

Quantitative Systems Immunology Summer School

Gene Regulatory Networks class

Laboratory exercises

Two sets of exercises are provided for this laboratory: data processing for GRN construction (ChIP-seq), and kinetic simulation/analysis of the behavior of a mutually repressing pair of genes determining HSC fate.

I recommend you start with the topic that most interests you and, rather than exploring all exercise options exhaustively, move on to the 2nd exercise roughly half-way through the laboratory session.

The exercises will be available at <http://CRdata.org> after the course so you can continue to explore them. Use the laboratory time to explore those issues that you feel you have not fully understood and would like to discuss with me. I also recommend you work in pairs because two heads are often better than one!

To start:

- Go to CRdata.org and Register as a user.
- After registration, log in to CRdata then go to My Account (button at the top right of the Home page) and set your Default Queue to 'Public' (i.e. the free service provided by CRdata for small jobs).

Feel free to explore the User Guide, or watch some of the videos on the “Take a Tour” page.

In the R Scripts tab, directory Public (under the column heading: Script type) contains 2 sets of laboratory exercises for this class: kinetic simulation (tagged with the labels: kinetic and QSI), and ChIP-seq data processing (tagged with ChIPseq and QSI). To find these files quickly Search for the associated tags (or simply click “view all records” and search visually).

Feel free to view the contents of each file by clicking “view” in the R Scripts tab.

To run a script, go to the Run analyses tab, then click on Add New Job – Wizard. You will be presented with a directory listing from which you can choose the appropriate script. Next, you will be presented with a number of dialogue boxes and choices.

Give a unique name to the run (so that you can find it in the Run Analyses directory later). Click the link “Help with chosen script...” for a brief description of what the script does.

For the inputs to the script, CRdata provides you with brief descriptions and recommended values (which you can change).

Two useful short cuts to use throughout the lab:

1. To run the same script multiple times with different inputs/parameters, you can replicate an earlier run by clicking “Clone” and then change the run settings as desired.
2. Instead of clicking Add new Job – Wizard, try Add New Job – Manual. As long as you remember at least part of the name of the script you want to run, CRdata will fill out the rest, making this a quicker way to launch new jobs.

For the ChIP-seq analysis part of the lab, you will work with the following files:

1. Estimate ChIPseq fragment length
2. ChIPseq coverage Browser Track & significance threshold
3. Generate significant peaks track for ChIPseq
4. Associate ChIPseq peaks with nearest genes

Run the above scripts *in the order listed*.

For DNA fragment length estimation you can choose between control and ctcf IP datasets, and between datasets for mouse chromosomes 10, 11 and 12. For the remaining scripts, we will just use chromosome 10 for simplicity. E.g. for the ChIPseq coverage Browser Track & significance threshold script, the inputs are the choice of experimental dataset to process (ctcf IP or control), and the estimated Average DNA Fragment Length (which you get from running the first script).

You can select different input values and re-run each script to see how the results change. The first time you run a script, you run it with the default (recommended) values. Once you have completed all dialog boxes, Click on Create Job. You will be returned to the Run analyses page and see your Job listed near the top of the page. Use your browser’s Refresh Button to update this screen until the Status of the job changes to Done (green tick mark).

Click on the Results link under the Actions column of the Jobs listing to see the output of the script. After running the ChIPseq coverage Browser Track & significance threshold script, note the significance threshold for the number of reads (nread), which you will need as an input value for the subsequent steps.

N.B. I have placed sample output files generated by scripts 2-4 in the Public data directory.

To see the Genome Browser track file that the script generated, and to explore the track on the UCSC Genome Browser, do the following:

- Go to the Data tab in CRdata.
- Click on the "My Private Files" directory name.
- Find the data file named something like:
 job-000000XXX-coverageOnChr10.bedGraph
- Click on the file name.
- Right click on the Hyperlink that appears and use "Save Link As" to save the file to your PC.
- Go to: <http://genome.ucsc.edu/>, then click on Genome Browser near the top left.
- Select the Mouse genome, assembly version July 2007 (NCBI37/mm9) and then click the button "Manage Custom Tracks".
- Click the Add Custom Tracks Button.
- Click Choose File and upload your file, then click Submit.
- In the Genome Browser page
 - Set your "R Track" display to Full.
 - Set the Evolutionary Conservation track's display to Pack and everything else to Dense.
- Click on Manage Custom Tracks.
- Click on "R Track", then in the Edit Configuration dialog box, paste this text:
 track name="ctcf ChIPseq" type=bedGraph visibility=full autoScale=on
 color=0,180,0 viewLimits=0:50
- Go back to the Genome Browser page and type "chr10" in the "position/search" box, then hit return.
- You should see a green histogram of ctcf reads per nucleotide at the top of the browser display window.
- You can zoom into a part of the chromosome by left-click-dragging the cursor on the chromosomal coordinate numbers at the top of the display.

Repeat the above procedures (as needed) for scripts 3 and 4.

Questions:

How consistent is the DNA fragment length estimation across chromosomes? How and why do the plots for control look different from the ctcf IP data?

How does the estimated average fragment length affect significant peak selection?

How could the association of peaks with candidate genes be improved?

For the kinetic modeling part of the lab, you will find three sets of simulation scripts in the Public directory.

1. The script Laslo et al Mac vs Neutrophil cell fate switch performs simple a time course simulation of the model described in the Supplementary Materials of Laslo et al, Cell 2006, 126:755–766 (see equations S2a, S2b, page 17). The equations are:

$$\frac{d(PU1)}{dt} = \frac{e_p}{1 + Gfi^{nr}} - PU1$$

$$\frac{d(Egr)}{dt} = \left(\frac{\alpha \cdot PU1}{1 + PU1} \right) \left(\frac{1}{1 + Gfi^{nr}} \right) - Egr$$

$$\frac{d(Gfi)}{dt} = \left(\frac{\alpha \cdot CEBP}{1 + CEBP} \right) \left(\frac{1}{1 + Egr^{nr}} \right) - Gfi$$

Here C/EBP α is held constant (at a user specified value, e_c) throughout the simulation. PU.1 starts low and is then switched to a high value. You can choose the values and timing of PU.1 activity. At the default parameter settings, the early period of the simulation (during which both C/EBP α and PU.1 are expressed at low levels) corresponds to the ‘primed’ state discussed in Laslo et al.

Change the level of C/EBP α from 0.1 to 0.5. You will note that the system no longer switches to (Egr > Gfi) when PU.1 is increased to 2. *To what value do you need to increase PU.1 in order to switch the state of system from (Gfi > Egr) to (Gfi < Egr)?*

Explore how combinations of C/EBP α and PU.1 expression levels and kinetic parameter values affect the behavior of the model.

2. To better understand the model of Laslo et al, I recommend you explore the much simpler mutual repression model of Cherry and Adler (J. theor. Biol. (2000) 203, 117–133). There are 4 scripts.
 - Cherry & Adler mutual repression GRN time-course simulation
 - Cherry & Adler - parameter requirements for bistability
 - Cherry & Adler steady states trajectories
 - Mutual Repression On-Off Thresholds (Cherry&Adler)

Start with the Cherry & Adler mutual repression GRN time-course simulation script. You can choose the initial conditions and parameter values and see how they affect the kinetic and steady state behavior of the system. The equations are:

$$\frac{dx}{dt} = \frac{k_1}{1 + y^n} - \mu_1 x$$

$$\frac{dy}{dt} = \frac{k_2}{1 + x^n} - \mu_2 y$$

You will note that for some parameter combinations, the above system is NOT bistable. The script Cherry & Adler - parameter requirements for bistability visualizes the parameter ranges which result in bistability (as described in the paper and my lecture). Using the output from this script, choose parameter combinations that you would expect to produce bistable and monostable systems, then test your hypotheses by running time course simulations with these parameter values.

The script Cherry & Adler steady states trajectories generates a summary plot that shows how the system behaves starting from a wide range of initial conditions. Explore how the system behavior changes with different parameter choices. For example, *what happens when you change the cooperativity parameter “n” to 1?*

As discussed in the lecture, bistable systems (such as this model) exhibit differing turning on and off thresholds (hysteresis). The script Mutual Repression On-Off Thresholds (Cherry&Adler) demonstrates this property when run with the default parameter settings. Change k1 and k2 to 10 to see another interesting behavior (refer to lecture slides for explanation).

3. An intriguing alternative model of how the steady states of mutually repressing genes may be switched is presented in Saka & Smith, BMC Developmental Biology 2007, 7:47. Here the mutually repressing genes are both downstream of a single regulator. Saka & Smith were particularly interested in the case where the shared regulator is a morphogen, but the model is general and the shared input (called M in the model) could just as easily be a cytokine or other signal.

The equations are:

$$\frac{dA}{dt} = \frac{k_a}{1 + B^\beta} \cdot \frac{M^\mu}{1 + M^\mu} - kd_a \cdot A$$

$$\frac{dB}{dt} = \frac{k_b}{1 + A^\alpha} \cdot \frac{M^\mu}{1 + M^\mu} - kd_b \cdot B$$

Note: For simplicity, Saka & Smith set $kd_a=kd_b=1$. The effects of changing kd_a and kd_b can be mimicked by changing k_a and k_b in the opposite direction.

There are four scripts for you to explore:

- Saka & Smith bistable morphogen detector model
This script performs a time-course simulation
- Saka & Smith model, plot trajectories & nullclines
This script plots the trajectories from a variety of initial states to steady state. It also plots the steady loci of the two genes so that you can verify the locations of the system steady states (crossing points of the 2 loci).
- Saka & Smith mutual repression GRN, scan M
This script varies the common input M across a user-specified range of values and plots the resulting steady states. With the default parameter settings, you should see the switch flip at about M=1.3.
- Saka & Smith morphogen-driven bistable array of cells
This script is essentially the same as preceding script, except that M is assumed to be a diffusible morphogen whose concentration is distributed exponentially across an array of cells. The simulation results are presented as both a graph and also a 2D plot mimicking a Fluorescence In-Situ Hybridization experiment.

How is it possible for a single input (M) to result in (activity of A) > (activity of B) at one concentration and (activity of A) < (activity of B) at another concentration?