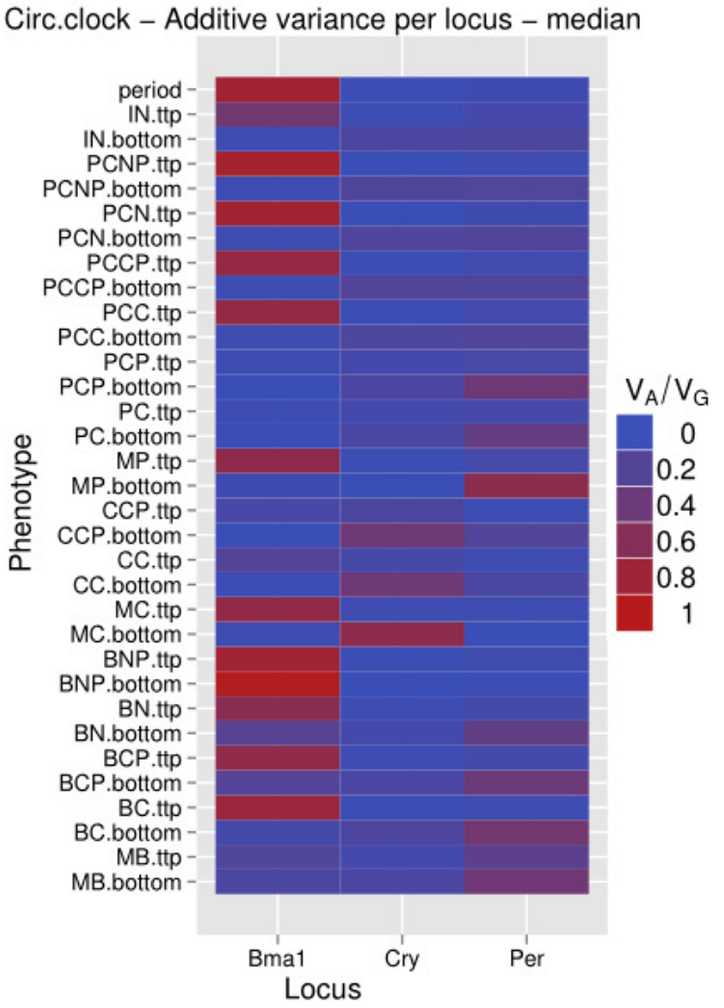
Genetic architecture – single locus, additive effects



- Local genetic variation along the pathway, variation in enzyme parameters have large effect on substrate concentration
- Distant genetic variation in glucose transporter, hexokinase and branch-gating enzymes

Genetic architecture – single locus, additive effects





- Some local genetic variation (*Cry* and *Per* on mRNA level), but generally all three genes explain variation distant phenotypes
- *Bmal1* explains period and time to peak for all complexes, *Cry* and *Per* explains bottom level for all complexes

Summary

- Challenge: understand variation in organisms as a function of genes and environment in a mechanistic sense
- Causally-cohesive genotype-phenotype (cGP) models
- cGP studies of gene regulatory networks produce clear patterns
  - Positive feedback increases both the amount and types of statistical interactions
  - The shape of the gene regulation function has large impact on epistasis, monotone GRF reduce the room for sign epistasis and statistical interactions
- Similar results observed for models of more complex systems
  - Circadian clock with several feedback loops produces much statistical epistasis and rich functional epistatic patterns
  - Glycolysis model without feedback and with monotonic enzyme kinetics shows almost no statistical interaction