COSC 348: Computing for Bioinformatics

Lecture 22:
Gene Regulatory Networks 2 - weight matrices

Lubica Benuskova

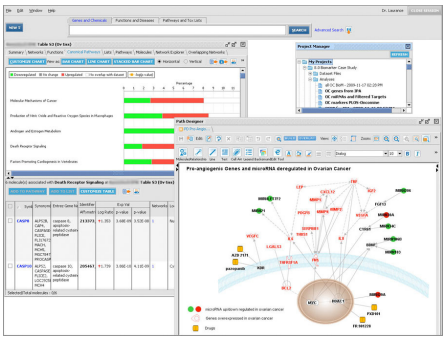
http://www.cs.otago.ac.nz/cosc348/

GRN inference and modelling GRN dynamics

- GRN inference means a reconstruction of GRN based on experimental data.
- Modelling **the GRN dynamics** means simulating the changes of gene expression levels over time.
- For that we need a particular model:
  - Boolean Networks √
  - **Weight matrices**
  - Differential equations
  - Bayesian Networks
  - other

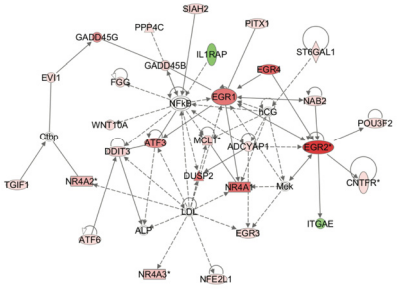
GRN inference by IPA

- Ingenuity® Pathway Analysis (IPA) software enables GRN inference based on their Ingenuity®Knowledge Base containing information on genes, proteins, chemicals, drugs, and molecular relationships.

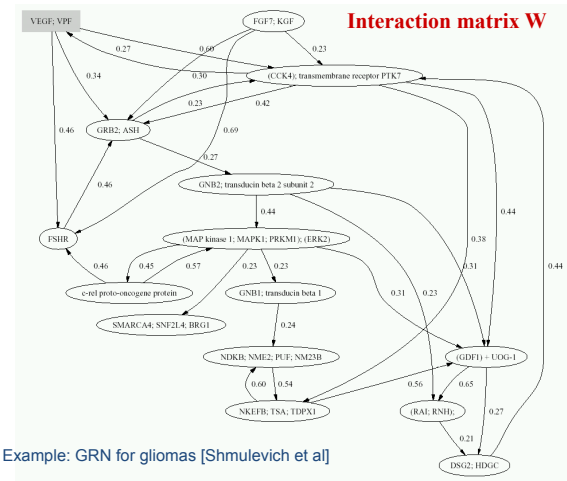


GRN as a weight matrix of interactions

- **W** = matrix of weighted interactions.
- Element of matrix $-1 < w_{ij} < +1$
- if $w_{ij} < 0$, then the influence is inhibitory, otherwise it's activatory.
- Concrete numbers quantify the strength of the influence.



The goal is to find the values of interaction coefficients in the matrix W so that we can predict the next state (expressions of genes)



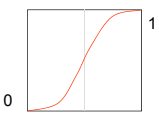
Example: GRN for gliomas [Shmulevich et al]

Update rule for gene expressions once we know **W**

- Regulatory interactions between genes are modelled with a weight matrix **W** such that the expression level (concentration of mRNA) of gene *i* at time *t*+1 is calculated as:

$$m_i(t+1) = f\left(\sum_{j=1}^{N} w_{ij}\, m_j(t)\right)$$

  - $m_j(t)$ = normalized expression level of gene *j* at time *t*
  - w_{ij} = regulatory influence of gene *j* on gene *i* (assumed to be constant!), either activation ($w_{ij} < 0$) or inhibition ($w_{ij} > 0$)
  - *f* is a nonlinear function, e.g. a sigmoid:



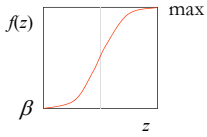
Update rule with the bias

- The β_i parameter represents the **basal expression** level of the gene in the absence of any regulative input (its "**bias**")
- Thus the transition equation (update rule) now reads:

$$m_i(t+1) = f\left(\sum_{j=1}^{N}\left(w_{ij}\, m_j(t) + \beta_i\right)\right)$$

- Where f is a sigmoid function:

$$f(z) = \frac{1}{1+\exp(-z)}$$



Fitness function: goodness of fit to data

- We want to minimise the total error E, which is the sum of squared differences between measured and predicted values of gene expression:

$$E = \sum_{\forall i}\sum_{\forall t}\left(m_i^{real}(t) - m_i^{predicted}(t)\right)^2$$

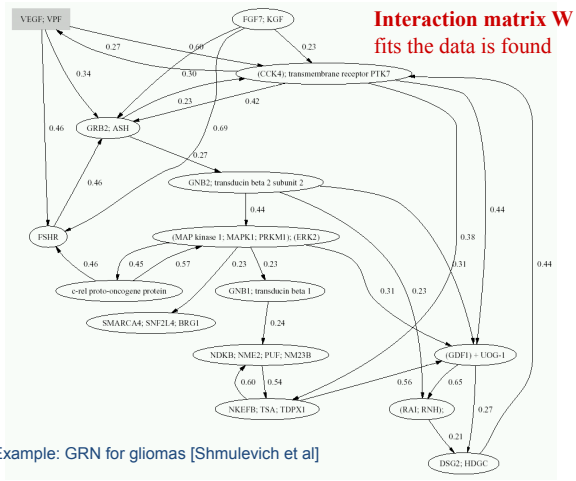
r	g_1	g_2	g_3	g_4	g_5	g_6
t_1	1.7	1.5	1.2	-0.3	1.4	1.6
t_2	1.8	-0.7	1.3	0.8	-0.1	1.7
t_3	-1.8	0.4	1.7	1.8	0.6	-0.4
t_4	-1.7	-1.4	0.9	0.5	-1.8	-0.2
t_5	0.0	1.9	-1.9	1.7	1.6	-0.5

Expression of six genes measured at 5 different time steps. Thus, i = 1…6 and t = 1…5

Finding **W** by means of a genetic algorithm

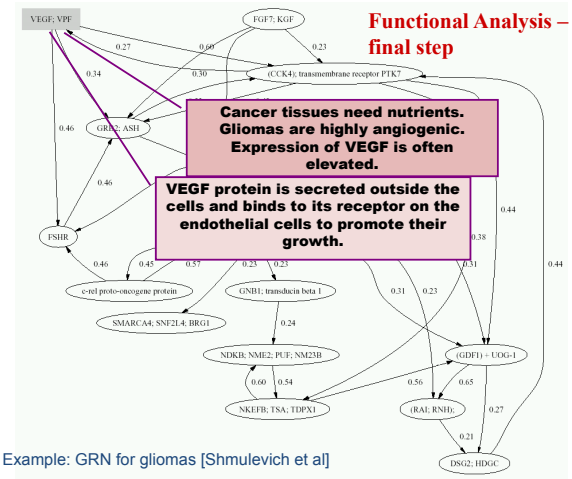
- Start with a large "population" of randomly generated solutions to a problem, i.e. many random interaction matrices **W**
- Repeatedly do the following:
  - Evaluate each of the attempted solutions for the goodness of match with the temporal expression of each gene
  - Use the best solutions to generate a new population through crossover and mutation.
- Quit when you have a satisfactory solution (or you run out of time).

Interaction matrix W that fits the data is found

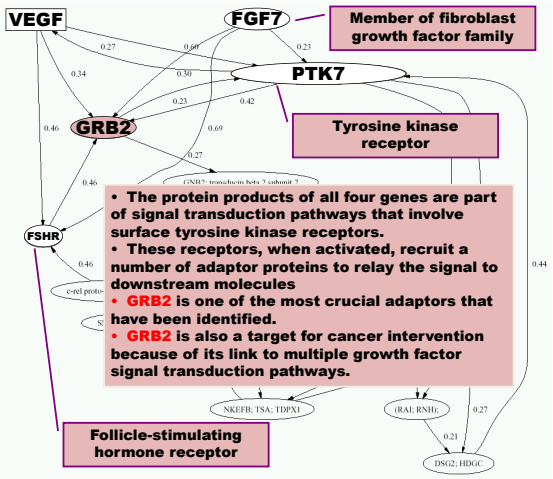


Example: GRN for gliomas [Shmulevich et al]

Functional Analysis – the final step

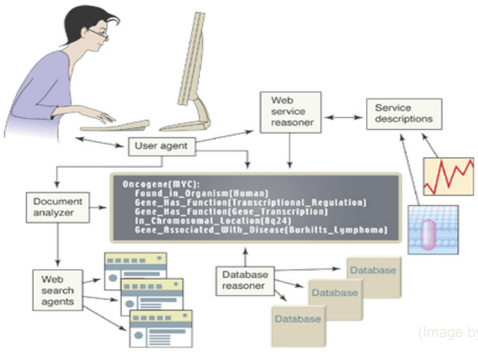


Example: GRN for gliomas [Shmulevich et al]



Interpretation of results

- We want to make sense of the selected genes and their interactions, ideally at one integrated place.



(Image by Jim Hendler)

Gene Ontology (GO)

- Gene Ontology (GO) provides a controlled vocabulary to describe gene and gene product attributes in any organism.

- Recent ontology statistics: As of ontology version 1.1423, dated 13:09:2010
  - 32560 terms, 99.3% defined;
  - 19489 biological processes
  - 2759 cellular components
  - 8867 molecular functions
  - **Etc. (links to dozens of microarray processing tools)**

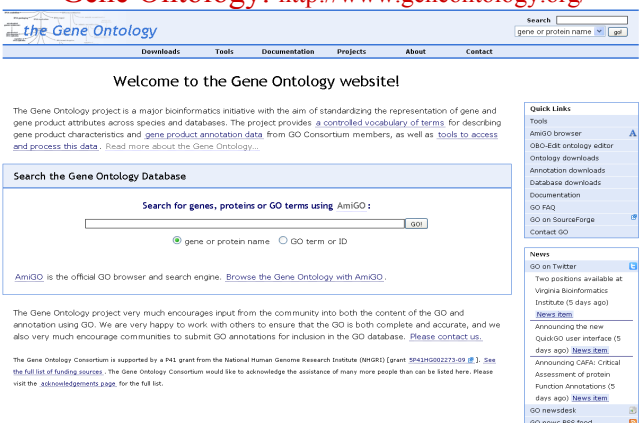
Ontology

- In philosophy, is a study of conceptions of reality and the nature of being. Ontology seeks to describe or posit the basic categories and relationships of being or existence to define entities and types of entities within its framework.

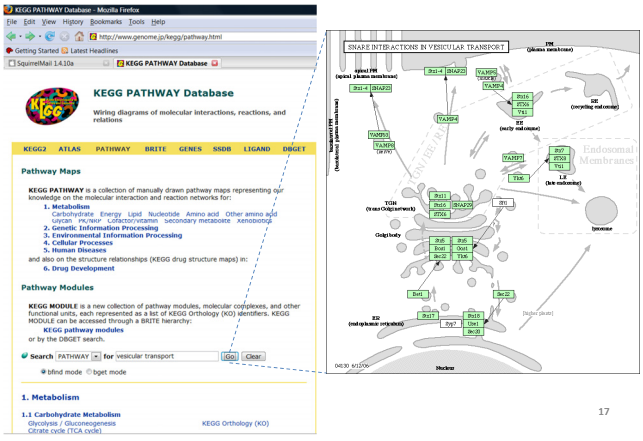
- Ontology in Computer Science: *Formal treatment of the concepts and their relationships in a given domain.*

- Applications
  - Representing and storing data (e.g., database schema)
  - Knowledge sharing within and between domains
  - Search and retrieval
  - Intelligent organization of data resources

Gene Ontology: http://www.geneontology.org/



KEGG: **Kyoto Encyclopedia of Genes and Genomes**



Weight matrices GRN: evaluation

- Delays of interactions between genes are **not** taken into account;

- The dynamics of GRN is discrete & synchronous. In reality this update is *asynchronous and continuous process*;

- Interaction coefficients are some abstract numbers behind which we should see chains of events leading from expression of one gene to other genes in the network, thus it's difficult to interpret them.

- Therefore it may be better to **replace w_{ij} with probabilities** – this is leading to other model(s) – like the **Bayesian networks** or probabilistic Boolean networks.