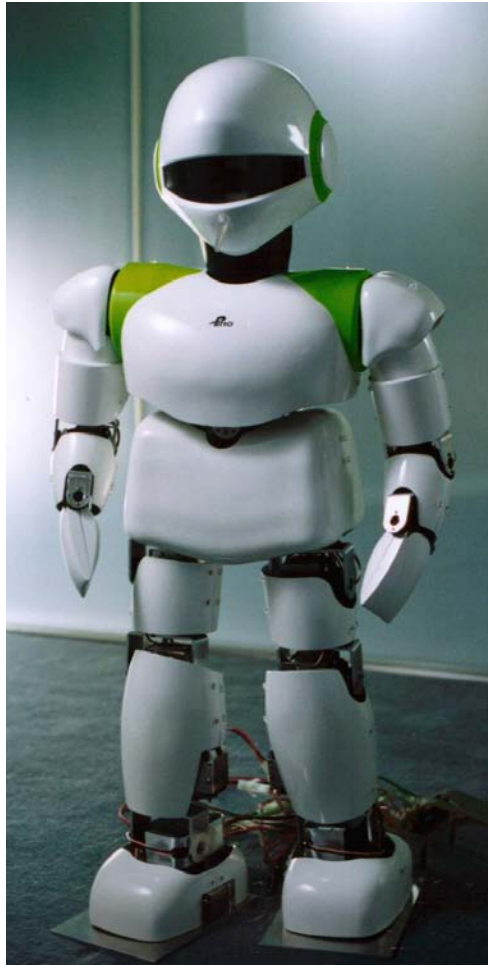


## Tele-presence for the next SBML re-union?



My thanks to:

Hiroaki Kitano, John Doyle *for their vision and support*

Michael Hucka *for 10 years of sustained, intelligent leadership*

Nicholas Le Novere *for BioModels & this meeting!*

Andrew Finney, Herbert Sauro, **and ALL SBML contributors**

# Intrinsic noise, dissipation cost, and robustness of cellular networks: The underlying energy landscape of MAPK signal transduction

Saul Lapidus\*, Bo Han\*, and Jin Wang\*\*\*

PNAS | April 22, 2008 | vol. 105 | no. 16 | 6039–6044

$$\begin{aligned} \frac{d\langle x_1^2 \rangle}{dt} &= -2a_1\langle x_1^2 \rangle \langle x_2 \rangle + a_1\langle x_1 \rangle \langle x_2 \rangle + d_1\langle x_3 \rangle + 2d_1\langle x_1 \rangle \langle x_3 \rangle + k_2\langle x_6 \rangle + 2k_2\langle x_1 \rangle \langle x_6 \rangle \\ \frac{d\langle x_2^2 \rangle}{dt} &= -2a_1\langle x_2^2 \rangle \langle x_1 \rangle + a_1\langle x_1 \rangle \langle x_2 \rangle + d_1\langle x_3 \rangle + k_1\langle x_3 \rangle + 2d_1\langle x_2 \rangle \langle x_3 \rangle + 2k_1\langle x_2 \rangle \langle x_3 \rangle \\ \frac{d\langle x_3^2 \rangle}{dt} &= -2d_1\langle x_3^2 \rangle - 2k_1\langle x_3^2 \rangle + a_1\langle x_1 \rangle \langle x_2 \rangle + d_1\langle x_3 \rangle + k_1\langle x_3 \rangle + 2a_1\langle x_1 \rangle \langle x_2 \rangle \langle x_3 \rangle \\ \frac{d\langle x_4^2 \rangle}{dt} &= k_1\langle x_3 \rangle + 2k_1\langle x_3 \rangle \langle x_4 \rangle - 2a_2\langle x_4^2 \rangle \langle x_5 \rangle + a_2\langle x_4 \rangle \langle x_5 \rangle + d_2\langle x_6 \rangle + 2d_2\langle x_4 \rangle \langle x_6 \rangle \\ &\quad - 2a_3\langle x_4^2 \rangle \langle x_7 \rangle \\ &\quad + a_3\langle x_4 \rangle \langle x_7 \rangle + d_3\langle x_8 \rangle + k_3\langle x_8 \rangle + 2d_3\langle x_4 \rangle \langle x_8 \rangle + 2k_3\langle x_4 \rangle \langle x_8 \rangle - 2a_5\langle x_4^2 \rangle \langle x_9 \rangle + a_5\langle x_4 \rangle \langle x_9 \rangle \\ &\quad + d_5\langle x_{12} \rangle \\ &\quad + k_5\langle x_{12} \rangle + 2d_5\langle x_4 \rangle \langle x_{12} \rangle + 2k_5\langle x_4 \rangle \langle x_{12} \rangle \\ \frac{d\langle x_5^2 \rangle}{dt} &= -2a_2\langle x_5^2 \rangle \langle x_4 \rangle + a_2\langle x_4 \rangle \langle x_5 \rangle + d_2\langle x_6 \rangle + k_2\langle x_6 \rangle + 2d_2\langle x_5 \rangle \langle x_6 \rangle + 2k_2\langle x_5 \rangle \langle x_6 \rangle \\ \frac{d\langle x_6^2 \rangle}{dt} &= -2d_2\langle x_6^2 \rangle - 2k_2\langle x_6^2 \rangle + a_2\langle x_4 \rangle \langle x_5 \rangle + d_2\langle x_6 \rangle + k_2\langle x_6 \rangle + 2a_2\langle x_4 \rangle \langle x_5 \rangle \langle x_6 \rangle \\ \frac{d\langle x_7^2 \rangle}{dt} &= -2a_3\langle x_7^2 \rangle \langle x_4 \rangle + a_3\langle x_4 \rangle \langle x_7 \rangle + d_3\langle x_8 \rangle + 2d_3\langle x_7 \rangle \langle x_8 \rangle + k_4\langle x_{11} \rangle + 2k_4\langle x_7 \rangle \langle x_{11} \rangle \\ \frac{d\langle x_8^2 \rangle}{dt} &= -2d_3\langle x_8^2 \rangle - 2k_3\langle x_8^2 \rangle + a_3\langle x_4 \rangle \langle x_7 \rangle + d_3\langle x_8 \rangle + k_3\langle x_8 \rangle + 2a_3\langle x_4 \rangle \langle x_7 \rangle \langle x_8 \rangle \\ \frac{d\langle x_9^2 \rangle}{dt} &= -2a_5\langle x_9^2 \rangle \langle x_4 \rangle + k_3\langle x_8 \rangle + a_5\langle x_4 \rangle \langle x_9 \rangle + 2k_3\langle x_8 \rangle \langle x_9 \rangle - 2a_4\langle x_9^2 \rangle \langle x_{10} \rangle + a_4\langle x_9 \rangle \langle x_{10} \rangle \\ &\quad + d_4\langle x_{11} \rangle \\ &\quad + 2d_4\langle x_9 \rangle \langle x_{11} \rangle + d_5\langle x_{12} \rangle + 2d_5\langle x_9 \rangle \langle x_{12} \rangle + k_6\langle x_{14} \rangle + 2k_6\langle x_9 \rangle \langle x_{14} \rangle \\ \frac{d\langle x_{10}^2 \rangle}{dt} &= -2a_4\langle x_{10}^2 \rangle \langle x_9 \rangle + a_4\langle x_9 \rangle \langle x_{10} \rangle + d_4\langle x_{11} \rangle + k_4\langle x_{11} \rangle + 2d_4\langle x_{10} \rangle \langle x_{11} \rangle + 2k_4\langle x_{10} \rangle \langle x_{11} \rangle \\ &\quad - 2a_6\langle x_{10}^2 \rangle \langle x_{13} \rangle + a_6\langle x_{10} \rangle \langle x_{13} \rangle + d_6\langle x_{14} \rangle + k_6\langle x_{14} \rangle + 2d_6\langle x_{10} \rangle \langle x_{14} \rangle + 2k_6\langle x_{10} \rangle \langle x_{14} \rangle \\ \frac{d\langle x_{11}^2 \rangle}{dt} &= -2d_4\langle x_{11}^2 \rangle - 2k_4\langle x_{11}^2 \rangle + a_4\langle x_9 \rangle \langle x_{10} \rangle + d_4\langle x_{11} \rangle + k_4\langle x_{11} \rangle + 2a_4\langle x_9 \rangle \langle x_{10} \rangle \langle x_{11} \rangle \\ \frac{d\langle x_{12}^2 \rangle}{dt} &= -2d_5\langle x_{12}^2 \rangle - 2k_5\langle x_{12}^2 \rangle + a_5\langle x_4 \rangle \langle x_9 \rangle + d_5\langle x_{12} \rangle + k_5\langle x_{12} \rangle + 2a_5\langle x_4 \rangle \langle x_9 \rangle \langle x_{12} \rangle \\ \frac{d\langle x_{13}^2 \rangle}{dt} &= -2a_6\langle x_{13}^2 \rangle \langle x_{10} \rangle + k_5\langle x_{12} \rangle + a_6\langle x_{10} \rangle \langle x_{13} \rangle + 2k_5\langle x_{12} \rangle \langle x_{13} \rangle + d_6\langle x_{14} \rangle + 2d_6\langle x_{13} \rangle \langle x_{14} \rangle \\ &\quad - 2a_7\langle x_{13}^2 \rangle \langle x_{15} \rangle + a_7\langle x_{13} \rangle \langle x_{15} \rangle + d_7\langle x_{16} \rangle + k_7\langle x_{16} \rangle + 2d_7\langle x_{13} \rangle \langle x_{16} \rangle + 2k_7\langle x_{13} \rangle \langle x_{16} \rangle \\ &\quad - 2a_9\langle x_{13}^2 \rangle \langle x_{17} \rangle \\ &\quad + a_9\langle x_{13} \rangle \langle x_{17} \rangle + d_9\langle x_{20} \rangle + k_9\langle x_{20} \rangle + 2d_9\langle x_{13} \rangle \langle x_{20} \rangle + 2k_9\langle x_{13} \rangle \langle x_{20} \rangle \\ \frac{d\langle x_{14}^2 \rangle}{dt} &= -2d_6\langle x_{14}^2 \rangle - 2k_6\langle x_{14}^2 \rangle + a_6\langle x_{10} \rangle \langle x_{13} \rangle + d_6\langle x_{14} \rangle + k_6\langle x_{14} \rangle + 2a_6\langle x_{10} \rangle \langle x_{13} \rangle \langle x_{14} \rangle \\ \frac{d\langle x_{15}^2 \rangle}{dt} &= -2a_7\langle x_{15}^2 \rangle \langle x_{13} \rangle + a_7\langle x_{13} \rangle \langle x_{15} \rangle + d_7\langle x_{16} \rangle + 2d_7\langle x_{15} \rangle \langle x_{16} \rangle + k_8\langle x_{19} \rangle + 2k_8\langle x_{15} \rangle \langle x_{19} \rangle \\ \frac{d\langle x_{16}^2 \rangle}{dt} &= -2d_7\langle x_{16}^2 \rangle - 2k_7\langle x_{16}^2 \rangle + a_7\langle x_{13} \rangle \langle x_{15} \rangle + d_7\langle x_{16} \rangle + k_7\langle x_{16} \rangle + 2a_7\langle x_{13} \rangle \langle x_{15} \rangle \langle x_{16} \rangle \\ \frac{d\langle x_{17}^2 \rangle}{dt} &= -2a_9\langle x_{17}^2 \rangle \langle x_{13} \rangle + k_7\langle x_{16} \rangle + a_9\langle x_{13} \rangle \langle x_{17} \rangle + 2k_7\langle x_{16} \rangle \langle x_{17} \rangle - 2a_8\langle x_{17}^2 \rangle \langle x_{18} \rangle \\ &\quad + a_8\langle x_{17} \rangle \langle x_{18} \rangle + d_8\langle x_{19} \rangle + 2d_8\langle x_{17} \rangle \langle x_{19} \rangle + d_9\langle x_{20} \rangle + 2d_9\langle x_{17} \rangle \langle x_{20} \rangle + k_{10}\langle x_{22} \rangle + 2k_{10}\langle x_{17} \rangle \langle x_{22} \rangle \\ \frac{d\langle x_{18}^2 \rangle}{dt} &= -2a_8\langle x_{18}^2 \rangle \langle x_{17} \rangle + a_8\langle x_{17} \rangle \langle x_{18} \rangle + d_8\langle x_{19} \rangle + k_8\langle x_{19} \rangle + 2d_8\langle x_{18} \rangle \langle x_{19} \rangle + 2k_8\langle x_{18} \rangle \langle x_{19} \rangle \\ &\quad - 2a_{10}\langle x_{18}^2 \rangle \langle x_{21} \rangle + a_{10}\langle x_{18} \rangle \langle x_{21} \rangle + d_{10}\langle x_{22} \rangle + k_{10}\langle x_{22} \rangle + 2d_{10}\langle x_{18} \rangle \langle x_{22} \rangle + 2k_{10}\langle x_{18} \rangle \langle x_{22} \rangle \\ \frac{d\langle x_{19}^2 \rangle}{dt} &= -2d_8\langle x_{19}^2 \rangle - 2k_8\langle x_{19}^2 \rangle + a_8\langle x_{17} \rangle \langle x_{18} \rangle + d_8\langle x_{19} \rangle + k_8\langle x_{19} \rangle + 2a_8\langle x_{17} \rangle \langle x_{18} \rangle \langle x_{19} \rangle \\ \frac{d\langle x_{20}^2 \rangle}{dt} &= -2d_9\langle x_{20}^2 \rangle - 2k_9\langle x_{20}^2 \rangle + a_9\langle x_{13} \rangle \langle x_{17} \rangle + d_9\langle x_{20} \rangle + k_9\langle x_{20} \rangle + 2a_9\langle x_{13} \rangle \langle x_{17} \rangle \langle x_{20} \rangle \\ \frac{d\langle x_{21}^2 \rangle}{dt} &= -2a_{10}\langle x_{21}^2 \rangle \langle x_{18} \rangle + k_9\langle x_{20} \rangle + a_{10}\langle x_{18} \rangle \langle x_{21} \rangle + 2k_9\langle x_{20} \rangle \langle x_{21} \rangle + d_{10}\langle x_{22} \rangle \\ &\quad + 2d_{10}\langle x_{21} \rangle \langle x_{22} \rangle \\ \frac{d\langle x_{22}^2 \rangle}{dt} &= -2d_{10}\langle x_{22}^2 \rangle - 2k_{10}\langle x_{22}^2 \rangle + a_{10}\langle x_{18} \rangle \langle x_{21} \rangle + d_{10}\langle x_{22} \rangle + k_{10}\langle x_{22} \rangle + 2a_{10}\langle x_{18} \rangle \langle x_{21} \rangle \langle x_{22} \rangle \end{aligned}$$

# Functional impact of global rare copy number variation in autism spectrum disorders

A list of authors and their affiliations appears at the end of the paper.

doi: 10.1038/nature09146

nature

SUPPLEMENTARY INFORMATION

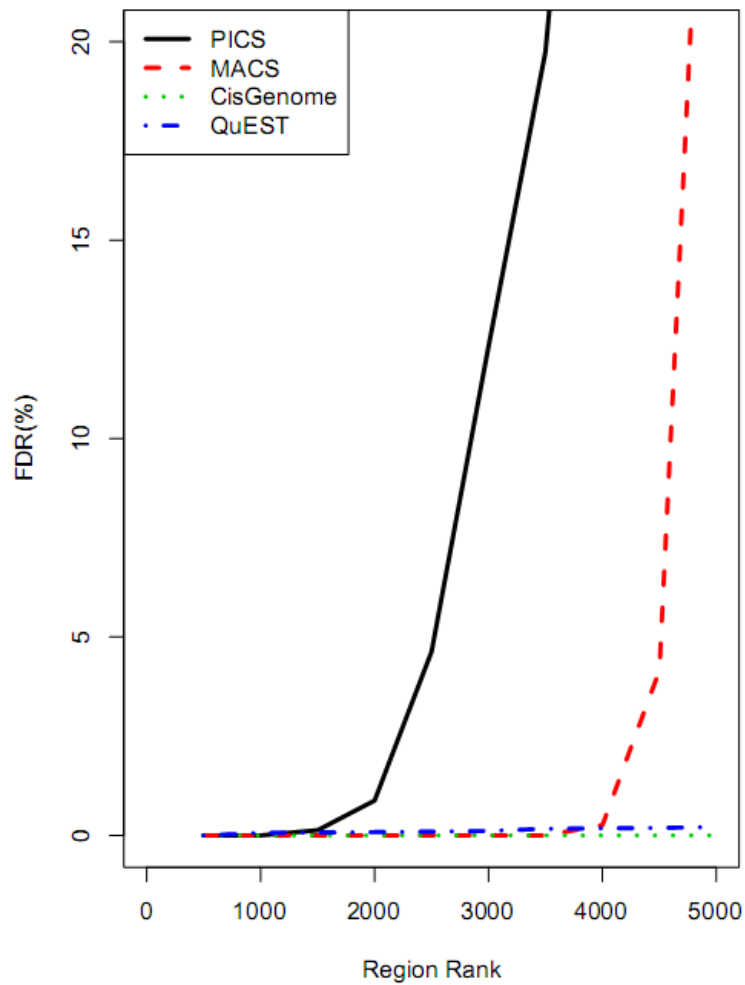
(+ 2 MS Excel files)

## CONTENTS

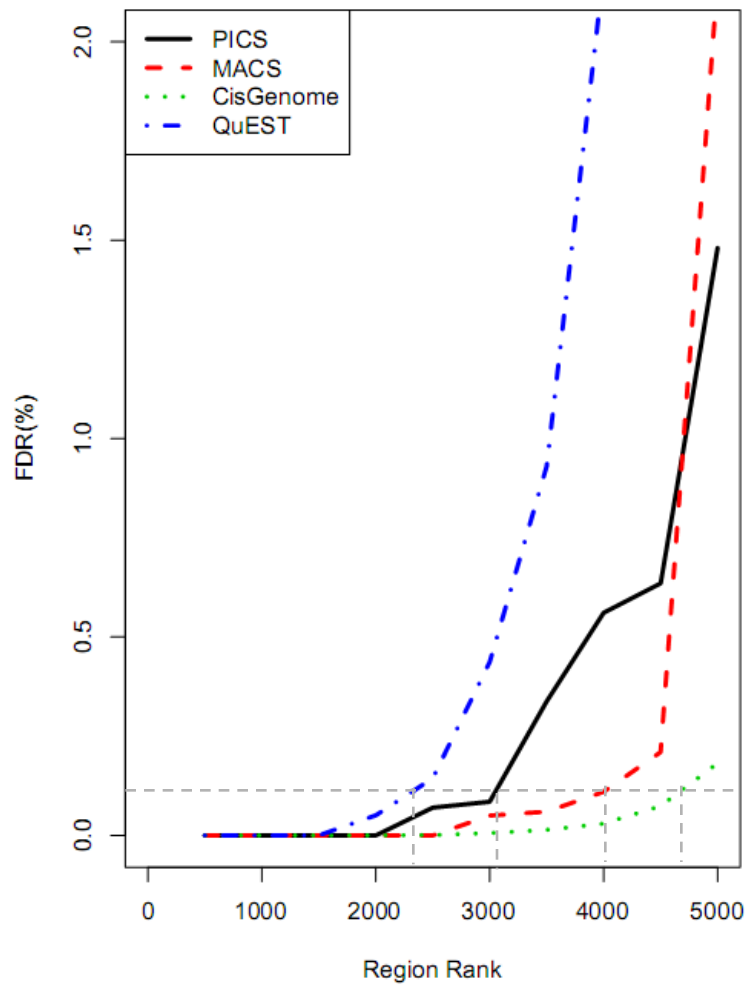
<b>A. AUTISM SPECTRUM DISORDER (ASD) SAMPLE AND CONTROL COLLECTIONS</b>	<b>4</b>
ASD samples	4
Control cohorts	4
<b>B. GENOTYPING AND DATA CLEANING</b>	<b>6</b>
SNP quality control	6
Intensity quality control for CNV detection	6
<b>C. CNV DETECTION AND QUALITY CONTROL EVALUATION</b>	<b>7</b>
Pilot experiment to evaluate the quality of detected stringent CNVs	8
<b>D. CNV VERIFICATION</b>	<b>9</b>
Confirmation using computational methods	9
Experimental CNV validation	9
<b>E. RARE CNV BURDEN ANALYSIS</b>	<b>11</b>
CNV global burden analysis	12
CNV Region (CNVR) burden analysis	12
CNV-based gene association test	13
Burden analysis for genes known to be implicated in ASD and/ or ID	13
Population attributable risk	13
<b>F. GENE-SET ENRICHMENT AND FUNCTIONAL MAP</b>	<b>14</b>
Analytical synopsis	14
Deriving gene sets	14
Gene-set enrichment test	15
Cryptic bias	15
Network visualization of gene-sets enriched for deletions: functional enrichment map	15
Expanded functional map between deletion-enriched gene sets & known ASD/ID genes	16
<b>G. SUPPLEMENTARY FIGURES AND TABLES</b>	<b>18</b>
<b>H. ACKNOWLEDGEMENTS</b>	<b>72</b>
<b>I. SUPPLEMENTARY REFERENCES</b>	<b>73</b>

Algorithm	Control?	Parameters	strands?	Map?	Significance score	FDR measure?	TF/histones?
CSPF	control or IP only	read length no orientation	merge strands no shift	N	simple height criteria	ROC curve (empirically)	both
XSET	IP only	fragment length orientation	merge strands no shift	Y	simple height criteria	FDR estimate using Poisson distribution	both
Mikkelsen et al.	IP only	no orientation	no merge no shift	Y	$p$ -values from permutations	no official FDR	both
MACS	control or IP only	fragment length orientation no duplicated reads	shift reads merge strands	N	Poisson $p$ -values	FDR estimate by peaks in control:IP	both
QuEST	control	orientation	shift reads merge strands	N	kernel density estimation	FDR estimate by permutations of the control	better for TF
FindPeaks	IP only	fragment length orientation	no merge no shift	N	simple height criteria	FDR estimate by permutations of the IP	both
SISSR	control or IP only	fragment length orientation	no merge no shift	N	compares reads on different strands	FDR estimate by peaks in background:IP	better for TF
Kharchenko et al.	control	orientation	no merge no shift	N	Poisson distribution	FDR estimate by permutations of the control	better for TF
PeakSeq	control	fragment length orientation	merge strands	Y	sample normalisation Binomial distribution	FDR estimate, $q$ - values (BH correction)	both
BayesPeak	control or IP only	fragment length orientation	no merge no shift	N	negative binomial distribution, Bayesian posterior probabilities	posterior enrichment probabilities	both

### GABP



### FOXA1



## Two types of complexity in high throughput data processing – (1) difficult math

.....  $l(\Theta|\mathbf{y})$  is the complete-data log-likelihood, given by

$$\begin{aligned} & l(\Theta|\mathbf{y}) \\ &= \sum_{d \in \{f,r\}} \sum_{i=1}^{n_d} \sum_{k=1}^G z_{dik} \left\{ \log \left[ w_k \mathcal{N} \left( d_i | \mu_{dk}, \frac{\sigma_{dk}^2}{u_{dik}} \right) \mathcal{G}a(u_{dik} | 2, 2) \right] \right\} \\ &= \sum_{d \in \{f,r\}} \sum_{i=1}^{n_d} \sum_{k=1}^G z_{dik} \left\{ \log w_k - \log \sigma_{dk} - \log \sqrt{2\pi} - \frac{u_{dik}}{2} \left( \frac{d_i - \mu_{dk}}{\sigma_{dk}} \right)^2 + \log u_{dik} - 2u_{dik} + \log 4 \right\} \end{aligned}$$

and  $l_{\text{prior}}$ , the log prior ‘penalty’ on  $(\delta, \sigma_f^2, \sigma_r^2)$ , is given as

$$l_{\text{prior}} = -\frac{1}{2} \sum_k \{ (\sigma_{fk}^{-2} + \sigma_{rk}^{-2}) [\rho(\delta_k - \xi)^2 + 2\beta] \} + \frac{2\alpha - 1}{2} \sum_k \{ \log(\sigma_{fk}^{-2} + \sigma_{rk}^{-2}) \}. \quad (6)$$

**E-Step:** Given the current estimate  $\Theta^-$  for  $\Theta$ , the conditional expectation of the penalized log complete data likelihood is given as

$$\begin{aligned} Q(\Theta|\Theta^-) &\stackrel{d}{=} \mathbb{E}[l(\Theta|\mathbf{y})|\Theta^-] + l_{\text{prior}} \\ &= \sum_{d \in \{f,r\}} \sum_{i=1}^{n_d} \sum_{k=1}^K \tilde{z}_{dik} \left\{ \log w_k - \log \sigma_{dk} - \frac{\tilde{u}_{dik}}{2} \left( \frac{d_i - \mu_{dk}}{\sigma_{dk}} \right)^2 \right\} + A \quad (7) \end{aligned}$$

where  $A$  is a constant with respect to the parameter vector  $\Theta$ . Given this, the E-step ([Peel and McLachlan, 2000](#)) consists of computing the following quantities

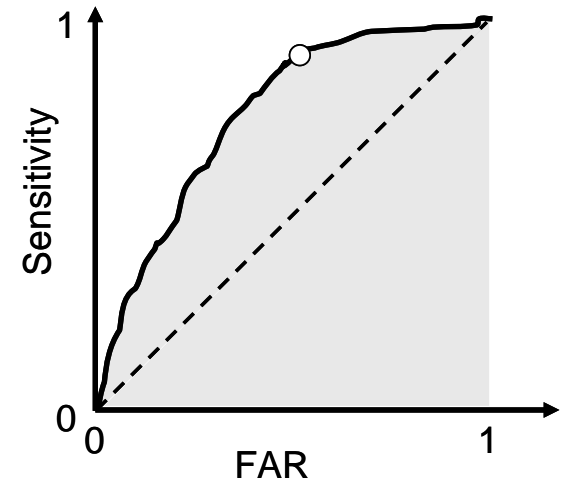
$$\tilde{z}_{dik} \stackrel{d}{=} \mathbb{E}(Z_{dki} | \mathbf{y}_{di}, \Theta^-) = \frac{w_k t_4(d_i | \mu_{dk}, \sigma_{dk})}{\sum_k w_k t_4(d_i | \mu_{dk}, \sigma_{dk})}, \quad (8)$$

$$\tilde{u}_{dik} \stackrel{d}{=} \mathbb{E}(U_{dik} | \mathbf{y}_{di}, z_{dik} = 1, \Theta^-) = \frac{5}{4 + (d_i - \mu_{dk})^2 / \sigma_{dk}^2}. \quad (9)$$

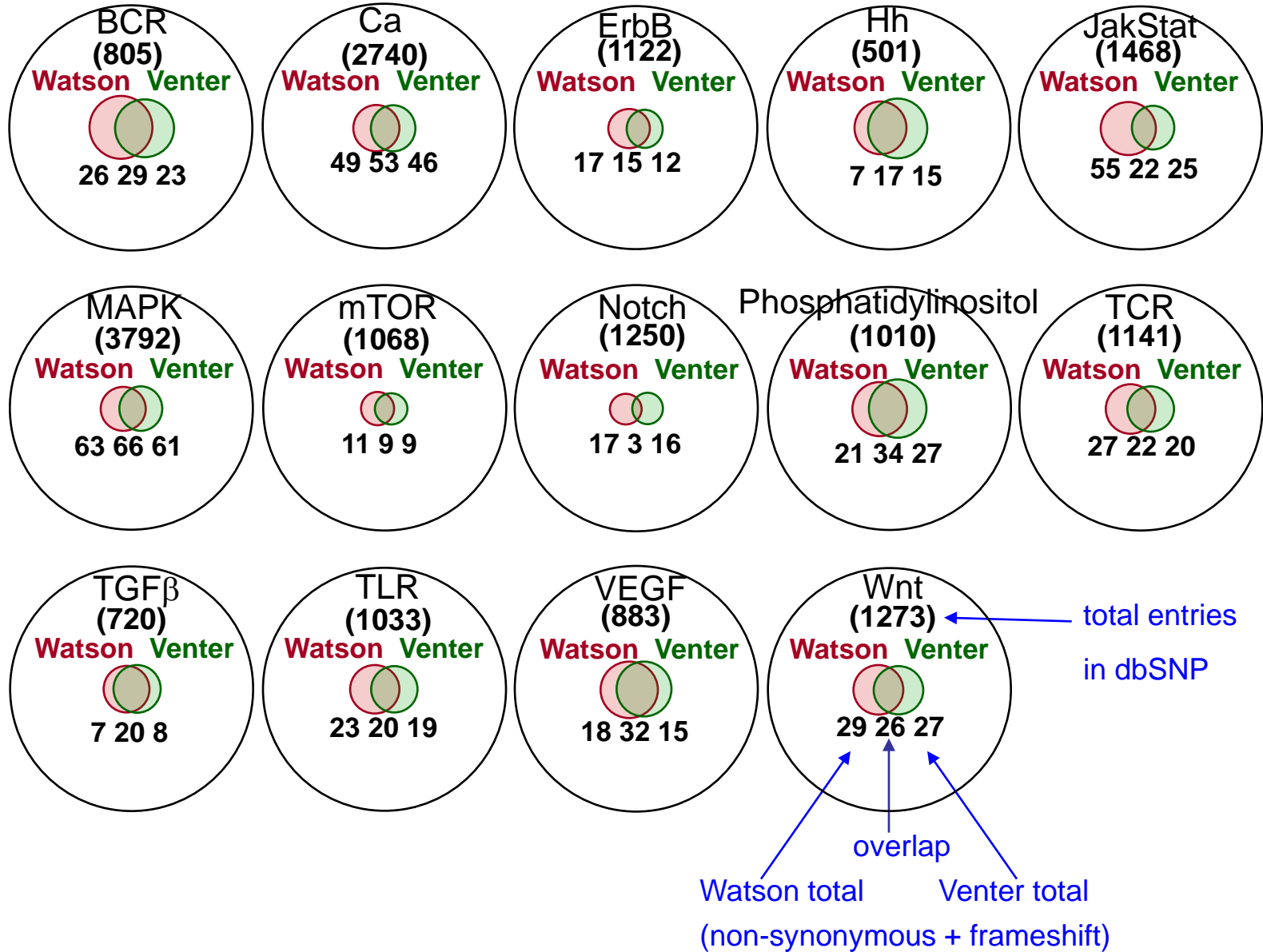
## Two types of complexity in high throughput data processing – (2) *difficult choices*

e.g. for ChIP-seq:

- Average fragment length
- Scan window size
- Mapability measure
- Read count significance threshold
- Background reads distribution
- Sensitivity vs. specificity



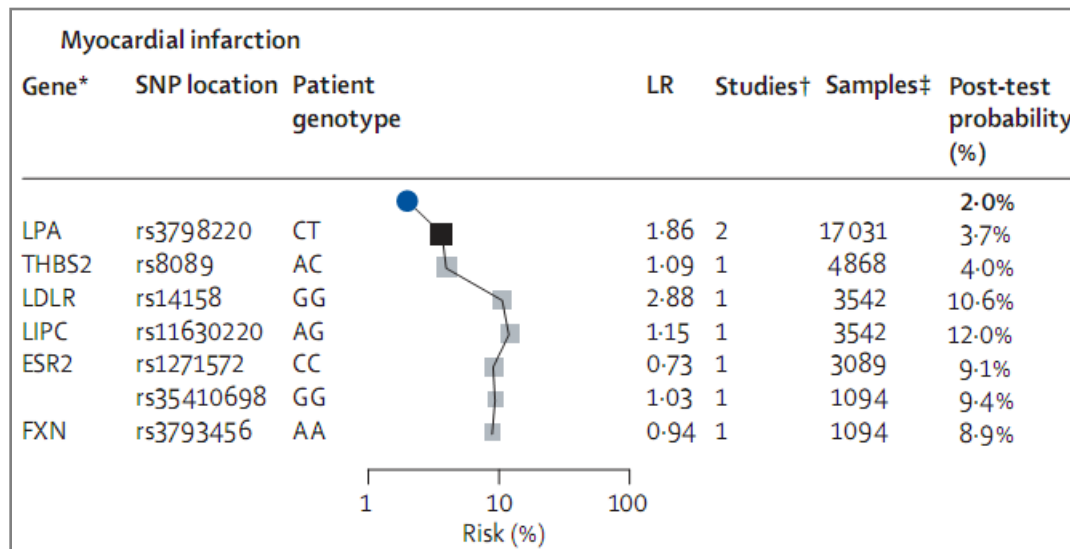
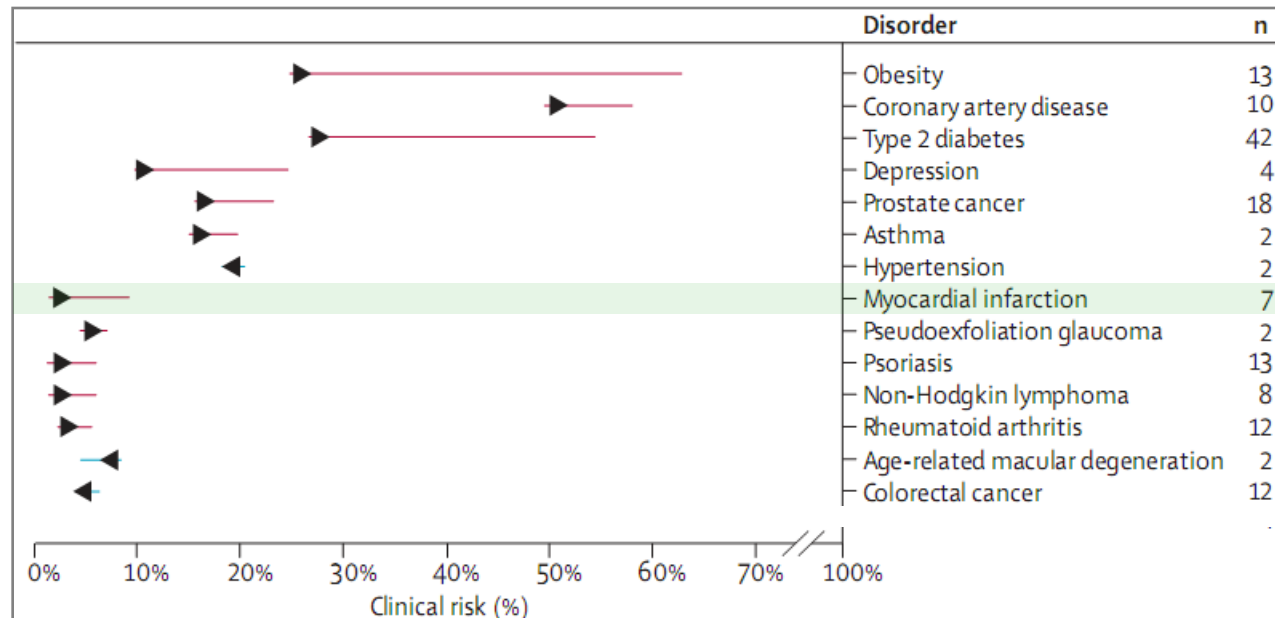
# DNA sequence variations affecting cellular signaling genes in two individuals



# Clinical assessment incorporating a personal genome

(Stephen Quake's genome)

www.thelancet.com Vol 375 May 1, 2010

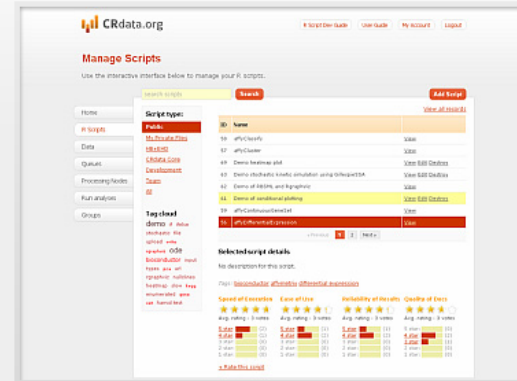


## The Better Way to Crunch data

- ✔ Use privately, or share data & R scripts
- ✔ Access the power of R through point & click menus
- ✔ Run your analysis on the Amazon cloud
  - free for small jobs, or
  - buy virtually unlimited resources directly from Amazon

Proceed as a guest user

Take a Tour

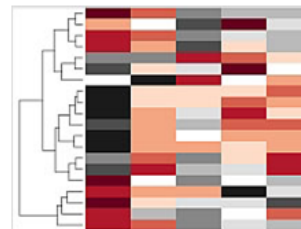
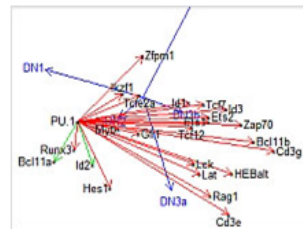


- Home
- R Scripts
- Data
- Queues
- Processing Nodes
- Run analyses
- Groups

## CRdata.org

Every day researchers lose valuable time waiting for computations on their desktops. CRdata.org frees your desktop for other work while our servers analyze your data in the cloud.

You can use CRdata as a private resource, share your data and R scripts with selected others, or with the whole community



To ensure, you have access to the latest resources, all R and Bioconductor libraries are updated nightly on CRdata.

**CRdata is free & open source**

[Programmer Guide](#)

**Get the source code**

Scripts:

create

view

Source code:

```
1 version <- '0.0.6'  
2 #<crdata_text>Script version: </crdata_text>  
3 #<crdata_object>version</crdata_object>  
4 CRDATA_SOURCE_R_FILES <- TRUE  
5 ## setup, if source'd in scripts  
6  
7 if (interactive() ||  
8     (exists("CRDATA_SOURCE_R_FILES") && get("CRDATA_SOURCE_R_FILES")))  
9 {  
10   require(Biobase)  
11   require(Category)  
12   require(KEGG.db)  
13   require(MLInterfaces)  
14   require(RColorBrewer)  
15   require(affy)  
16   require(annotate)  
17   require(bioDist)  
18   require(genefilter)  
19   require(limma)  
20   require(multicore)  
21   require(vsn)  
22 }
```

CRdata.org

http://crdata.org/r\_scripts/new

## Manage Scripts

Use the interactive interface below to manage your R scripts.

- Home
- R Scripts
- Data
- Queues
- Processing Nodes
- Run analyses
- Groups

### New Script

Source code

or [switch back to upload file](#)

Script Name

Tag list   
Needs to be separated by commas

Approx. CPU requirement

Visibility

Do you want to provide a [small tooltip](#) or a [full help page](#)?

[Create Rscript](#) or [Cancel](#)

---

#### Describe script inputs one by one

Parameter name

Display name

Type

Values   
Needs to be separated by commas

[Create Parameter](#)

#### Parameters list

Name	Actions
<i>No parameters.</i>	

Home

R Scripts

Data

Queues

Processing Nodes

Run analyses

Groups

## New Job

### Choose script to run

Laslo et al Mac vs Neutrophil cell fate switch

[Help with chosen script...](#)

Type at least 3 chars and select a name from the list

### Name your job

### Choose Jobs Queue

Public

## Parameters

### Initial (minimum) value of input to PU.1 (Float)

0.1

### Higher value of input to PU.1 (maxP in the script) (Float)

2.0

### Max expression rate of Egr (alpha) (Float)

5.0

### C/EBPalpha expression level (held constant) (Float)

0.1

### feedback nonlinearity coefficient (n, Hill-like) (Float)

4

### Time at which PU.1 switches to high (A.U.) (Integer)

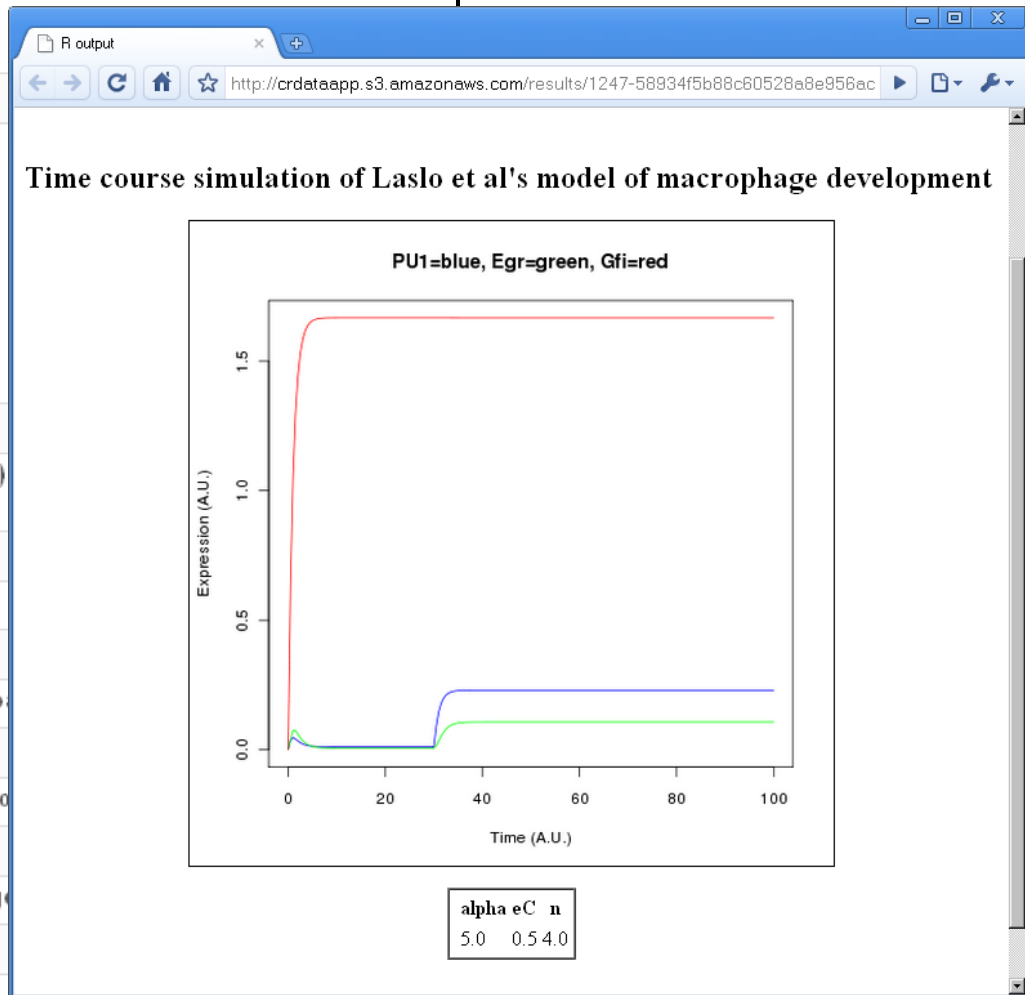
30

### Total time simulated (Integer)

100

Create Job

or [Cancel](#)



# Manage script files and directories

## Manage Scripts

Use the interactive interface below to manage your R scripts.











Script type:

 Public

 My Private Files

 CRdata Core

 Development

 Team

 raiiv\_crdata\_group

 All

Tag cloud

if ifelse qsi

graphviz rna-seq  
 graphviz slow  
 numerated data out  
 test test machine  
 learning grn  
 continuous  
 expression gillespie  
 sequence logos  
 module biomart

[View all records](#)

ID	Name	
94	sequenceDifferentialExperssion (RNA-seq)	<a href="#">View</a>
97	sequenceCountsPerTranscript (RNA-seq)	<a href="#">View</a>
56	affyDifferentialExpression	<a href="#">View</a>
59	affyContinuousGeneSet	<a href="#">View</a>
57	affyCluster	<a href="#">View</a>
58	affyClassify	<a href="#">View</a>
55	affyHeatmap	<a href="#">View</a>
107	PICS plot demo	<a href="#">View</a> <a href="#">Edit</a> <a href="#">Destroy</a>

« Previous **1** 2 3 4 5 Next »

### Selected script details

affyDifferentialExpression

R Documentation

## Two-Group Differential Expression (t-test) on Affymetrix CEL Files

### Description

This script creates a 'top table' of probesets that are differentially expressed between CEL files that have been assigned to one of two groups.

### Usage

```
affyDifferentialExpression(dataset, preprocess = c("rma", "vsnrma"),
                          covariate)
```

### Arguments

**dataset** A 'zip' archive containing (a) Affymetrix CEL files and (b) a 'targets.csv' file. The 'targets.csv' file is a comma-separated table, as might be exported from Excel. The file *must* contain a column 'FileName' that references each CEL file in the archive to be included in the analysis.

Tags: [bioconductor](#) [affymetrix](#) [differential expression](#)

### History:

When	Who	JobName	Time	Status
2010-07-24 18:46:34 UTC		affyDifferentialExpression	51s	Done
2010-07-24 18:44:40 UTC	Hamid SUDO			Update
2010-07-01 21:39:05 UTC	Hamid CRdata team	cont+1 test	55s	Done
2010-06-28 00:58:59 UTC	Martin Morgan			Update
2010-06-26 14:55:08 UTC	Martin Morgan			Update

### Speed of Execution

★★★★★  
Avg. rating - 3 votes

5 star: (2)  
 4 star: (1)  
 3 star: (0)  
 2 star: (0)  
 1 star: (0)

### Ease of Use

★★★★☆  
Avg. rating - 3 votes

5 star: (1)  
 4 star: (2)  
 3 star: (0)  
 2 star: (0)  
 1 star: (0)

### Reliability of Results

★★★★★  
Avg. rating - 3 votes

5 star: (1)  
 4 star: (2)  
 3 star: (0)  
 2 star: (0)  
 1 star: (0)

### Quality of Docs

★★★★★  
Avg. rating - 3 votes

5 star: (0)  
 4 star: (2)  
 3 star: (1)  
 2 star: (0)  
 1 star: (0)

[+ Rate this script](#)

# Manage Jobs Queues

Use the interactive interface below to manage your Jobs Queues.

Home

R Scripts

Data

Queues

Processing Nodes

Run analyses

Groups

## New Queue

### Name

Auto-scale

### Minimum nr. of processing nodes to be associated

### Maximum nr. of processing nodes to be associated

### Set nr. of jobs after auto-scale process for nodes start

### Set maximum idle time for a processing node

### Set maximum waiting time for a job to be processed (in minutes)

### AWS Key

### Visibility

[Create Queue](#)

or [Cancel](#)

# Manage Processing Nodes

Use the interactive interface below to manage your Processing Nodes.

Home

R Scripts

Data

Queues

Processing Nodes

Run analyses

Groups

## New Processing Node

### Choose jobs queue

### AWS Key

### Ec2 instance ami

### Ec2 instance type

or [manually create a node](#)

[Create Processing Node](#)

or [Cancel](#)

## Summary

Empowering individuals, labs, and consortia

An analysis portal for (all) public data

A place to explore and replicate published work

A place to compare quality of available data & methods



To come

Support for multiple languages & Clouds

Strong data encryption for individual genomes

Interactive graphics

Semantic searches

Thanks to:

NHLBI Grant: HL089102

Rajiv Dulepet

Michael Angerman

Constantin Georgescu

Ellen Rothenberg & lab (Caltech)

Barbara Wold & lab (Caltech)

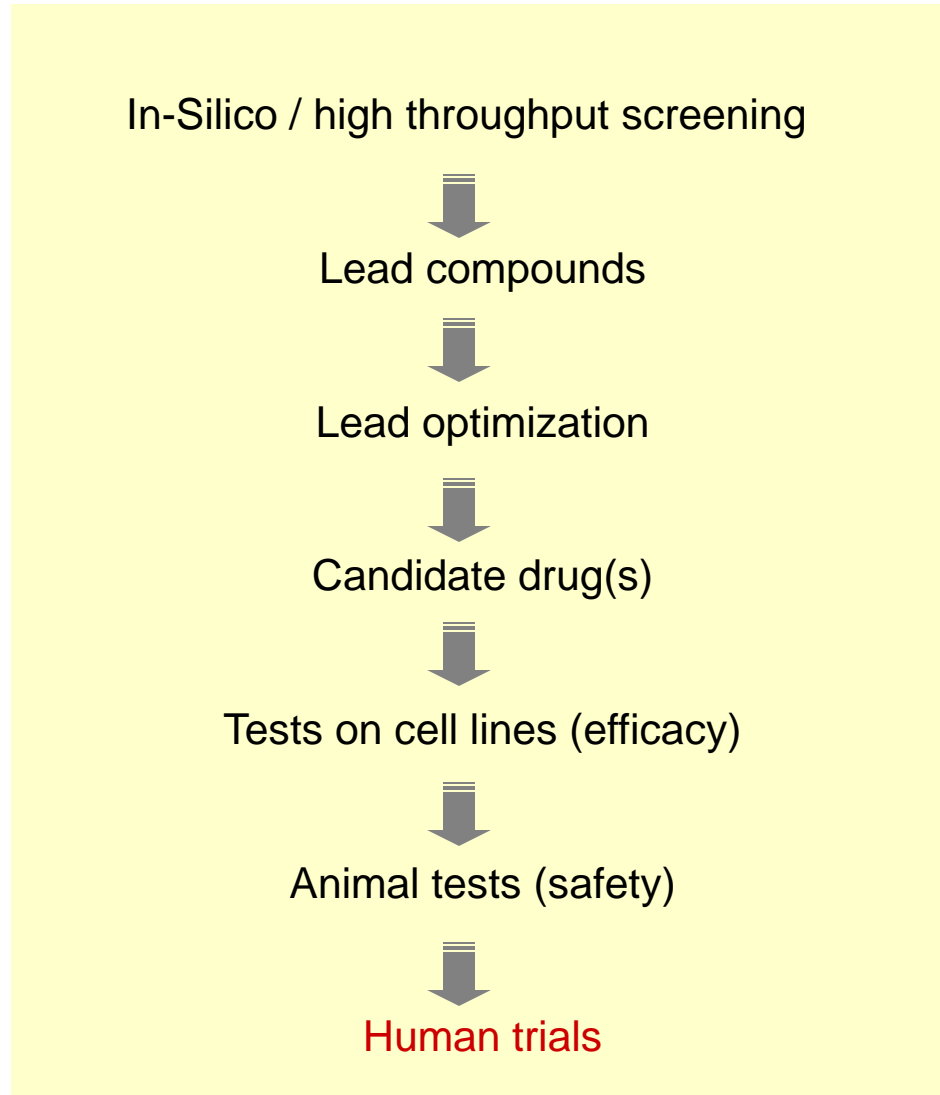
Irwin Benstein & lab (FHCRC)

Guy Naor, Cosmin Andriescu (SparkTech Soft)

Martin Morgan (FHCRC)

# Inter-individual variations are only accounted for at clinical trials & beyond

## Typical drug development process



# ENCODE Transcription Factor Binding Sites by ChIP-seq from HudsonAlpha Institute

Cell Line	GMI2878	K562	hESC	S3	HepG2	A549	BE2_C	GMI2891	GMI2892	Junkat	PANC-1	PFSK-1	SK-N-MC	SK-N-SH_RA	U87	Cell Line
Factor	Tier1	Tier1	Tier2	Tier2	Tier2											Factor

BATF	+ -	✓														
BCL3	+ -	✓														
BCL11A	+ -	✓														
BHLHE40	+ -				✓											
CTCF	+ -					✓										
EBF	+ -	✓														
Egr-1	+ -	✓	✓													
FOSL2	+ -				✓											
FOXP2	+ -															
GABP	+ -	✓	✓		✓											
GR	+ -					✓										
HEY1	+ -		✓		✓											
JmD	+ -				✓											
IRE4	+ -	✓														
NRSF	+ -	✓	✓	✓	✓			✓								
p300	+ -	✓			✓											
PAX5-C20	+ -	✓														
PAX5-N19	+ -	✓														
Pbx3	+ -	✓														
Pol2	+ -	✓	✓	✓	✓	✓		✓								
Pol2-4H8	+ -	✓	✓	✓												
POU2F2	+ -	✓														
PUI1	+ -	✓	✓													
RXRα	+ -				✓											
Sim3Ak-20	+ -	✓	✓		✓											
SIX5	+ -		✓													
SRF	+ -	✓	✓		✓											
SPI1	+ -	✓														
TAF1	+ -	✓	✓	✓	✓	✓		✓	✓							
TCF12	+ -	✓														
USF-1	+ -	✓	✓		✓	✓										
ZBTB33	+ -	✓			✓											
Input	+ -	✓	✓	✓	✓	✓	✓	✓	✓							
Factor	+ -	+ -	+ -	+ -	+ -	+ -	+ -	+ -	+ -							
Cell Line	GMI2878	K562	HL-hESC	HeLa-S3	HepG2	A549	BE2_C	GMI2891	GMI2892							
	Tier1	Tier1	Tier2	Tier2	Tier2											
+ -	All															

## ENCODE Caltech RNA-seq (Barbara Wold's lab)

Maximum display mode:   [Reset to defaults](#)

Select views ([help](#)):

[Plus Raw Signal](#)  [Minus Raw Signal](#)  [Raw Signal](#)  [Splice Sites](#)

Select subtracks by read type and cell line:

Replicate:  1  2  3  4

+ - All	Single Strand-Specific Paired			
	Read Type	32nt	75nt	75nt
Cell Line	+ -	+ -	+ -	
GMI2878	+ -	✓	✓	✓
HLhESC	+ -			
HeLaS3	+ -			
HepG2	+ -			
HUVEC	+ -			
K562	+ -			
NHEK	+ -			

## ENCODE Histone Modifications by Broad Institute ChIP-seq

Maximum display mode:   [Reset to defaults](#)

Select views ([help](#)):

Peaks  [Signal](#)

Peaks Configuration

Minimum P-Value (-log 10):  (0 to 100)

Select subtracks by cell line and antibody:

+ - All	Cell Line	GMI2878	HL-hESC	HepG2	HMEC	HSMIM	HUVEC	K562	NHEK	NHLF	Cell Line	All + -
Antibody	+ -	+ -	+ -	+ -	+ -	+ -	+ -	+ -	+ -	+ -	Antibody	+ -
CTCF	+ -	✓						✓			+ -	CTCF
H3K4me1	+ -	✓						✓			+ -	H3K4me1
H3K4me2	+ -										+ -	H3K4me2
H3K4me3	+ -	✓						✓			+ -	H3K4me3
H3K9ac	+ -										+ -	H3K9ac
H3K9me1	+ -										+ -	H3K9me1
H3K27ac	+ -	✓						✓			+ -	H3K27ac
H3K27me3	+ -	✓						✓			+ -	H3K27me3
H3K36me3	+ -	✓						✓			+ -	H3K36me3
H4K20me1	+ -										+ -	H4K20me1
Pol2(b)	+ -										+ -	Pol2(b)
Input Control	+ -										+ -	Input Control

List subtracks:  only selected/visible  all (24 of 177 selected)

Cell Line<sup>11</sup> Antibody<sup>12</sup> Views<sup>13</sup>

✓	GMI2878	CTCF	Peaks	ENCODE Histone Mods, Broad ChIP-seq Peaks (CTCF, GMI2878) ...	<a href="#">schema</a>	2009-10-
✓	GMI2878	CTCF	Signal	ENCODE Histone Mods, Broad ChIP-seq Signal (CTCF, GMI2878) ...	<a href="#">schema</a>	2009-10-



## Manage Jobs


Use the interactive interface below to manage your analysis runs.


[Add Job - Wizard](#)

[Add Job - Manual](#)
[View All](#)
[Delete](#)
[Home](#)
[R Scripts](#)
[Data](#)
[Queues](#)
[Processing Nodes](#)

[Run analyses](#)
[Groups](#)
 New  Submitted  Running  Done  Failed  Cancelled  Pending

<input type="checkbox"/>	ID	Description	Created at	Status	Running Time	Actions
<input type="checkbox"/>	1050	ChIPseq fragment length estimation	2010-07-06 20:06:28 UTC	✓	31s	<a href="#">Results</a> <a href="#">Feedback</a> <a href="#">Logs</a> <a href="#">Clone</a> <a href="#">Delete</a>
<input type="checkbox"/>	1049	ChIPseq fragment length estimation	2010-07-06 19:35:47 UTC	✓	31s	<a href="#">Results</a> <a href="#">Feedback</a> <a href="#">Logs</a> <a href="#">Clone</a> <a href="#">Delete</a>
<input type="checkbox"/>	1048	chipseq	2010-07-05 23:47:52 UTC	✓	39s	<a href="#">Results</a> <a href="#">Feedback</a> <a href="#">Logs</a> <a href="#">Clone</a> <a href="#">Delete</a>
<input type="checkbox"/>	1047	chipseq	2010-07-05 23:14:56 UTC	✓	38s	<a href="#">Results</a> <a href="#">Feedback</a> <a href="#">Logs</a> <a href="#">Clone</a> <a href="#">Delete</a>
<input type="checkbox"/>	1045	qqplot test	2010-07-05 18:59:07 UTC	✓	4s	<a href="#">Results</a> <a href="#">Feedback</a> <a href="#">Logs</a> <a href="#">Clone</a> <a href="#">Delete</a>
<input type="checkbox"/>	1044	qqplot test	2010-07-05 18:54:27 UTC	✓	4s	<a href="#">Results</a> <a href="#">Feedback</a> <a href="#">Logs</a> <a href="#">Clone</a> <a href="#">Delete</a>
<input type="checkbox"/>	1043	Hmisc xYplot	2010-07-04 17:20:07 UTC	✓	3s	<a href="#">Results</a> <a href="#">Feedback</a> <a href="#">Logs</a> <a href="#">Clone</a> <a href="#">Delete</a>
<input type="checkbox"/>	1042	chipseq package plot tests	2010-07-04 17:12:23 UTC	✓	1m 10s	<a href="#">Results</a> <a href="#">Feedback</a> <a href="#">Logs</a> <a href="#">Clone</a> <a href="#">Delete</a>
<input type="checkbox"/>	1041	Seq Count	2010-07-04 01:10:16 UTC	✓	59s	<a href="#">Results</a> <a href="#">Feedback</a> <a href="#">Logs</a> <a href="#">Clone</a> <a href="#">Delete</a>
<input type="checkbox"/>	1039	ChIPseqR plot test	2010-07-03 21:24:41 UTC	✓	2h 30m 47s	<a href="#">Results</a> <a href="#">Feedback</a> <a href="#">Logs</a> <a href="#">Clone</a> <a href="#">Delete</a>
<input type="checkbox"/>	1040	PICS plot test	2010-07-03 21:27:57 UTC	✓	57s	<a href="#">Results</a> <a href="#">Feedback</a> <a href="#">Logs</a> <a href="#">Clone</a> <a href="#">Delete</a>
<input type="checkbox"/>	1038	IRanges plot test	2010-07-03 21:20:42 UTC	✓	6s	<a href="#">Results</a> <a href="#">Feedback</a> <a href="#">Logs</a> <a href="#">Clone</a> <a href="#">Delete</a>
<input type="checkbox"/>	1037	Genominator plot test	2010-07-03 21:16:50 UTC	✓	26s	<a href="#">Results</a> <a href="#">Feedback</a> <a href="#">Logs</a> <a href="#">Clone</a> <a href="#">Delete</a>
<input type="checkbox"/>	1036	GenomeGraphs test plots	2010-07-03 21:14:15 UTC	✓	27s	<a href="#">Results</a> <a href="#">Feedback</a> <a href="#">Logs</a> <a href="#">Clone</a> <a href="#">Delete</a>



Home  
R Scripts  
Data  
Queues  
Processing Nodes  
Run analyses  
Groups

## New Job - Step 1 from 2 (script selection)

seq

**Script type:**

- [Public](#)
- [My Private Files](#)
- [CRdata Core](#)
- [Development](#)
- [Team](#)
- [rajiv\\_crdata\\_group](#)

[View all records](#)

ID	Name	
94	sequenceDifferentialExperssion (RNA-seq)	<a href="#">View</a>
97	sequenceCountsPerTranscript (RNA-seq)	<a href="#">View</a>
107	PICS plot demo	<a href="#">View</a> <a href="#">Edit</a> <a href="#">Destroy</a>
109	Estimate ChIPseq fragment length	<a href="#">View</a> <a href="#">Edit</a> <a href="#">Destroy</a>

Home  
R Scripts  
Data  
Queues  
Processing Nodes  
Run analyses  
Groups

## New Job - Step 2 from 2 (job information)

**Name your job**

Estimate ChIP-seq DNA fragment length

**Choose Jobs Queue**

Public

**Parameters**

choose the IP type (List)

ctcf

select chromosome (List)

chr10  
chr10  
chr11  
chr12

or [Cancel](#)

for ChIPseq [View](#) [Edit](#) [Destroy](#)

arest genes [View](#) [Edit](#) [Destroy](#)

k & significance threshold [View](#) [Edit](#) [Destroy](#)

n of a PWM [View](#)

vious **1** 2 [Next >](#)

seq data, we can estimate DNA fragment lengths using  
ng on mouse chromosomes 10, 11 and 12.

age manual at:  
ges/release/bioc/html/chipseq.html for more details.

**History:**

When	Who	JobName	Time	Status
2010-07-23 21:02:22 UTC	Hamid SUDO	ChIPseq Frag length	39s	Done

Job: test30

R script: test from rajiv

### Job Timing

Submitted: 2010-06-15 07:25:38 UTC  
 Started: 2010-06-15 07:25:44 UTC  
 Completed: 2010-06-15 07:25:59 UTC

### R Script

[test from rajiv](#)

Parameters:  
coTFs: [TFs](#)

[Go back to jobs list](#)

[View job](#)

STATUS: done

## Manage Data

Use the interactive interface below to manage your data.

search datasets

[View all records](#)

Home

R Scripts

Data

Queues

Processing Nodes

Run analyses

Groups

**Dataset type:**

Public

**My Private Files**

CRdata Core

Development

Team

All

**Tag cloud**

demo affymetrix  
chip-seq qsi cel  
rna-seq pwm  
bioconductor

ID	Name	
175	job-0000001136-coverageOnChr10.bedGraph	<a href="#">View</a> <a href="#">Edit</a> <a href="#">Destroy</a>
177	job-0000001138-targetGenes.csv	<a href="#">View</a> <a href="#">Edit</a> <a href="#">Destroy</a>
176	job-0000001137-PeaksChr10.bedGraph	<a href="#">View</a> <a href="#">Edit</a> <a href="#">Destroy</a>
114	BAM Sequence Alignment File	<a href="#">View</a>
108	Sequence Differential Expression Input	<a href="#">View</a>
103	Sample CEL file subset	<a href="#">View</a>
72	rajiv_public_test_data	<a href="#">View</a>
66	TFs	<a href="#">View</a>

« Previous 1 2 Next »

### Selected dataset details

This is a sample output file from the script: "Generate significant peaks track for ChIPseq". The locations of statistically significant ChIPseq peaks are marked with a bar of unit height.

Tags: [hb](#) [qsi](#) [chip-seq](#)

### History:

When	Who	JobName	Time	Status
2010-07-18 18:09:29 UTC	Hamid Bolouri			Update
2010-07-18 18:08:43 UTC	Hamid Bolouri			Update
2010-07-08 16:38:37 UTC				Create

**Speed of Execution** Avg. rating - 0 votes

**Ease of Use** Avg. rating - 0 votes

**Quality of Docs** Avg. rating - 0 votes

**Quality of Data** Avg. rating - 0 votes

5 star: (0) 4 star: (0) 3 star: (0) 2 star: (0) 1 star: (0)

[+ Rate this dataset](#)

## Dataset: TFs

### Description:

Dataset file: <http://crdataapp.s3.amazonaws.com/data/1bc86620-57d2-012d-7a7c-1231390c5015/TFs.txt>

Jobs: [509](#) [516](#) [525](#) [565](#) [590](#) [639](#) [640](#) [651](#) [993](#) [1057](#) [1112](#) [1113](#) [1114](#) [1118](#) [1119](#) [1122](#) [1125](#)

### Tags:

### Ratings by you:

Speed of Execution:

Ease of Use:

Quality of Docs:

Quality of Data:

### Title

### Comment

## Create & delete processing queues

### Manage Jobs Queues

Use the interactive interface below to manage your Jobs Queues.

search jobs queues

<input type="checkbox"/>	ID	Name	Security	Processing Nodes	Actions
<input type="checkbox"/>	5	<a href="#">A test shared queue for Hamid's 2 groups</a>	<a href="#">CRdata Core Development Team</a>		<a href="#">Edit</a> <a href="#">Destroy</a>
<input type="checkbox"/>	14	<a href="#">H@Caltech Q - shared with H@ISB</a>	<a href="#">CRdata Core Development Team</a>		<a href="#">Edit</a> <a href="#">Destroy</a>
<input type="checkbox"/>	19	<a href="#">Public</a>	public	i-6c913e06	<a href="#">Edit</a> <a href="#">Destroy</a>

## Create & delete processing nodes for queues

### Manage Processing Nodes

Use the interactive interface below to manage your Processing Nodes.

search processing nodes

<input type="checkbox"/>	ID	Node Identifier	IP Address	Status	Queue	Running jobs	Actions
<input type="checkbox"/>	<a href="#">100</a>	<a href="#">i-6c913e06</a>	174.129.86.120	Activated	<a href="#">Public</a>		

## Manage Groups

Use the interactive interface below to manage your Groups.






ID	Name	Description	Users	Actions
57	<a href="#">EddieTestGroup</a>	<a href="#">eddie and Ron</a>	1	<a href="#">Members</a> <a href="#">Join</a>
25	<a href="#">Bernstein lab, FHCRC</a>	<a href="#">Data and R code used within Irwin Bernstein's lab</a>	4	<a href="#">Members</a> <a href="#">Leave</a>
24	<a href="#">HB+EHD</a>	<a href="#">Models and data shared by HB and EHD</a>	3	<a href="#">Members</a> <a href="#">Leave</a>
6	<a href="#">CRdata Core Development Team</a>	<a href="#">CRdata core developers' group</a>	4	<a href="#">Members</a>
13	<a href="#">rajiv crdata group</a>		3	<a href="#">Members</a> <a href="#">Leave</a>
8	<a href="#">angermanmichael's group</a>		2	<a href="#">Members</a> <a href="#">Join</a>

- Home
- R Scripts
- Data
- Queues
- Processing Nodes
- Run analyses
- Groups

**Name:** Example Group

**Description:** About Example Group...

**Read Me:** More about Example Group

---

**Members:**

1. Hamid Bolouri [Cancel membership](#)

## Manage Account

Use the interactive interface below to manage your account settings.

- Home
- R Scripts
- Data
- Queues
- Processing Nodes
- Run analyses
- Groups

*Name:* **Hamid Bolouri**

*Email:* hbolouri@gmail.com

Settings for:

[Account](#) [Notifications](#) [AWS Keys](#) [Default availability zone](#) [Default queue](#)

CRdata.org [R Script Dev Guide](#) [User Guide](#) [My Account](#) [Logout](#)

### Manage AWS Keys

Use the interactive interface below to manage your AWS Keys.

Home

R Scripts

Data

Queues

Processing Nodes

Run analyses

Groups

#### New AWS Key

Name

Access key

Secret access key

[Create AWS Key](#) or [Cancel](#)

[Users' Blog](#) [About Us](#) [Privacy Policy](#) [MIT License](#)