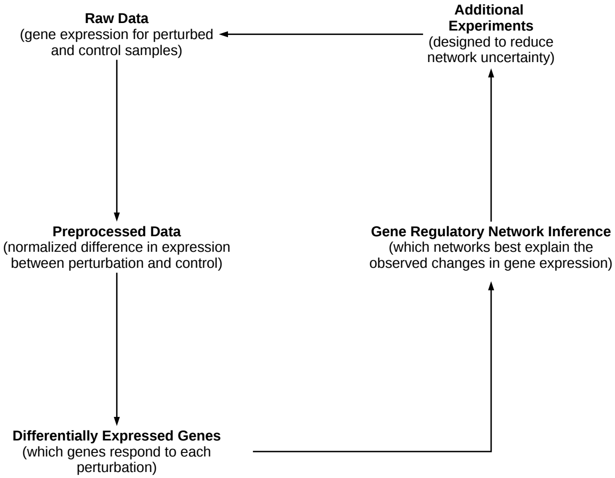
A Computational Bayesian Approach to Gene Regulatory Network Estimation

Matthew N McCall

7/19/2013

Workflow



Raw Data
(gene expression for perturbed and control samples)

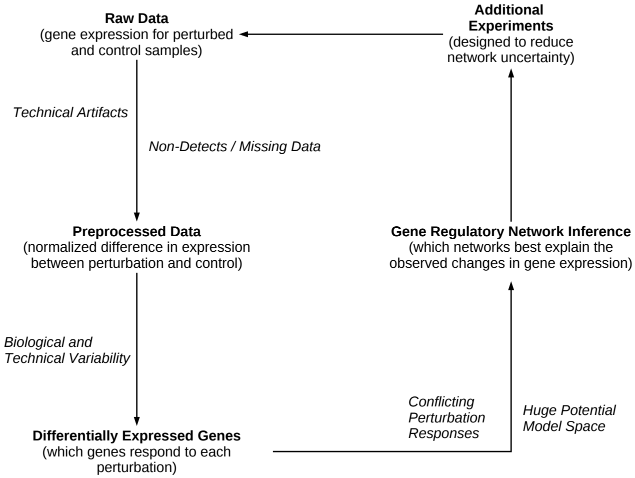
Preprocessed Data
(normalized difference in expression between perturbation and control)

Differentially Expressed Genes
(which genes respond to each perturbation)

Gene Regulatory Network Inference
(which networks best explain the observed changes in gene expression)

Additional Experiments
(designed to reduce network uncertainty)

Workflow



Raw Data
(gene expression for perturbed and control samples)

Technical Artifacts

Non-Detects / Missing Data

Preprocessed Data
(normalized difference in expression between perturbation and control)

Biological and Technical Variability

Differentially Expressed Genes
(which genes respond to each perturbation)

Conflicting Perturbation Responses

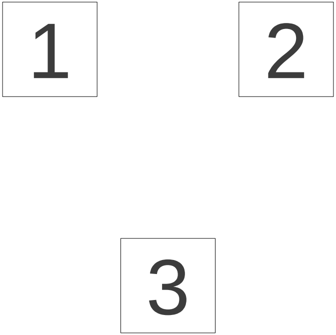
Huge Potential Model Space

Gene Regulatory Network Inference
(which networks best explain the observed changes in gene expression)

Additional Experiments
(designed to reduce network uncertainty)

Toy Example

How does the expression of each of these genes affect the other genes?



Ternary Network Assumptions

- Each gene can take on one of three expression states: -1 (down) / 0 (baseline) / 1 (up).
- The state of the genes in the network at time $t+1$ is a deterministic function of the state of the genes in the network at time t. These functions are called the *transition functions*.
- This implies that from any initial state, the network will eventually reach a set of states that will repeat infinitely often, called an *attractor*.
- The observed data represent a measurement of the steady state behavior of the network – closely related to the attractor structure.

Toy Example

Transition functions:

Gene 1		**Gene 3**		
		-1	0	1
Gene 2	-1	?	?	?
	0	?	0	?
	1	?	?	?

Gene 2		**Gene 3**		
		-1	0	1
Gene 1	-1	?	?	?
	0	?	0	?
	1	?	?	?

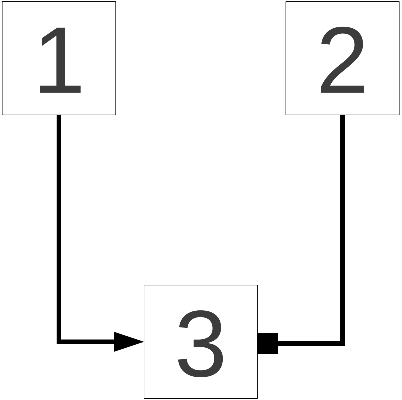
Gene 3		**Gene 2**		
		-1	0	1
Gene 1	-1	?	?	?
	0	?	0	?
	1	?	?	?

Toy Example

Six perturbation experiments. Perturbations are in bold:

	Exp1	Exp2	Exp3	Exp4	Exp5	Exp6
Gene 1	**1**	0	0	**-1**	0	0
Gene 2	0	**1**	0	0	**-1**	0
Gene 3	1	-1	**1**	-1	1	**-1**

Toy Example



Toy Example

Transition functions:

Gene 1		**Gene 3**		
		-1	0	1
Gene 2	-1	?	0	?
	0	0	0	0
	1	?	0	?

Gene 2		**Gene 3**		
		-1	0	1
Gene 1	-1	?	0	?
	0	0	0	0
	1	?	0	?

Gene 3		**Gene 2**		
		-1	0	1
Gene 1	-1	?	-1	?
	0	1	0	-1
	1	?	1	?

Some Inference Challenges

- There are often many network structures that fit the observed data equally well.
- The model space is huge. For the 3 gene example, there are $\approx 2.82x10^{11}$ possible networks. In the ones I work with, there are $\approx 2.48x10^{763}$.
- One approach to this is select the most parsimonious network that fits the observed data well.
- But there is no real reason to assume that a simpler network is better.

Inference

- In most experiments, it is the attractors which are observed, rather than the transition functions. Therefore, it is natural to see the problem as a search for networks constrained by a given set of attractors.
- We consider a computational Bayesian approach in which the model can be scored by comparing its attractors (under experimental perturbations) to the observed data. The objective is therefore to sample from a posterior density of plausible models.

Inference

- Define a distance that measures the compatibility of an attractor with the observed steady state data.
- Search the space of potential networks seeking to minimize the distance measure – we use a variation on simulated annealing.
- Define a posterior density on the model space and sample from that density. The posterior density is approximately a uniform distribution on the networks that achieve a score in the neighborhood of the minimum score.

Real Data Example

- A small set of genes, called cooperation response genes (CRGs), that respond synergistically to loss-of-function p53 and Ras activation. Subsequently, shown to be involved in tumor formation significantly more often than non-CRGs.
- Cells transformed to malignancy by mutant p53 and activated Ras. These cells are subsequently perturbed with the aim of restoring CRG expression to levels found in non-transformed parental cells via retrovirus-mediated re-expression of corresponding cDNAs and shRNA-dependent stable knock-down.
- The current data consist of perturbations of 20 individual CRGs and 3 double perturbations of selected CRG pairs.

Network Input Data

Probabilities of up- / down-regulation in response to perturbations:

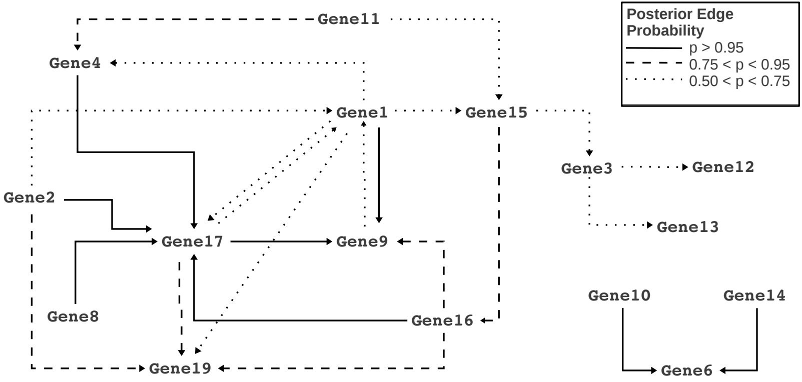
	Ex1	Ex2	Ex3	Ex4	Ex5	Ex6	Ex7	Ex8	Ex9	Ex10	Ex11	Ex12	Ex13	Ex14	Ex15	Ex16	Ex17	Ex18	Ex19	Ex20	Ex21	Ex22	Ex23
Gene1		0.09	-0.04	-0.04	-0.03	-0.03	-0.04	-0.03	0.20	0.09	-0.06	0.12	-0.03	0.10	-1.00	0.10	-0.03	-0.03	-0.03	-0.03		0.09	
Gene2	0.09		0.11	-0.13	0.09	-0.12	0.09	0.12	0.09	0.10	0.10	-0.12	-0.14	0.09	-0.13	0.09	0.10	-0.12	-0.12	0.09	0.13	0.12	0.25
Gene3	0.09	0.09		-0.07	0.09	-0.06	0.09	0.27	0.09	-0.06	0.09	0.09	-0.06	0.09	-0.98	0.09	0.09	0.09	0.09	-0.06	-0.06	0.09	0.11
Gene4	0.10	0.43	0.20		0.08	-0.11	-0.12	0.08	0.20	-0.12	0.12	0.08	-0.14	-0.12	-0.12	0.13	0.08	0.08	0.08	0.08	-0.11	0.11	0.09
Gene5	-0.14	-0.12	-0.09	-0.28		-0.09	-0.09	0.10	-0.12	0.10	-0.09	-0.09	-0.14	-0.09	0.11	-0.10	0.10	-0.09	-0.09	-0.10	0.10	0.11	-0.09
Gene6	-0.03	0.07	0.06	0.06	0.06		0.07	0.07	-0.03	-1.00	0.18	-0.03	-0.03	-1.00	-0.03	0.09	-0.03	-0.35	0.06	-0.03	0.06	0.07	0.07
Gene7	-0.10	-0.11	-0.19	-0.11	0.11	0.10		-0.10	0.13	0.13	-0.10	0.10	-0.11	0.12	-0.10	-0.10	0.10	0.11	0.11	0.11	-0.12	0.10	0.16
Gene8	0.15	0.06	-0.08	-0.09	0.06	0.06	0.07		0.99	-0.08	0.08	0.07	0.06	-0.12	-0.17	0.06	0.05	0.06	0.06	0.06	-0.10	0.06	0.07
Gene9	0.06	-0.12	-0.11	-0.11	0.05	-0.12	-0.15	0.06		0.05	-0.11	0.05	0.05	0.06	0.96	-0.12	-0.11	0.06	0.05	0.07			1.00
Gene10	0.12	0.10	-0.12	-0.14	0.10	-0.12	-0.12	0.19	0.19		-0.12	-0.12	-0.13	-0.13	0.10	0.10	0.10	0.11	0.10	-0.12	-0.13	0.12	0.13
Gene11	0.11	0.29	0.28	-0.10	0.09	-0.10	-0.11	-0.19	0.08	0.11		0.10	-0.12	-0.10	-0.10	0.11	-0.10	-0.10	0.09	0.09	0.09	0.09	0.09
Gene12	0.12	0.11	0.12	-0.18	-0.10	-0.10	-0.10	-0.19	0.11	0.15	0.11		-0.10	0.11	-0.10	0.14	-0.10	-0.10	-0.13	-0.12	0.11	-0.10	0.11
Gene13	-0.20	-0.09	-0.09	-0.34	-0.08	-0.08	-0.08	-0.08	-0.10	0.08	-0.09	-0.14		0.07	-0.12	-0.14	-0.09	0.07	-0.08	-0.08	-0.10	-0.09	0.06
Gene14	0.10	0.10	0.10	0.10	0.14	0.10	0.11	0.15	0.11	-0.16	0.10	0.18	-0.14		-0.11	-0.11	-0.11	-0.11	0.11	0.10	-0.11	-0.11	0.10
Gene15	0.09	-0.08	-0.15	-0.34	-0.11	-0.12	-0.08	0.09	0.11	0.09	0.11	-0.12	-0.13	-0.08		0.09	0.11	-0.10	-0.08	-0.11	0.10	0.11	0.22
Gene16	-0.07	0.11	0.11	-0.07	0.11	-0.06	-0.09	0.12	-0.08	-0.08	-0.07	-0.06	-0.08	0.12	-0.90		-0.06	-0.06	-0.06	-0.06	-0.07	0.13	0.15
Gene17	0.14	-1.00	0.05	0.99	-0.03	-0.03	0.05	1.00	0.05	-1.00	-0.03	-0.03	0.05	0.06	-1.00	-1.00		-0.04	-0.05	-0.04	0.06		
Gene18	-0.09	-0.24	-0.19	-0.35	-0.09	0.10	0.11	-0.08	-0.08	0.11	-0.11	-0.09	-0.09	-0.09	0.16	-0.12	0.11		-0.09	-0.08	-0.08	-0.08	0.10
Gene19	0.03	0.03	0.03	-0.09	0.03	0.03	0.03	-0.06	-0.05	0.04	0.10	-0.06	-0.06	-0.05	1.00	0.03	-0.05	-0.09		0.03	-0.05	0.04	0.03
Gene20	0.13	0.16	0.07	-0.17	-0.11	-0.10	-0.10	0.06	0.89	-0.10	0.07	0.06	-0.16	-0.11	0.08	0.07	0.06	0.06	-0.10		-0.10	0.10	0.07

Network Fit

Posterior probabilities of attractors resulting from 3 perturbations:

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18	G19	G20	Post-Prob
Ex1	**1.00**	**0.09**	**0.09**	**0.10**	**-0.14**	**-0.03**	**-0.10**	**0.15**	**0.06**	**0.12**	**0.11**	**0.12**	**-0.20**	**0.10**	**0.09**	**-0.07**	**0.14**	**-0.09**	**0.03**	**0.13**	–
Ex1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
Ex2	**0.09**	**1.00**	**0.09**	**0.43**	**-0.12**	**0.07**	**-0.11**	**0.06**	**-0.13**	**0.10**	**0.29**	**0.11**	**-0.09**	**0.10**	**-0.08**	**0.11**	**-1.00**	**-0.24**	**0.03**	**0.16**	–
Ex2	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-1.00	0.00	0.00	0.00	0.99
Ex3	**-0.03**	**0.12**	**0.27**	**0.08**	**0.10**	**0.07**	**-0.10**	**1.00**	**0.06**	**0.19**	**-0.19**	**-0.19**	**-0.08**	**0.16**	**0.09**	**0.12**	**1.00**	**-0.08**	**-0.06**	**0.06**	–
Ex3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	-1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.87
Ex3	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.10
Ex3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02

Posterior Edge Probability Graph



Where to Find More Info

- Methodology: Almudevar et al. (2011) SAGMB, 10(1):14
- Biology: McMurray et al. (2008) Nature, 453:1112-1116
- R/BioC package: `ternarynet`
- Website: `mnmccall.com`

Thanks

- Anthony Almudevar
- Helene McMurray
- Hartmut Land