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)
See also:

http://models.cellml.org

http://www.ebi.ac.uk/biomodels-main/
High throughput molecular and cell biology

### Technology characteristics
- Small sample sizes
- Commoditized equipment
- Low-cost per experiment
- Global (genome-wide) data
- Rapidly evolving

### Currently dominated by
- Re-sequencing
- ChIP-seq
- RNA-seq

### But also
- Proteomics, Metabolomics
- Screens (RNAi, drugs, mutants)
- Biomarker assays
- Single-cell assays
  - microfluidics, FACS, microscopy

### Computational implications
- Large, noisy datasets
  - Complex analysis needed
  - Results are statistical & parameter/method dependent

- Accessible to individual labs
  - Global data, focused hypotheses
  - Bursts of heavy computing
  - Opportunity: iterative data exploration

- Large-scale cataloguing projects
  - (TGCA, ENCODE, Epigenome, 1000 genomes, Bacteriome...)
  - Need to share data + analysis
  - Opportunity: 3\textsuperscript{rd}-party data mining
Genome and transcriptome mapping

Mapping large numbers of sequence reads to a reference genome or transcriptome is the first step in an analysis. Upon data upload, DNAnexus automatically maps all your sequence data to your genome of choice, allowing you to understand data quality from a mapping perspective, visualize your mappings, download the mappings in several standard as well as custom-defined formats, or perform downstream analyses within DNAnexus. For transcriptome-based sequencing, DNAnexus enables you to map to gene annotation databases of your choice, or to "splice-omes" that consider all possible junctions between annotated exons.
The Elastic-R AJAX Workbench

http://biocep-distrib.r-forge.r-project.org/doc.html
Create customized workspaces with tabbed, dockable and floating tool windows.
ChIP-seq Overview

1. Cross-link proteins to DNA

2. Fragment chromatin to 100-150bp

3. Immunoprecipitate antibody-bound DNA fragments

4. Reverse cross-links and sequence fragment ends

5. Map sequence reads to genome

6. Identify genomic regions with enriched number of mapped reads

Barbara Wold’s lab, Science, 2007, 16(5830):1497-502
Kharchenko, Tolstorukov and Park

Ly9 expression is repressed in the bone marrow

Data Suzanne Furuyama & Irwin Bernstein, FHCRC

http://biogps.gnf.org/
A typical pipeline:

(Hes1 ChIPseq in HSCs Bernstein lab, FHCRC)

> 500 lines of R

Uses 12 R “packages”

Includes > 10 major user-defined parameters

(e.g. # read mismatches, Baye’s PP, ncRNAs?)

> 200 interim files
Two types of complexity in high throughput data processing – (1) difficult math

\[ l(\Theta|y) \] is the complete-data log-likelihood, given by

\[
\begin{align*}
    l(\Theta|y) &= \sum_{d \in \{f,r\}} \sum_{i=1}^{n_d} \sum_{k=1}^{G} z_{dik} \left\{ \log \left[ w_k N \left( d_i | \mu_d, \sigma_d^2 / u_{dik} \right) \right] \right\} \\
    &= \sum_{d \in \{f,r\}} \sum_{i=1}^{n_d} \sum_{k=1}^{G} z_{dik} \left\{ \log w_k - \log \sigma_d - \log \sqrt{2\pi} - \frac{u_{dik}}{2} \frac{(d_i - \mu_d)^2}{\sigma_d^2} + \log u_{dik} - 2u_{dik} + \log 4 \right\}
\end{align*}
\]

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

\[
l_{\text{prior}} = \frac{1}{2} \sum_k \left\{ (\sigma_{f,k}^{-2} + \sigma_{r,k}^{-2}) [\rho(\delta_k - \xi)^2 + 2\beta] \right\} + \frac{2\alpha - 1}{2} \sum_k \left\{ (\sigma_{f,k}^{-2} + \sigma_{r,k}^{-2}) \right\}.
\]

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

\[
Q(\Theta|\Theta^-) = \mathbb{E}[l(\Theta|y)|\Theta^-] + l_{\text{prior}}
\]

\[
= \sum_{d \in \{f,r\}} \sum_{i=1}^{n_d} \sum_{k=1}^{K} \tilde{z}_{dik} \left\{ \log w_k - \log \sigma_d - \frac{\tilde{u}_{dik}}{2} \frac{(d_i - \mu_d)^2}{\sigma_d^2} \right\} + A
\]

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

\[
\tilde{z}_{dik} \overset{d}{=} \mathbb{E}(Z_{dki}|y_{di}, \Theta^-) = \frac{w_kt_4(d_i | \mu_d, \sigma_d)}{\sum_k w_kt_4(d_i | \mu_d, \sigma_d)}
\]

\[
\tilde{u}_{dik} \overset{d}{=} \mathbb{E}(U_{dik}|y_{di}, z_{dik} = 1, \Theta^-) = \frac{5}{4 + (d_i - \mu_d)^2 / \sigma_d^2}.
\]
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 fragments (excluding 120bp adapters) are asymmetrically distributed ~ 50bp–280bp
(data: Suzanne Furuyama, Bernstein lab, FHCRC)
Estimating the average DNA fragment length from the strand-specific tag shift

Sarkar, Gentleman, Lawrence, Zhang, Yao

Kharchenko, Tolstorukov and Park

Assuming average fragment length=75bp

Assuming average fragment length=136bp
<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Control?</th>
<th>Parameters</th>
<th>strands?</th>
<th>Map?</th>
<th>Significance score</th>
<th>FDR measure?</th>
<th>TF/histones?</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSPF</td>
<td>control or IP only</td>
<td>read length no orientation</td>
<td>merge strands no shift</td>
<td>N</td>
<td>simple height criteria</td>
<td>ROC curve (empirically)</td>
<td>both</td>
</tr>
<tr>
<td>XSET</td>
<td>IP only</td>
<td>fragment length orientation</td>
<td>merge strands no shift</td>
<td>Y</td>
<td>simple height criteria</td>
<td>FDR estimate using Poisson distribution</td>
<td>both</td>
</tr>
<tr>
<td>Mikkelsen et al.</td>
<td>IP only</td>
<td>no orientation</td>
<td>no merge no shift</td>
<td>Y</td>
<td>p-values from permutations</td>
<td>no official FDR</td>
<td>both</td>
</tr>
<tr>
<td>MACS</td>
<td>control or IP only</td>
<td>fragment length orientation no duplicated reads</td>
<td>shift reads merge strands</td>
<td>N</td>
<td>Poisson p-values</td>
<td>FDR estimate by peaks in control:IP</td>
<td>both</td>
</tr>
<tr>
<td>QuEST</td>
<td>control</td>
<td>orientation</td>
<td>shift reads merge strands</td>
<td>N</td>
<td>kernel density estimation</td>
<td>FDR estimate by permutations of the control</td>
<td>better for TF</td>
</tr>
<tr>
<td>FindPeaks</td>
<td>IP only</td>
<td>fragment length orientation</td>
<td>no merge no shift</td>
<td>N</td>
<td>simple height criteria</td>
<td>FDR estimate by permutations of the IP</td>
<td>both</td>
</tr>
<tr>
<td>SISSR</td>
<td>control or IP only</td>
<td>fragment length orientation</td>
<td>no merge no shift</td>
<td>N</td>
<td>compares reads on different strands</td>
<td>FDR estimate by peaks in background:IP</td>
<td>better for TF</td>
</tr>
<tr>
<td>Kharchenko et al.</td>
<td>control</td>
<td>orientation</td>
<td>no merge no shift</td>
<td>N</td>
<td>Poisson distribution</td>
<td>FDR estimate by permutations of the control</td>
<td>better for TF</td>
</tr>
<tr>
<td>PeakSeq</td>
<td>control</td>
<td>fragment length orientation</td>
<td>merge strands</td>
<td>Y</td>
<td>sample normalisation Binomial distribution</td>
<td>FDR estimate, q-values (BH correction)</td>
<td>both</td>
</tr>
<tr>
<td>BayesPeak</td>
<td>control or IP only</td>
<td>fragment length orientation</td>
<td>no merge no shift</td>
<td>N</td>
<td>negative binomial distribution, Bayesian posterior probabilities</td>
<td>posterior enrichment probabilities</td>
<td>both</td>
</tr>
</tbody>
</table>

Spyrou et al, BMC Bioinformatics 2009, 10:299
Zhang et al, Biometrics 2010, Jun 1st [Epub ahead of print]
Predicting gene expression from epigenomics
(collaboration with Jingli Zhang and Ellen Rothenberg, Caltech)

Regression on all data

Using all Principle Components
Correlation coefficient=0.72

Using only top 2 PCs
Correlation coefficient=0.70

Leave One Out cross validation

Predicted DN3/DN1 expression
Actual DN3/DN1 expression
Transfac matrices associated with stage-specific H3K4me2 regions

Where TFs have multiple matrices, they fall close to each other in the plot

Except for PU.1 (blue background)

Dahl et al, Spi-B can functionally replace PU.1 in myeloid but not lymphoid development

1000-fold enrichment of predictions:

2.73Mbp out of 2.5Gbp selected

612,626 hits for 647 Transfac matrices

Putative enhancers for 8,556 genes
### ENCODE Transcription Factor Binding Sites by ChIP-seq from HudsonAlpha Institute

<table>
<thead>
<tr>
<th>Cell Line</th>
<th>GM12878 Tier1</th>
<th>K562 Tier1</th>
<th>hESC Tier2</th>
<th>S3 Tier2</th>
<th>HepG2 Tier2</th>
<th>A549</th>
<th>BE2-C</th>
<th>GM12891</th>
<th>GM12892</th>
<th>J521</th>
<th>Panc.1</th>
<th>PFSK-1</th>
<th>SK-N-MC</th>
<th>SK-N-SH_RA</th>
<th>US7</th>
<th>Cell Line</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BATF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCL3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCL11A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BHLH40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTCF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EBF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Egr-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOXL2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOXO4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GABP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEY1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JunD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRE1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NRSF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAX5-C20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAX5-N19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pbx3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pol2-4H8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pou2F2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RARA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sm3al-20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIX5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAF1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCF12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USF1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZBTB33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Input</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

- Maximum display mode: [on, off, submit, export to default]

Select subtracks by road type and cell line:

- Replicate: 1, 2, 3, 4

### ENCODE Histone Modifications by Broad Institute ChIP-seq

- Maximum display mode: [on, off, submit, export to default]

Select subtracks by cell line and antibody:

- Antibody: CTCF, H3K4me1, H3K4me2, H3K4me3, H3K27ac, H3K27me3, H3K36me3, H3K36me2, H3K36ac, H3K20me1, Pol2-4H8, Input Control

List subtracks: only selected/visible, all (24 out of 177 selected)
NIH Roadmap Epigenomics consortium

292 public datasets so far
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

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

Users' Blog  About Us  Privacy Policy  MIT License
Bringing computation to the lab bench

- Point & Click user interface
- Hide the math, highlight and enable choices
- Large-scale computing resources available on demand
- No software installation/maintenance
- No IT infrastructure
  - No need for sys-admin, obsolescence, space, cooling, etc.

Empowering the stakeholders

- Enable experimental biologists to *interactively* explore their data
  - Alternative analysis methods
  - Alternative parameter choices
  - Alternative views of results/data
- Relieve computational biologists from non-specialist tasks
  - Allow focus on strengths, e.g. method development
Scalability

• Number of users (jobs being run)
• Number/size of scripts and datasets
• Size of individual jobs (CPU + memory required)
• Cost
  Processor Queues auto-scale with demand
  Users can have private nodes and shared nodes

Evolvability

• Diversity of users: freely available to all, open source
• Diversity of content: data & scripts provided by the community
• Community curated
• Inclusive architecture supports
  multiple OS (Windows, Unix/Linux, Mac)
  multiple languages (Perl, Python, Java coming)
## Manage Jobs

Use the interactive interface below to manage your analysis runs.

### Jobs Table

<table>
<thead>
<tr>
<th>ID</th>
<th>Description</th>
<th>Created at</th>
<th>Status</th>
<th>Running Time</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1050</td>
<td>ChipSeq fragment length estimation</td>
<td>2010-07-06 20:06:28 UTC</td>
<td>✔️</td>
<td>31s</td>
<td>Results, Feedback, Log, Clone, Delete</td>
</tr>
<tr>
<td>1048</td>
<td>ChipSeq fragment length estimation</td>
<td>2010-07-06 19:35:47 UTC</td>
<td>✔️</td>
<td>31s</td>
<td>Results, Feedback, Log, Clone, Delete</td>
</tr>
<tr>
<td>1047</td>
<td>chisqex</td>
<td>2010-07-05 23:47:52 UTC</td>
<td>✔️</td>
<td>39s</td>
<td>Results, Feedback, Log, Clone, Delete</td>
</tr>
<tr>
<td>1045</td>
<td>chisqex</td>
<td>2010-07-05 23:14:56 UTC</td>
<td>✔️</td>
<td>38s</td>
<td>Results, Feedback, Log, Clone, Delete</td>
</tr>
<tr>
<td>1044</td>
<td>qplot test</td>
<td>2010-07-05 16:59:07 UTC</td>
<td>✔️</td>
<td>4s</td>
<td>Results, Feedback, Log, Clone, Delete</td>
</tr>
<tr>
<td>1044</td>
<td>qplot test</td>
<td>2010-07-06 18:54:27 UTC</td>
<td>✔️</td>
<td>4s</td>
<td>Results, Feedback, Log, Clone, Delete</td>
</tr>
<tr>
<td>1043</td>
<td>Hmisc xaplot</td>
<td>2010-07-04 17:20:07 UTC</td>
<td>✔️</td>
<td>3s</td>
<td>Results, Feedback, Log, Clone, Delete</td>
</tr>
<tr>
<td>1042</td>
<td>chipseq package plot tests</td>
<td>2010-07-04 17:12:23 UTC</td>
<td>✔️</td>
<td>1m 10s</td>
<td>Results, Feedback, Log, Clone, Delete</td>
</tr>
<tr>
<td>1041</td>
<td>Seq Count</td>
<td>2010-07-04 01:10:16 UTC</td>
<td>✔️</td>
<td>59s</td>
<td>Results, Feedback, Log, Clone, Delete</td>
</tr>
<tr>
<td>1039</td>
<td>ChipSeqR plot test</td>
<td>2010-07-03 21:24:41 UTC</td>
<td>✔️</td>
<td>2h 30m 47s</td>
<td>Results, Feedback, Log, Clone, Delete</td>
</tr>
<tr>
<td>1040</td>
<td>PICS plot test</td>
<td>2010-07-03 21:27:57 UTC</td>
<td>✔️</td>
<td>57s</td>
<td>Results, Feedback, Log, Clone, Delete</td>
</tr>
<tr>
<td>1038</td>
<td>IRanges plot test</td>
<td>2010-07-03 21:20:42 UTC</td>
<td>✔️</td>
<td>6s</td>
<td>Results, Feedback, Log, Clone, Delete</td>
</tr>
<tr>
<td>1037</td>
<td>Genominator plot test</td>
<td>2010-07-03 21:16:50 UTC</td>
<td>✔️</td>
<td>26s</td>
<td>Results, Feedback, Log, Clone, Delete</td>
</tr>
<tr>
<td>1036</td>
<td>GenomeGraphs test plots</td>
<td>2010-07-03 21:14:15 UTC</td>
<td>✔️</td>
<td>27s</td>
<td>Results, Feedback, Log, Clone, Delete</td>
</tr>
</tbody>
</table>
New Job - Step 1 from 2 (script selection)

Script type:
- Public
- My Private Files
- CRdata Core
- Development
- Team
- raji crdata group

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>View</th>
<th>Edit</th>
<th>Destroy</th>
</tr>
</thead>
<tbody>
<tr>
<td>94</td>
<td>sequenceDifferentialExpression (RNA-seq)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>sequenceCountsPerTranscript (RNA-seq)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>PICS plot demo</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>Estimate ChIPseq fragment length</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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)
cctf
select chromosome (List)
chr10
chr11
chr12

Create Job or Cancel

History:
<table>
<thead>
<tr>
<th>When</th>
<th>Who</th>
<th>JobName</th>
<th>Time</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010-07-23 21:02:22 UTC</td>
<td>Hamid SUDO</td>
<td>ChIPseq Frag length</td>
<td>39s</td>
<td>Done</td>
</tr>
</tbody>
</table>
New Job

Choose script to run
Laslo et al Mac vs Neutrophil cell fate switch

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.

Filter by:
- Script type: Public
- Tags: Homo sapiens

- CRdata group
- My Private Files
- CRdata Core Development Team

Tag cloud:
- qsi
- false

History:

<table>
<thead>
<tr>
<th>When</th>
<th>Who</th>
<th>JobName</th>
<th>Time</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010-07-24 18:46:34 UTC</td>
<td>Hamid SUDO</td>
<td>affyDifferentialExpression</td>
<td>51s</td>
<td>Done</td>
</tr>
<tr>
<td>2010-07-24 18:44:40 UTC</td>
<td>Hamid SUDO</td>
<td>affyDifferentialExpression</td>
<td>51s</td>
<td>Done</td>
</tr>
<tr>
<td>2010-07-01 21:39:05 UTC</td>
<td>CRdata team</td>
<td>affyDifferentialExpression</td>
<td>51s</td>
<td>Done</td>
</tr>
<tr>
<td>2010-06-28 00:58:59 UTC</td>
<td>Martin Morgan</td>
<td>affyDifferentialExpression</td>
<td>51s</td>
<td>Done</td>
</tr>
</tbody>
</table>

Selected script details:

**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("ras", "omnraa"),
                          correct = TRUE)
```

**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.
Scripts:

create

view

Source code:

```r
version <- '0.0.6'
{
  require(Bioconductor)
  require(Genomewide)
  require(KEGG.db)
  require(RColorBrewer)
  require(splines)
  require(vsn)
}
```
Manage Jobs Queues

Use the interactive interface below to manage your Jobs Queues.

Create & delete processing queues

Manage Processing Nodes

Use the interactive interface below to manage your Processing Nodes.

Create & delete processing nodes for queues
Manage Jobs Queues

Use the interactive interface below to manage your Jobs Queues.

New Queue

Name
Hamid's Private Q

Auto-scale

Minimum nr. of processing nodes to be associated
0

Maximum nr. of processing nodes to be associated
0

Set nr. of jobs after auto-scale process for nodes starts
0

Set maximum idle time for a processing node
0

Set maximum waiting time for a job to be processed (in minutes)
0

AWS Key
HB private keys

Visibility
private

Create Queue or Cancel

Edit Queue

Name
Public

Auto-scale

Minimum nr. of processing nodes to be associated
3

Maximum nr. of processing nodes to be associated
6

Set nr. of jobs after auto-scale process for nodes starts
2

Set maximum idle time for a processing node
15

Set maximum waiting time for a job to be processed (in minutes)
3

AWS Key
Public Queue Keys

Visibility
public

Update Queue or Cancel
Choose memory and processing node type

Manage Processing Nodes

Use the interactive interface below to manage your Processing Nodes.

New Processing Node

Choose jobs queue

Hamid's private Q

AWS Key

HB private keys

Ec2 instance ami

64 bit

Ec2 instance type

m2.4xlarge

or manually create a node

Create Processing Node or Cancel
Manage Groups

Name: UW ENCODE DNase1 HS sites

Description: Data files and analysis scripts integrating 53 datasets to predict genome-wide. This group is open to all.

Read Me:

Please:

1. Note that some of the data here is available under the ENCODE embargo. If you are an approved member, you will be notified when the embargo date ends (check the UCSC ENCODE website for dates).

2. I have coded the scripts so that you can change the analysis parameters. However, the scripts can take a significant amount of time to process, so please launch 2 or more nodes and associated nodes to run these scripts.

I used data from all the UW ENCODE DNase1 assays except for cancer cell lines (I did not use these because they do affect HS loci). To remove some files from the analysis, or to add more files, I used the fileNames.csv* file (one filename per row, use MS Excel or a plain text editor to edit it, then reload the file into CRdata as a private file so as not to confuse other users.

Hamid Bolouri, August 2010

Members:
1. amir feizi  
2. Dimos Kapetis  
3. Hamid Bolouri  
4. Mikael Huss

Invite user to join group

Selected script details

Help text for CRdata.org script: “Consensus UW ENCODE DNase1 HS regions”

This script combines data from multiple UW ENCODE project DNase1 Hyper Sensitivity assays to predict genome-wide human cis-regulatory modules. See:

http://hgdownload.cse.ucsc.edu/goldenPath/hg18/encode/seq/wuEncodeUwDnaseSeq/

and

http://genome.ucsc.edu/cgi-bin/hgTrackU?db=hg18&wd=wuEncodeUwDnaseSeq

for files and project/assay details.

Input Parameters:

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Default Value</th>
<th>Parameter Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>fileNames</td>
<td>UW Encode DNase1 HS sites fileNames.csv</td>
<td>This is a list of filenames that were downloaded from the ENCODE UCSC FTP site with the script: “upload UW ENCODE DNase1 HS sites files”. The file names must exactly match those in your CRdata directory. This file should be plain text saved in comma separated values (.CSV) format, with one line per file name. For example, see the default file provided. Peaks with -log10(p-value) less than threshold are deemed too frequent to be true positives. The default value of 20 was selected on the basis of histograms of p-values showing a distinct and large population of peaks below this value.</td>
</tr>
<tr>
<td>p-value threshold</td>
<td>20</td>
<td>We remove non condition-specific peaks co-occurring in 11 or more datasets. Since there are typically 2 replicates per cell type, this corresponds to peaks co-occurring in 5-6 cell types.</td>
</tr>
<tr>
<td>co-occurrence threshold</td>
<td>11</td>
<td>We remove peaks spanning 5kb or longer because known cis-regulatory modules are usually shorter. These longer HS regions may indicate transcribed regions and/or structural features of chromosomes.</td>
</tr>
<tr>
<td>max peak width</td>
<td>5000</td>
<td>Annotate peaks within this distance from the nearest TSS.</td>
</tr>
<tr>
<td>distance to TSS</td>
<td>2000</td>
<td>Peaks that are less than this number of base pairs apart are merged.</td>
</tr>
<tr>
<td>min gap size</td>
<td>100</td>
<td>If true generates an output file annotating all gene-peak associations found</td>
</tr>
<tr>
<td>printAll</td>
<td>True</td>
<td>If true generates an output file annotating peaks less than 20kb downstream of nearest gene</td>
</tr>
<tr>
<td>printUpstream</td>
<td>True</td>
<td>If true generates an output file annotating peaks less than 20kb upstream of TSS</td>
</tr>
<tr>
<td>printDownstream</td>
<td>True</td>
<td>If true generates an output file listing names of all the genes associated with peaks</td>
</tr>
<tr>
<td>printUniqueGenes</td>
<td>False</td>
<td>If true generates an output file listing names of all the genes associated with peaks</td>
</tr>
</tbody>
</table>
New Job

Choose script to run
- Predict UW ENCODE DNase1 cis-regulatory modules
- Type at least 3 chars and select a name from the list.

Name your job

Choose Jobs Queue
- Bernstein lab Q (Paused)

Parameters

fileNames (Dataset)
- or upload a dataset file
- printAll (Boolean)
- printUpstream (Boolean)
- printDownstream (Boolean)
- printUniqueGenes (Boolean)
- merge peaks with gaps smaller than (bp) (Integer)
- 100

filter out DNase1 regions larger than (bp) (Integer)
- 5000

filter out DNase1 regions occurring > than (x) files (Integer)
- 11

Jsig10(p-value) selection threshold for DNase1 peaks (Float)
- 20

Annotate DNase1 peaks < minDistance (bp) from TSS (Integer)
- 2000

Create Job or Cancel

Dataset type:

Public
- My Private Files

UW ENCODE
- DNase1 HS sites

Bernstein lab
- FHCRC
- CRdata Core
- Development Team
- All

Tag cloud
- dnasel
- hb
- uw
- encode

Search datasets

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>View</th>
<th>Edit</th>
<th>Destroy</th>
</tr>
</thead>
<tbody>
<tr>
<td>393</td>
<td>peaksWithinMinDistanceDownstream.txt</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>392</td>
<td>peaksWithinMinDistanceUpstream.txt</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>309</td>
<td>UW Encode DNase1 HS sites fileNames.csv</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>325</td>
<td>Rep1Hom.narrowPeak.gz</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>326</td>
<td>Rep2Hnfneo.narrowPeak.gz</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>327</td>
<td>Rep2Hre.narrowPeak.gz</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>328</td>
<td>Rep1Hae.narrowPeak.gz</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>329</td>
<td>Rep1Huvcl.narrowPeak.gz</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Selected dataset details

Sample output file listing selected DNase1 HS regions less than 2000bp downstream of associated genes (see documentation for Bioconductor package ChIPpeakAnno for annotation details).

Tags: [hb encode uw dnasel]

History:

<table>
<thead>
<tr>
<th>When</th>
<th>Who</th>
<th>JobName</th>
<th>Time</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010-08-22 20:39:24 UTC</td>
<td>Hamid Bolouri</td>
<td></td>
<td>Update</td>
<td></td>
</tr>
<tr>
<td>2010-08-22 10:37:50 UTC</td>
<td></td>
<td></td>
<td>Create</td>
<td></td>
</tr>
</tbody>
</table>

Speed of Execution

Ease of Use

Quality of Docs

Quality of Data

+ Rate this dataset

* * * * * (5 star)  * * * * * (4 star)  * * * * (3 star)  * * * (2 star)  * (1 star)
Summary

Empowering bench biologists, 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
189 registered users since launch (August 8th 2010)

<table>
<thead>
<tr>
<th>City</th>
<th>Visits</th>
<th>Pages/Visit</th>
<th>Avg. Time on Site</th>
<th>% New Visits</th>
<th>Bounce Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>407</td>
<td>10.09</td>
<td>00:10:53</td>
<td>58.72%</td>
<td>40.54%</td>
</tr>
<tr>
<td></td>
<td>123</td>
<td>6.07</td>
<td>00:03:57</td>
<td>26.83%</td>
<td>46.34%</td>
</tr>
<tr>
<td></td>
<td>118</td>
<td>3.54</td>
<td>00:02:16</td>
<td>30.51%</td>
<td>64.41%</td>
</tr>
<tr>
<td></td>
<td>66</td>
<td>1.00</td>
<td>00:00:00</td>
<td>0.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td></td>
<td>44</td>
<td>5.02</td>
<td>00:04:17</td>
<td>68.18%</td>
<td>38.64%</td>
</tr>
<tr>
<td></td>
<td>41</td>
<td>1.02</td>
<td>00:00:05</td>
<td>0.00%</td>
<td>97.56%</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>2.67</td>
<td>00:01:07</td>
<td>23.08%</td>
<td>92.31%</td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>3.42</td>
<td>00:01:32</td>
<td>48.39%</td>
<td>38.71%</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>1.93</td>
<td>00:00:13</td>
<td>6.90%</td>
<td>93.10%</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>8.75</td>
<td>00:07:05</td>
<td>10.71%</td>
<td>60.71%</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>2.18</td>
<td>00:00:45</td>
<td>3.57%</td>
<td>82.14%</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>13.56</td>
<td>00:08:36</td>
<td>24.00%</td>
<td>52.00%</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>3.83</td>
<td>00:01:30</td>
<td>97.50%</td>
<td>50.00%</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>1.58</td>
<td>00:00:43</td>
<td>8.33%</td>
<td>91.67%</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>4.14</td>
<td>00:01:28</td>
<td>71.43%</td>
<td>66.67%</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>3.29</td>
<td>00:00:35</td>
<td>57.14%</td>
<td>57.14%</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>1.75</td>
<td>00:01:01</td>
<td>40.00%</td>
<td>65.00%</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>9.61</td>
<td>00:06:34</td>
<td>33.33%</td>
<td>27.79%</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>3.81</td>
<td>00:04:06</td>
<td>93.75%</td>
<td>56.25%</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>4.00</td>
<td>00:01:23</td>
<td>33.33%</td>
<td>6.67%</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>2.86</td>
<td>00:00:26</td>
<td>57.14%</td>
<td>50.00%</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>7.00</td>
<td>00:03:49</td>
<td>50.00%</td>
<td>35.71%</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>8.43</td>
<td>00:03:04</td>
<td>42.86%</td>
<td>35.71%</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>4.79</td>
<td>00:01:35</td>
<td>28.57%</td>
<td>14.29%</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>1.00</td>
<td>00:00:00</td>
<td>0.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>1.00</td>
<td>00:00:00</td>
<td>0.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>3.46</td>
<td>00:04:26</td>
<td>69.23%</td>
<td>46.15%</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>4.33</td>
<td>00:04:11</td>
<td>58.33%</td>
<td>33.33%</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>2.67</td>
<td>00:02:13</td>
<td>8.33%</td>
<td>50.00%</td>
</tr>
</tbody>
</table>
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)

**THE INCIDENCE OF ALKAPTONURIA: A STUDY IN CHEMICAL INDIVIDUALITY**

ARCHIBALD E. GARROD

Physician to the Hospital for Sick Children, Great Ormondstreet, Demonstrator of Chemical Pathology at St. Bartholomew’s Hospital

Eosinophil counts \( \left( \frac{X - \mu}{\sigma} \right) \) per unit volume in 4,458 Icelandic people


Proportion of individuals

% TH1 (INF-\( \gamma \) expressing) cells among CD4+ cells in blood

Duramad et al, Cancer Epidemiology, Biomarkers and Prevention, 2004, 13(9):1452–8
143 healthy middle-aged blood donors


Wnt pathway targets of environmental exposures

Comparative Toxicogenomics Database http://ctd.mdibl.org/
Thrombi in male sudden coronary death

- Acute thrombus
- Plaque rupture
- Plaque erosion
- Stable plaque

Age group

The human microbiome


The Human Microbiome Project
NY Times, 12th July 2010
Clinical assessment incorporating a personal genome

(Stephen Quake’s genome)

www.thelancet.com Vol 375 May 1, 2010

---

**Myocardial infarction**

<table>
<thead>
<tr>
<th>Gene*</th>
<th>SNP location</th>
<th>Patient genotype</th>
<th>LR</th>
<th>Studies†</th>
<th>Samples‡</th>
<th>Post-test probability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPA</td>
<td>rs3798220</td>
<td>CT</td>
<td>1.86</td>
<td>2</td>
<td>17031</td>
<td>2.0%</td>
</tr>
<tr>
<td>THBS2</td>
<td>rs8089</td>
<td>AC</td>
<td>1.09</td>
<td>1</td>
<td>4868</td>
<td>4.0%</td>
</tr>
<tr>
<td>LDLR</td>
<td>rs14158</td>
<td>GG</td>
<td>2.88</td>
<td>1</td>
<td>3542</td>
<td>10.6%</td>
</tr>
<tr>
<td>LIPE</td>
<td>rs11630220</td>
<td>AG</td>
<td>1.15</td>
<td>1</td>
<td>3542</td>
<td>12.0%</td>
</tr>
<tr>
<td>ESR2</td>
<td>rs1271572</td>
<td>CC</td>
<td>0.73</td>
<td>1</td>
<td>3089</td>
<td>9.1%</td>
</tr>
<tr>
<td>FXN</td>
<td>rs35410698</td>
<td>GG</td>
<td>1.03</td>
<td>1</td>
<td>1094</td>
<td>9.4%</td>
</tr>
<tr>
<td></td>
<td>rs3793456</td>
<td>AA</td>
<td>0.94</td>
<td>1</td>
<td>1094</td>
<td>8.9%</td>
</tr>
</tbody>
</table>
DNA sequence variations affecting cellular signaling genes in two individuals

**BCR** (805) Watson Venter
26 29 23

**Ca** (2740) Watson Venter
49 53 46

**ErbB** (1122) Watson Venter
17 15 12

**Hh** (501) Watson Venter
7 17 15

**JakStat** (1468) Watson Venter
55 22 25

**MAPK** (3792) Watson Venter
63 66 61

**mTOR** (1068) Watson Venter
11 9 9

**Notch** (1250) Watson Venter
17 3 16

**Phosphatidylinositol** (1010) Watson Venter
21 34 27

**TCR** (1141) Watson Venter
27 22 20

**TGFβ** (720) Watson Venter
7 20 8

**TLR** (1033) Watson Venter
23 20 19

**VEGF** (883) Watson Venter
18 32 15

**Wnt** (1273) Watson Venter
29 26 27

---

Watson total
(1468 total entries in dbSNP)

Venter total

Overlapping genes:

- BCR, Ca, ErbB, Hh, JakStat
- MAPK, mTOR, Notch, Phosphatidylinositol, TCR
- TGFβ, TLR, VEGF, Wnt

---

Watson + Venter overlap
(non-synonymous + frameshift)
To predict an individual’s genomic susceptibilities, we will need to integrate data from:

- Genome annotation DBs
- RNA structure & function DBs
- Protein structure & function DBs
- Pathway structure & function DBs
- Mutation effect prediction algorithms
- Environmental effects DBs
- Drug effects and interactions DBs
- Electronic health records
- Pathway-based calculation of interaction effects
- ...

What is the appropriate computational infrastructure?
# Manage Groups

Use the interactive interface below to manage your Groups.

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

**Name:** Example Group  
**Description:** About Example Group...  
**Read Me:** More about Example Group  
**Members:**  
1. Hamid Bolouri [Cancel membership]  
   
[Invite user to join group]