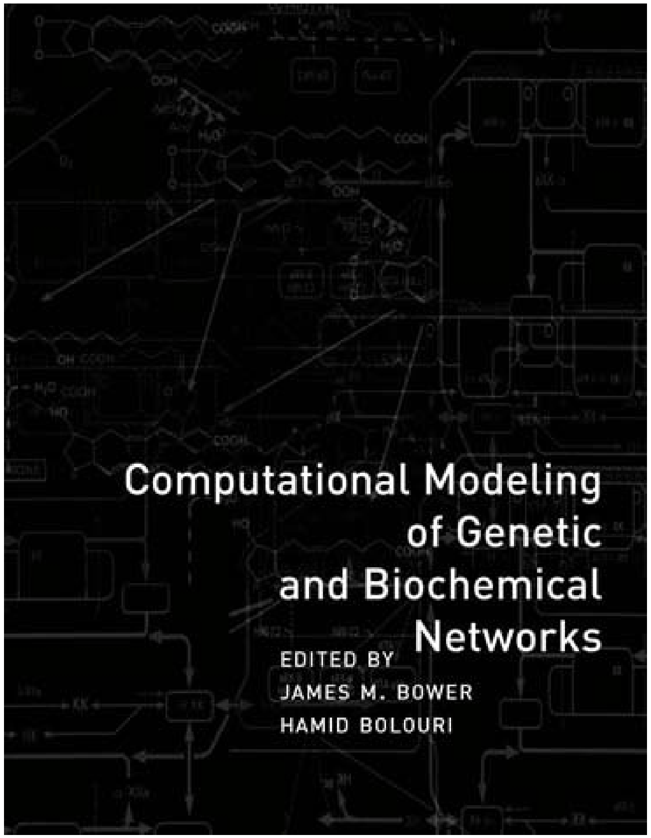
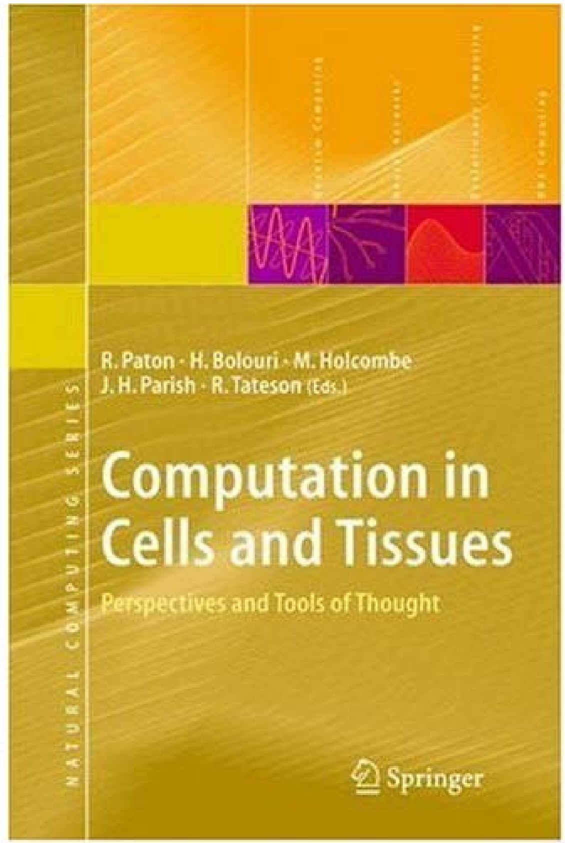
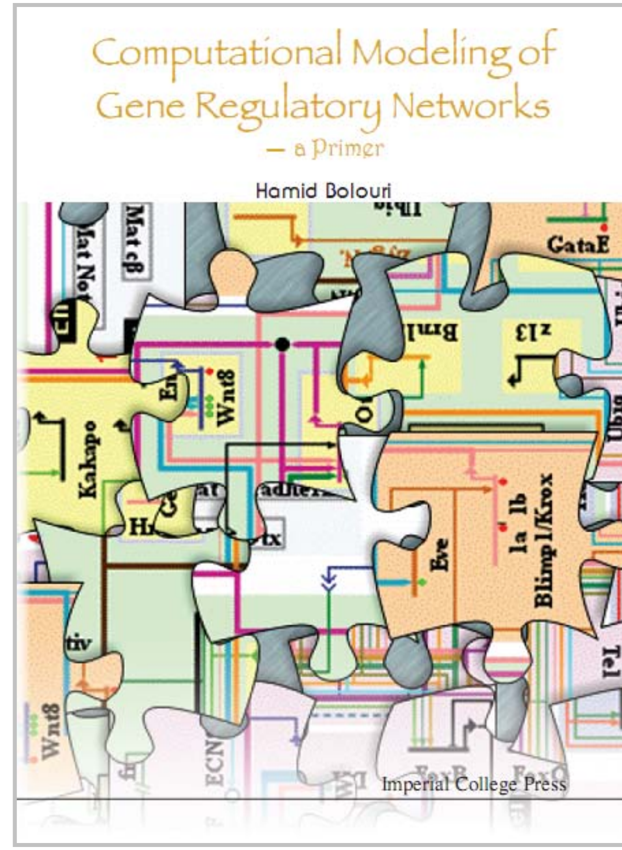
Using intensive workshops for bioinformatics training



2001



2004



2008

http://www.its.caltech.edu/~hbolouri/

MotifMogul

[Mogul Homepage] [FAQ] [Example Run] [News] [Comments]

Organism: yeast (S.cerevisiae)

Upload a reference sequence in Fasta Format: GAL10.fa [Browse...]

Motif Scanners	☑ Fuzznuc	☐ MotifScanner	☐ AHAB
Single Scan	☐ Verbumculus ☐ YMF	☑ PARS [Palindrome] ☑ DREAM	☐ rmes.poisson ☐ rmes.gaussian
Co-regulated Upload sequences in Fasta Format: GAL10-coregulated.fa [Browse...]	☑ AlignACE ☐ AlignACE-Long ☐ MEME ☐ MEME-Long	☐ Sampler ☑ Sampler-Long ☐ MotifSampler ☐ MotifSampler-Long	☐ WConsensus ☑ WConsensus-Long ☑ BioProspector ☐ BioProspector-Long ☑ BioProspector-Palindrome
Comparative: Upload a sequence in Fasta Format: GAL10-bayanus.fa [Browse...]	☑ Seqcomp	☑ DBA	☐ Bayesaligner
Phylogenetic: Upload sequences in Fasta Format: GAL10-phylo.fa		☑ AlignACE-Phylo	☐ MotifSampler-Phylo

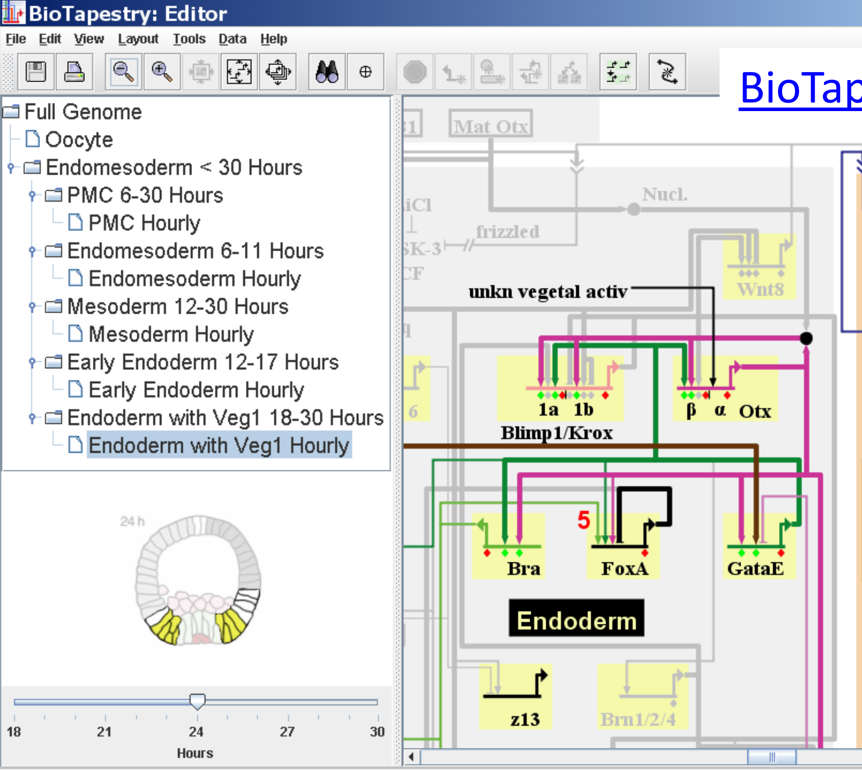
Press [- OK -]

Innate Immune Database (IIDB)

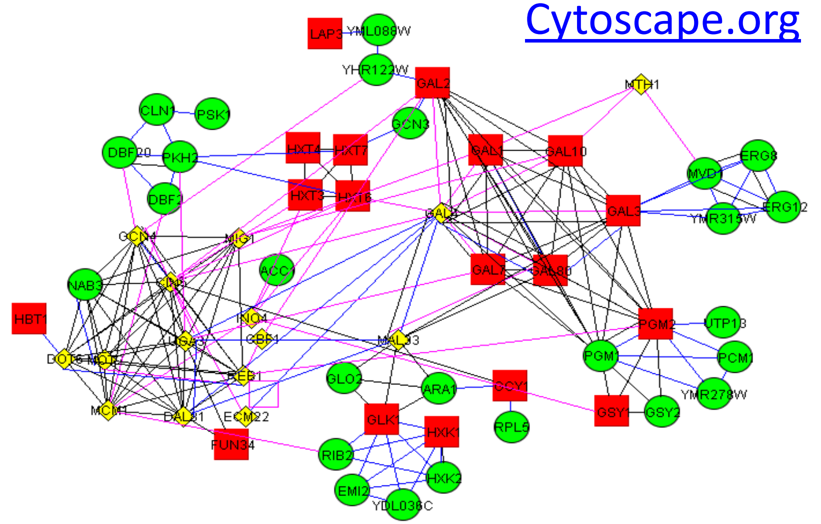
The development of IIDB is supported by a grant from the National Institute of Allergy and Infectious Disease(NIAID), a division of the National Institutes of Health (NIH).

Your Favorite Gene Search for a Gene Gene Aliases **List of Annotated Genes** NCBI mm5 Version 33 ENSEMBL Version 29e	**Computationally Predicted Co-regulated Genes** **ISB Co-regulated Gene Clusters** Annotated by: NCBI | ENSEMBL **LPS Responsive Gene Clusters** from Nilsson R et. al., Supplemental Data Annotated by: NCBI | ENSEMBL	**Advanced Analyses** Search for TFBS Search Genes for Shared TFBS Create Gene Groups by GO - Annotation Get a Sequence File **Explore ChIP-chip Data:** ATF3

IIDB Home | About IIDB | How to Use IIDB | IIDB Tutorial | Site Map | Questions/Contact



BioTapestry.org



Cytoscape.org

SBML.org

The Systems Biology Markup Language

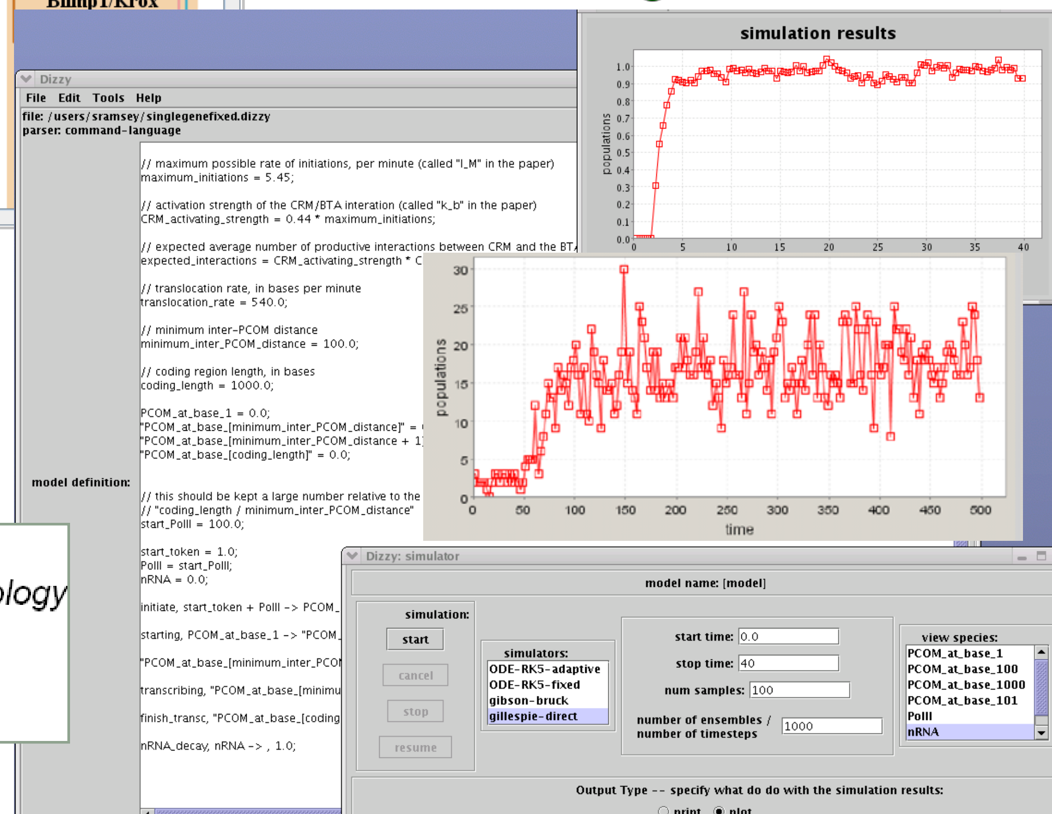
Gene Regulatory Networks for Development

Directors: Eric Davidson, *California Institute of Technology*

David McClay , *Duke University*

Course Date: October 13 - 24, 2009

MBL | Biological Discovery in Woods Hole



http://magnet.systemsbiology.net/software/Dizzy/

Alternative Talk Title:

Mistakes, I've made a few…

Characteristics of bioinformatics in systems biology:

- Inherently cross-disciplinary in content and approach
  - Algorithm design, e.g. sequence assembly and alignment
  - Statistics, e.g. transcription factor binding site prediction
  - Engineering, e.g. simulation modeling, sensitivity analysis
  - Math/physics, e.g. network inference, graph theory
  - …
- Rapidly growing
  - New biotechnologies → new data interpretation needs
  - New bioinformatics tools and methodologies → rapid obsolescence
  - Few degree courses cover the full spectrum of topics
  - Tension between the need for immediate results and fundamental research
  - Need for win-win collaborations among wet and dry scientists

Characteristics of intensive bioinformatics workshops

- a few days to a few weeks (often off-site)

- typical student is a researcher

  grad students, post docs, industry professionals, faculty

- mixture of student backgrounds

  ‘dry’ / ‘desk’ scientists

  (computer scientists, physicists, applied math, statisticians, engineers…)

  &

  ‘wet’ / ‘bench’ scientists

  (molecular/cell biologists, bioengineers, pharmacologists, biochemists…)

- many students have relevant advanced specialists skills

  (in a narrowly focused area)

- students are motivated and intelligent

  uneasy about showing ignorance

  eager to share expertise

Lessons learnt (so far)

- No time to build up conceptual infrastructure

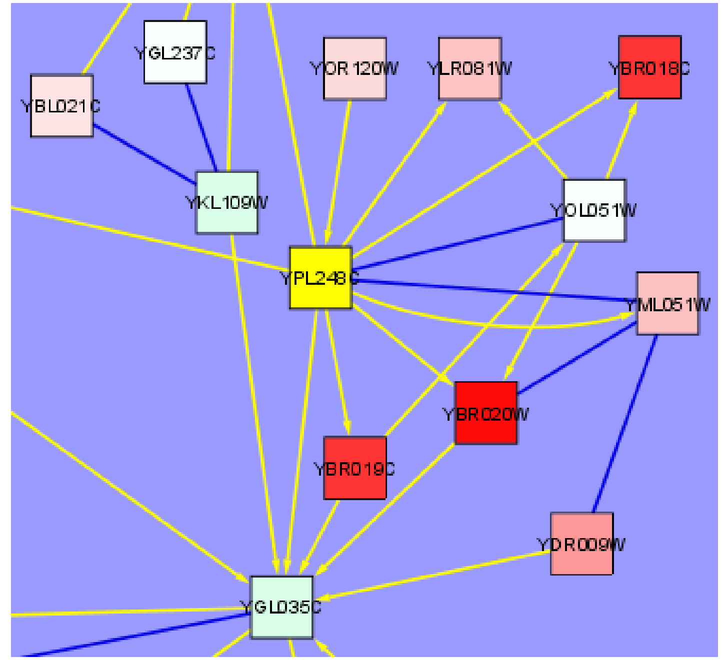
  ∴ Use examples from students' own background to build intuitive understanding

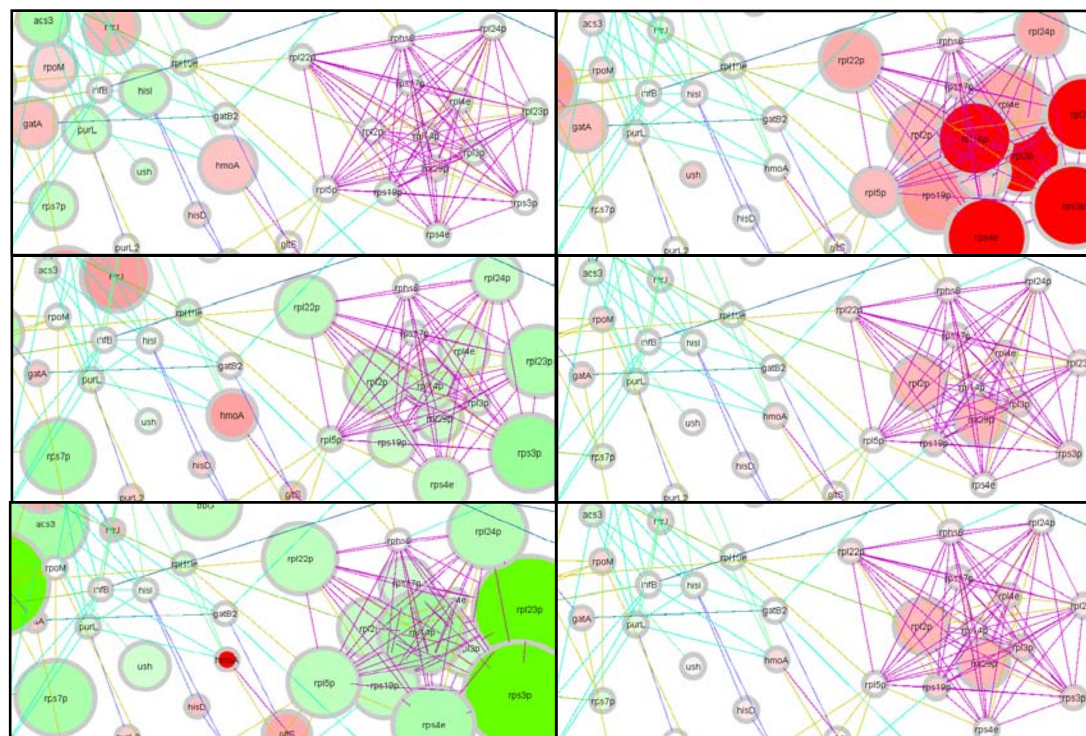
  - E.g. for dry scientists: examples of useful non-mathematical models

Red sky at night,
Sailor's delight;
Red sky at morning,
Sailor's warning.

  - Make an intuitive connection ↑ , then present specific examples ↓

Declarative/data-driven/implicit modeling

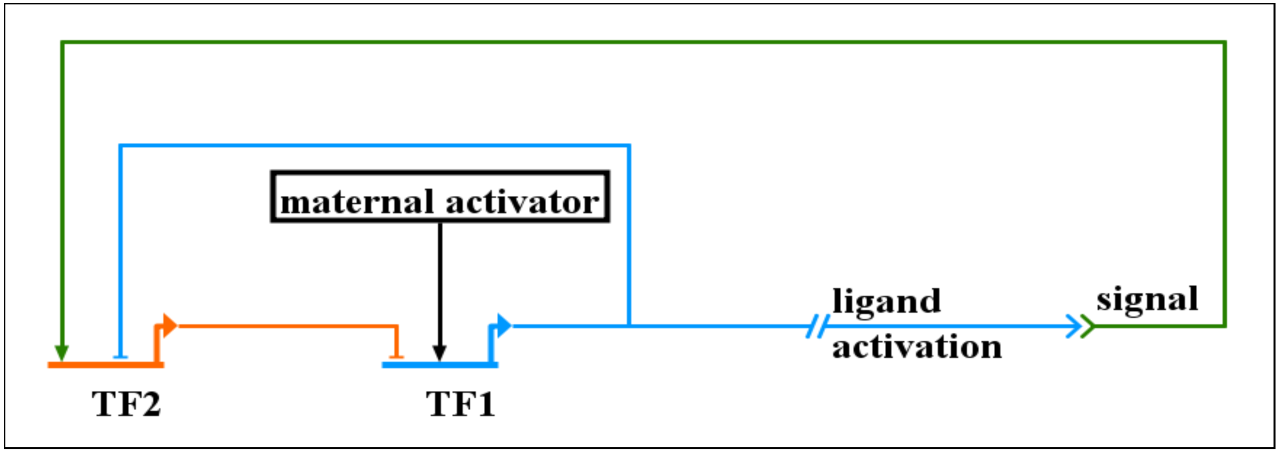




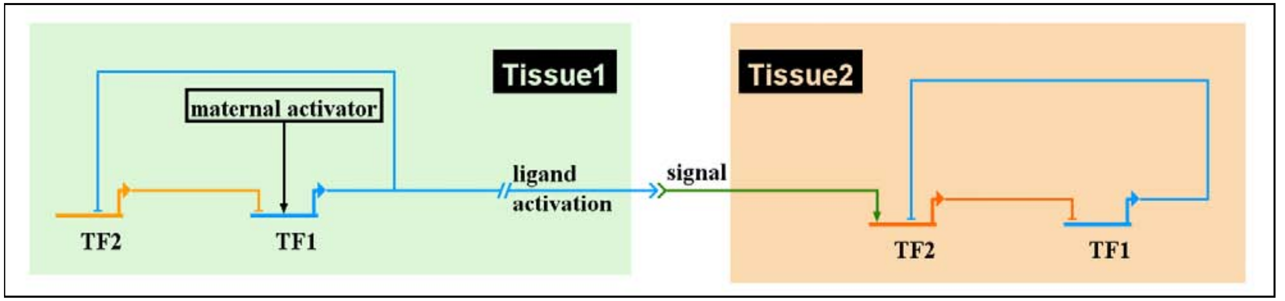
http://cytoscape.org/

Declarative GRN modeling

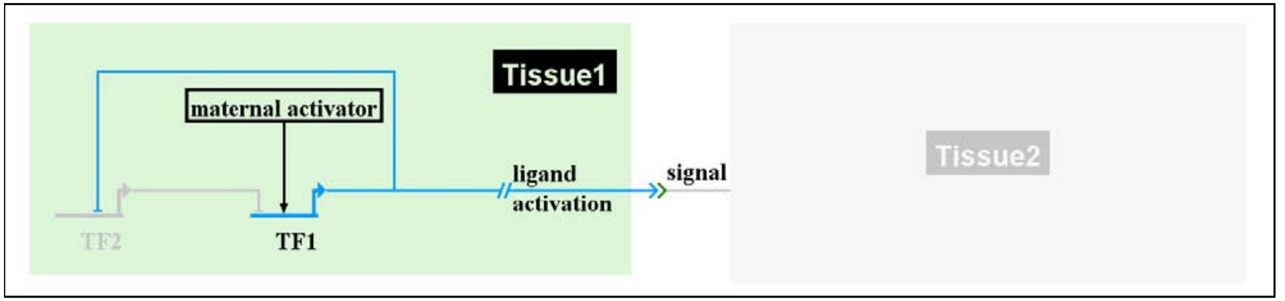
View from the Genome



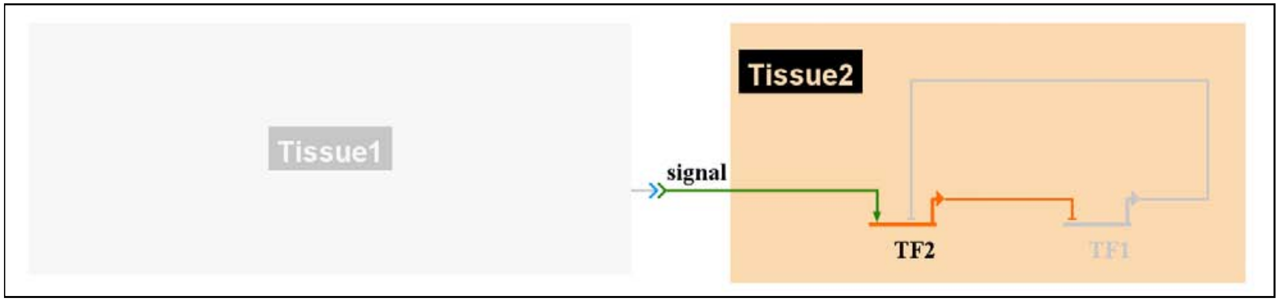
View from All Nuclei



View from T1 Nuclei



View from T2 Nuclei

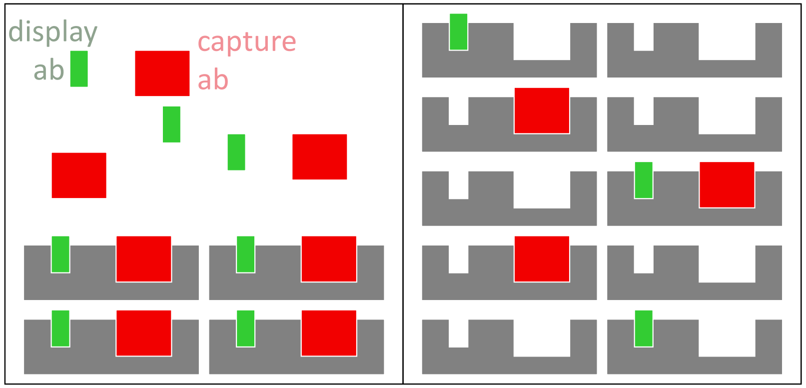


http://www.biotapestry.org/

Lessons learnt

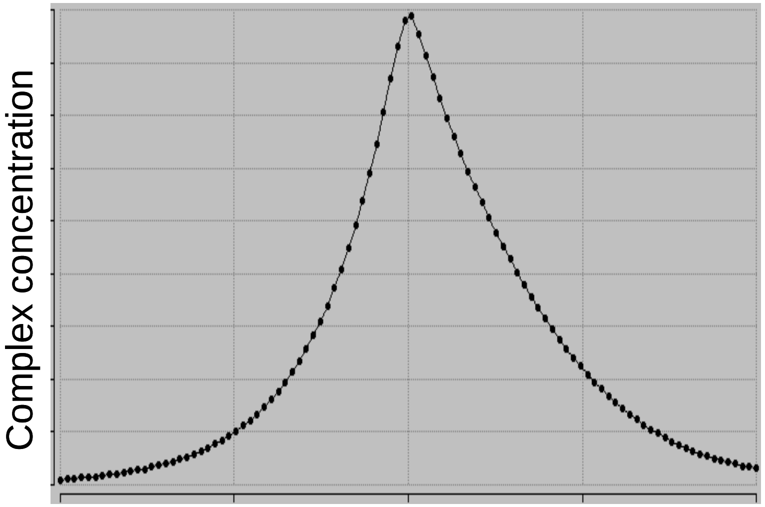
- No time to build up conceptual infrastructure
  - Use examples from students' own background to build intuitive understanding
  - E.g. for bench scientists: the need for mathematical models

Too much analyte in 1-step sandwich immunoassays



[analyte]<<[antibodies] [analyte]>>[antibodies]

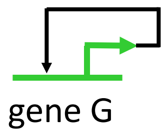
e.g. J Autoimmun (1988) 1: 109-17.



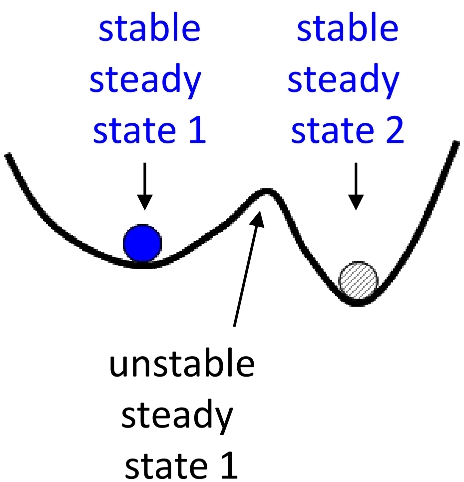
  - Make an intuitive connection ↑ , then present specific examples ↓

Simple gene regulatory systems with complex behavior

Toy model:



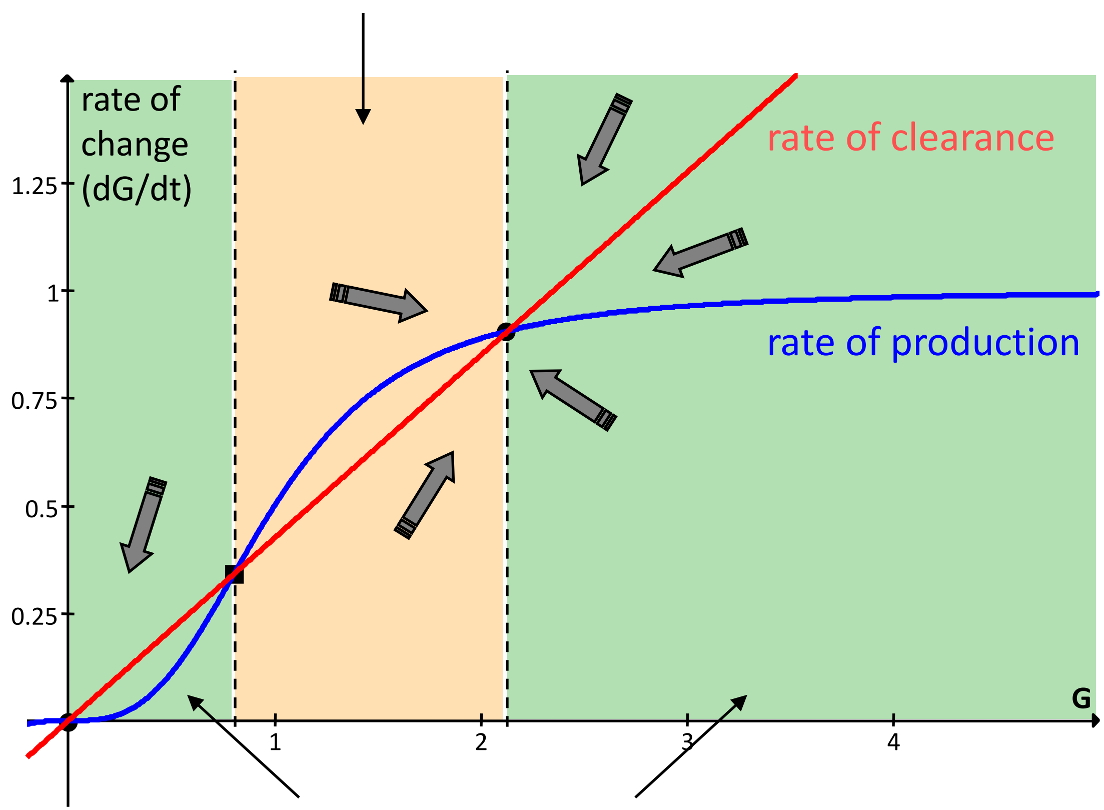
$$\frac{dG}{dt} = k_t \cdot \frac{G^N}{K + G^N} - k_d \cdot G$$



At Steady states,
production rate = clearance rate

- ● stable steady states
- ■ unstable steady state

rate of production > rate of clearance → G will increase over time

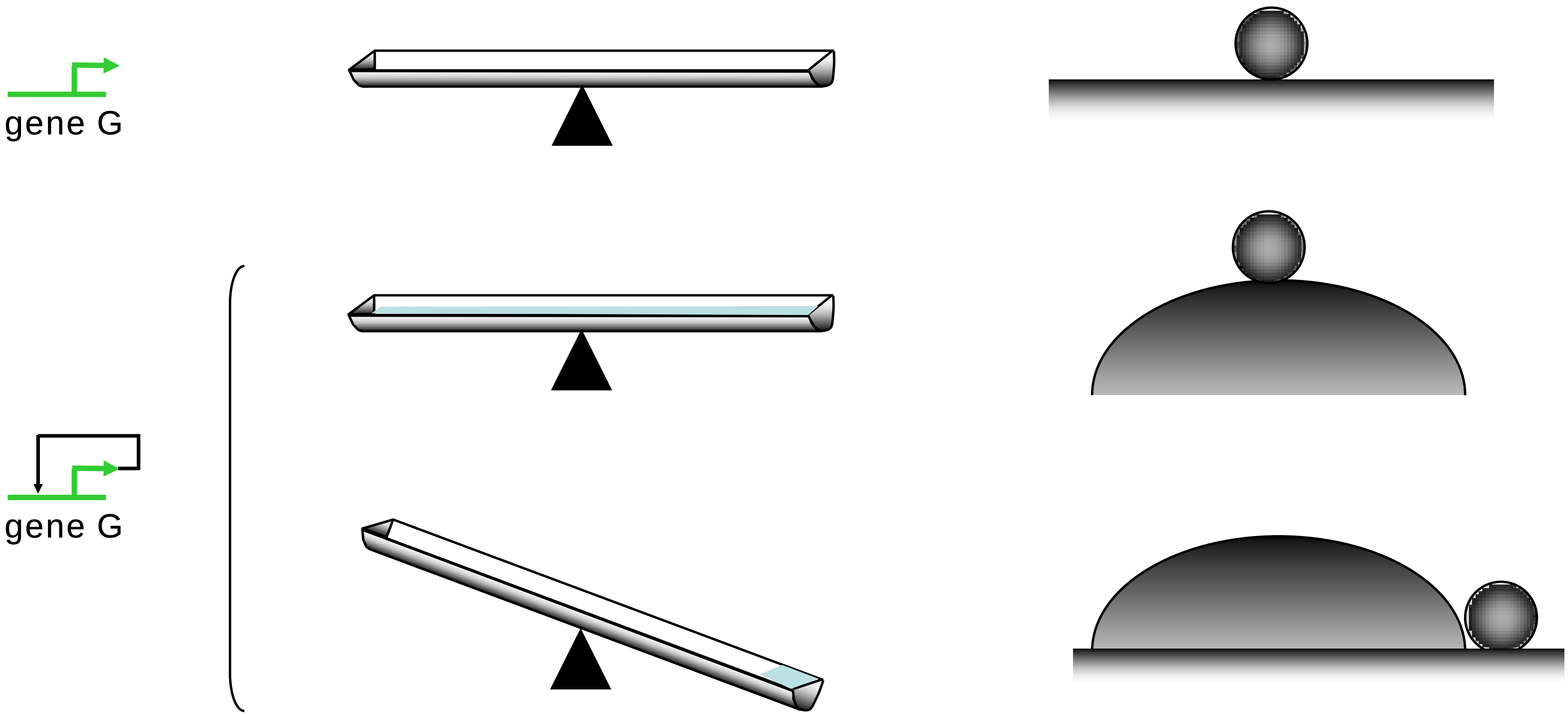


rate of clearance > rate of production → G will decrease over time

Lessons learnt

- Use specific examples to teach general principles
- Follow up each topic with a hands on (computational) lab-exercise
  - Mixed groups of bench and desk scientists
    - Challenges and leadership opportunities for both types
  - Graded difficulty set of objectives
    - Relax those feeling ‘challenged’. Provide a challenge for the (inevitable) expert(s) in the class.
  - Eg1. Data integration using Cytoscape/BioTapestry
    - Curate papers/DBs, interpret data, build network, discover the unexpected
  - Eg2. Searching a sequence for putative transcription factor binding sites
    - Students download weight matrices and sequences, scan sequences, rank findings
  - Eg3. Characterizing network ‘building blocks’ with simulation modeling
    - Use intuitive modeling environments
      - http://magnet.systemsbiology.net/software/Dizzy/ http://www.berkeleymadonna.com/
  - Lab time is a great opportunity to circulate among students and learn their needs

Risky ventures 1: getting caught up in analogies



Add negative feedback (e.g. siRNA)?

Fascinating, but too easy to go off-topic (Rube Goldberg contraptions)

Risky ventures 2:

Assuming background knowledge

Theorist vs. experimentalist jokes.

Experimentalists observe things that cannot be explained
Theoreticians explain things that cannot be observed

The only true science is physics, everything else is stamp collecting

Lab group sizes > 3 (silent and dominant partners)

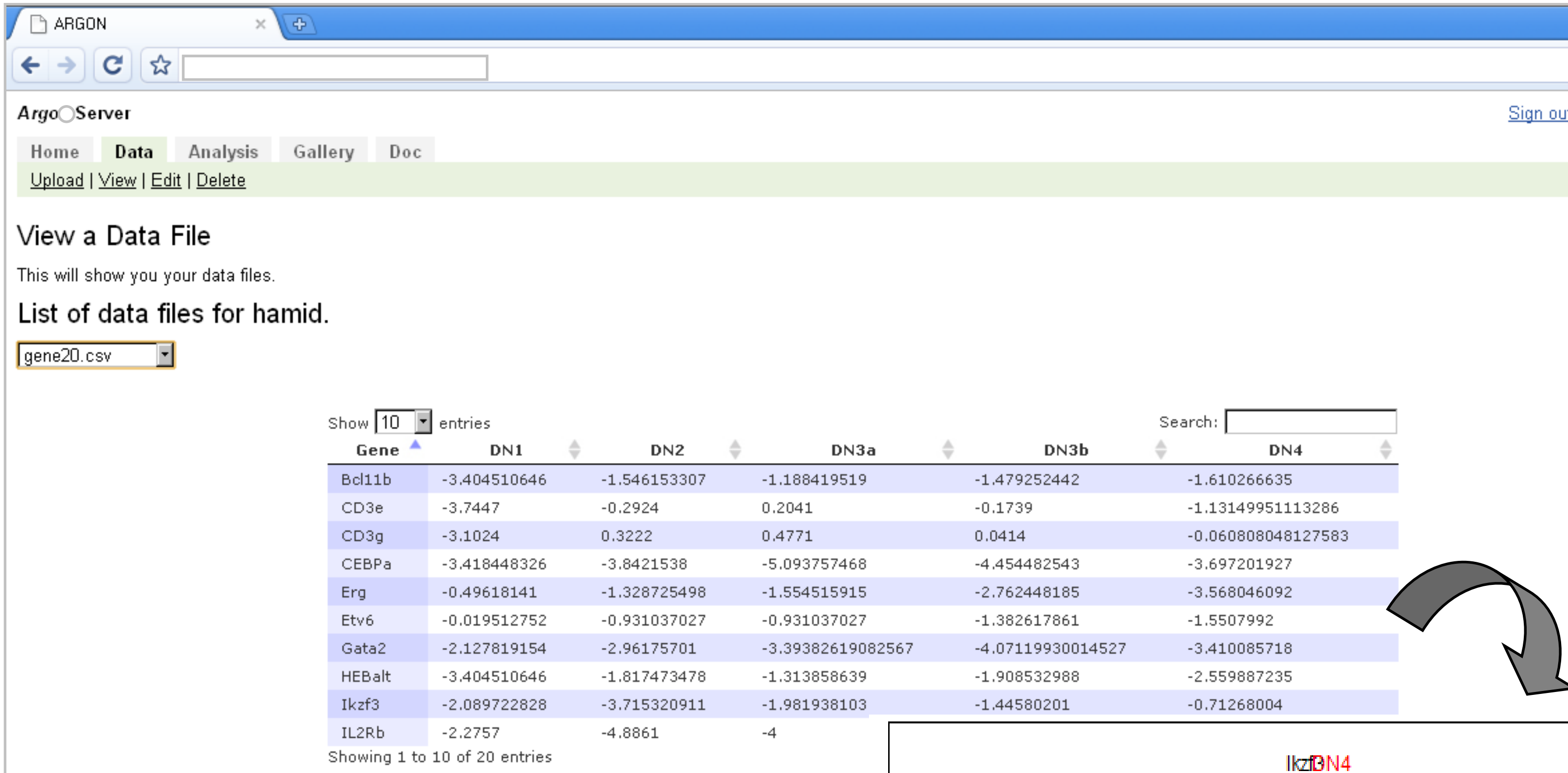
Mini competitions and prizes for good questions

Worth experimenting with:

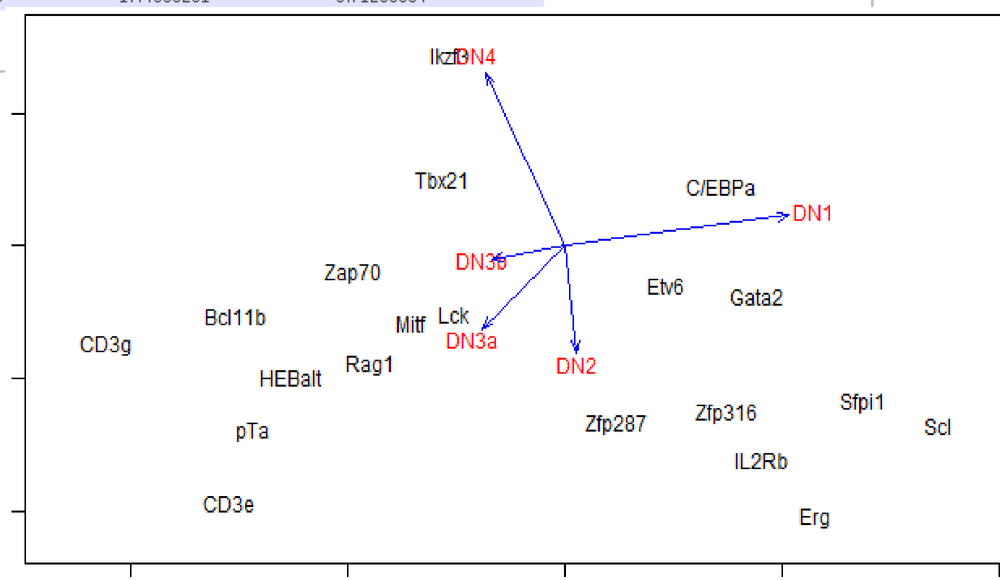
Student project presentations if time/numbers permit

Surgery hours in the bar

Need intuitive, learn-by-doing bioinformatics tools



Gene	DN1	DN2	DN3a	DN3b	DN4
Bcl11b	-3.404510646	-1.546153307	-1.188419519	-1.479252442	-1.610266635
CD3e	-3.7447	-0.2924	0.2041	-0.1739	-1.13149951113286
CD3g	-3.1024	0.3222	0.4771	0.0414	-0.060808048127583
CEBPa	-3.418448326	-3.8421538	-5.093757468	-4.454482543	-3.697201927
Erg	-0.49618141	-1.328725498	-1.554515915	-2.762448185	-3.568046092
Etv6	-0.019512752	-0.931037027	-0.931037027	-1.382617861	-1.5507992
Gata2	-2.127819154	-2.96175701	-3.39382619082567	-4.07119930014527	-3.410085718
HEBalt	-3.404510646	-1.817473478	-1.313858639	-1.908532988	-2.559887235
Ikzf3	-2.089722828	-3.715320911	-1.981938103	-1.44580201	-0.71268004
IL2Rb	-2.2757	-4.8861	-4		



Web-based menu-driven instead of

```
plsf=plsr(ymat~xmat,ncomp=4,scale=T)
covs=cov(cbind((3*ymat),xmat), plsf$scores)
...
```

http://www.its.caltech.edu/~hbolouri/