ChromNet: Learning the human chromatin network from all ENCODE ChIP-seq data

http://chromnet.cs.washington.edu

ENCODE users meeting 2016
The chromatin network*
What we call the network of factors that interact to regulate the genome

*Lundberg et al. Genome Biology 2016
Motivating example
Illustrating the value of conditional dependence

True interactions

Factors are variables

Genomic positions are samples

Co-occurrences

Conditional dependence

\[ A = \beta_B B + \beta_C C + \beta_D D \]
Measuring many factors is important
Conditional dependence get better with a more complete model

What if we didn’t measure C? Then A and B get connected.
Group graphical model
Making conditional dependence robust to redundancy

Without redundancy $\Sigma^{-1}$
With redundancy $\Sigma^{-1}$

Group Graphical Model $G$

Redundant variables cause lost or unstable connections.
Measuring accuracy improvements
Recovery of known interactions and relationships

Within cell types
- GroupGM
- Inverse Correlation
- Correlation

Between cell types
- GroupGM
- Inverse Correlation
- Correlation

Validated novel interaction
MYC
HCFC1

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>K562 (236)</th>
<th>GM12878 (143)</th>
<th>H1-hESC (84)</th>
<th>HepG2 (115)</th>
<th>A549 (94)</th>
<th>HeLa-S3 (85)</th>
<th>MCF-7 (55)</th>
<th>SK-N-SH (44)</th>
<th>Ishikawa (25)</th>
<th>HCT116 (28)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fold Enrichment over Random</td>
<td>3.2</td>
<td>2.5</td>
<td>2.0</td>
<td>1.5</td>
<td>1.0</td>
<td>2.5</td>
<td>3.0</td>
<td>3.5</td>
<td>4.0</td>
<td>4.5</td>
</tr>
</tbody>
</table>

Binning Protein-protein interaction enrichment
GroupGM

Global network protein processing
(B)
Learning/visualization
(C)
Evaluation, validation, and analysis

HCFC1
Context specificity
Identifying which genomic regions drive an interaction

CTCF and SIX5 associate in the presence of ZNF143.

We removed ZNF143 from ChromNet then checked which genomic regions drove the CTCF-SIX5 edge.

They matched the held-out ZNF143 better than even the CTCF and SIX5 tracks themselves!
Global network
All cells types integrated into a single network.
Explore, and even integrate your own data!
http://chromnet.cs.washington.edu
This work was supported by a National Science Foundation (NSF) Graduate Research Fellowship to SML (DGE-1256082), a NSF grant to SIL (DBI-1355899), a Natural Sciences and Engineering Research Council of Canada grant to MMH (RGPIN-2015-03948), a Canada Research Chair in Molecular Oncology to LZP, a Canadian Institute for Health Research grant to LZP and BR (MOP-275788), and a Canadian Breast Cancer Foundation Ontario Region Doctoral Fellowship to WBT. Cloud computing resources for this research were generously provided by Google.
PLA results
Following up on the MYC-HCFC1 network edge
Histone mark / writer connections

fold enrichment over random

edge density in recovered network

- GroupGM
- Inverse correlation
- Correlation