Identifying Cooperativity Among Transcription Factors Controlling the Cell Cycle In Yeast

Nilanjana Banerjee\textsuperscript{1,2} and Michael Q. Zhang\textsuperscript{1}

Nucleic Acids Research, 2003, Vol.31, No.23

\textsuperscript{1}Cold Spring Harbor Laboratory \textsuperscript{2}George Mason University

Presented by...
Mark Connell
with Julie Blackwood and Tristan Richards
Overview

- Transcription factors, gene expression
- Banerjee’s and Zhang’s goals
- Determining cooperativity
  - Comparing TFs bound alone and together
- Assessing biological relevance
  - Test ability to identify target genes
- Web tool to explore cooperativity
- Discussion
- My research
Transcription factors

Precise transcriptional control is essential to gene expression and regulation
- Control through cooperativity
- Three different modes of transcription factor interactions considered

http://www.accessexcellence.org/AB/GG/control_Express.html
Other relevant work

- Previous work studying cooperativity used computationally derived motif combinations... accurate?
- Genome-wide location data provides more reliable model for TF binding site interactions (*in vivo*)
- Genome-wide gene expression data
Banerjee’s and Zhang’s Goals:

- Use novel method to reveal how multiple TFs cooperate to regulate transcription in cell cycle
- Combine genome-wide gene expression data and chromatin immuno-precipitation (ChIP-chip) location data
- Compare coupled and independent binding of TFs to assess cooperativity
- Create a graphical web tool for researchers about cooperativity
Methods overview

- Use ChIP-chip genome-wide binding for 113 yeast regulators (TFs)
- Determine which gene pairs’ expressions are correlated
- Determine cooperativity using expression correlation scores for pairs of TFs versus TFs alone
- Use multivariate hypergeometric distribution to assess significant cooperativity
Selecting TF groups

- For all pairs of TFs, selected three different sets of target genes based on significant binding:
  - TF A but not B \((A \cap \neg B)\)
  - TF B but not A \((-A \cap B)\)
  - both TFs A and B \((A \cap B)\)
  - (at least five genes required for each set)

- Assumption: TFs are cooperative when BOTH A and B bind

For what genes is it essential that both A and B bind?
Correlation of gene pairs

- Calculated expression correlation scores
  - Quantify clustering of expression profiles
  - Performed for each set of genes
- Score \((EC_G)\) is fraction of gene pairs in a set \(G\) with correlation higher than threshold \(\lambda T\)
  - Threshold: 95\(^{th}\) percentile correlation value for all the pair-wise correlations between 1000 randomly chosen genes

High score, high correlation in a set
Assessing significant cooperativity

- Look at scores for different sets
- A combination of TFs were considered cooperative if:
  - $(EC_G)_A \cap B >> (EC_G)_{-A} \cap B$
  - $(EC_G)_A \cap B >> (EC_G)_A \cap -B$
Determining correlation of gene pairs

From ChIP Data

Set of promoters with binding of TF

Expression Profile

Expression Correlation Score

EC=0.05

EC=0.22!!

EC=0.06
Multivariate Hypergeometric Distribution(!)

\[ P(m_a, m_{ab}, m_b \mid n_a, n_{ab}, n_b, N, M) = \frac{C_{n_a}^m C_{n_{ab}}^{m_{ab}} C_{n_b}^{m_b}}{C_N^M} \]

\( N = n_a + n_b + n_{ab} ; M = m_a + m_b + m_{ab} \)

\( N \) is the total number of gene pairs, composed of the 3 sets

\( M \) is the total number of correlated gene pairs

\[ C_N^M = \binom{N}{M} = \frac{N!}{M!(N-M)!} \]
More MHD...

- Consider ALL possible combinations of $x_a + x_b + x_{ab}$, such that...

\[ \sum_{i=\{a,b,ab\}} x_i = \sum_{i=\{a,b,ab\}} m_i \]

And... the sum of all probabilities calculated where $x_{ab} > m_{ab}$, the cooperativity p-value becomes...

\[ P_C = P(x_{ab} \geq m_{ab}) = \sum_{x_{ab} \geq m_{ab}} P(m_a, m_{ab}, m_b \mid n_a, n_{ab}, n_b, N, M) \]
More MHD...

- Multiple hypothesis tests requires corrections to represent true alpha level
  - Bonferroni correction (fewer cooperative)
  - Holm’s Correction
- Also consider TF pairs that were not significantly cooperative after Holm’s correction included because biological evidence of potential cooperativity
Results: cooperativity!

- 31 out of 261 possible TF pairs showed significant cooperativity before p-value adjustments or due to biological relevance
  - 12 out of 31 TF pairs backed by literature
  - Depending on correction method, 12-14 TF pairs significantly cooperative by p-value
- 6 of 7 most cooperative pairs have literature evidence

Results from method are somewhat accurate!
Results: cooperativity!

- As expected, many of cooperative TF pairs come from cell cycle
- Other TF pairs are part of environmental stress response, metabolism
- Some TF pairs do not belong to defined functional groups, need further characterization
An example: mcm1 and ndd1

- Statistically significant
  - 2\textsuperscript{nd} smallest Pc value out of 261 TF pairs
  - Valid cooperativity considering both alpha level corrections

- Biologically significant
  - Mcm1 does not “appreciably activate” either the SWI5 or CLB2 promoter
  - Research suggests that ndd1 is required for switching on genes during G2-M
Assessing biological relevance of target genes

- An Open Reading Frame (ORF) is a sequence of codons (from DNA or RNA)
- Compared sets of target genes of Mcm1 and Fkh2 derived via:
  - ChIP- and expression-based analysis
  - Motif- and expression-based analysis
- Compared these results with experimentally established target genes (Clb2 cluster)
  - ChIP method identified more targets
- Not a lot of overlap with Clb2 cluster...
Overlapping ORFS

Motif-based (30)

ChIP-chip (34)

Cdc20
Ace2
Clb2

Bud4
Rax2
Mtl1
Alk1
YIL158W
YJL051W

Clb2 Cluster (36)

7 1 10
Relevance of target genes

- Data for ChIP-positive target genes of Mcm1 and Fkh2
- Expression correlation of target genes of core profile of known targets
- Known targets circled in red
- Genes close to that set are likely additional targets of Mcm1 and Fkh2
The Visualization Tool

- Select any pairwise combination of TFs
  - Generates three lists of ORFs whose promoters have binding to TFs, representing the different sets of TFs A and B (A ∩ B, -A ∩ B, and A ∩ -B)
  - Creates a graph of each set’s gene expression profiles
  - Links for each ORF to explore biological relevance
- Select a significance level (Pc)
  - Creates TF cooperativity network between different functional categories
Revisiting mcm1 and ndd1: Expression profiles

MCM1+NDD1+, EC=0.69

MCM1+NDD1-, EC=0.34

NDD1+MCM1-, EC=0.36
Cooperativity network

- User defined significance levels
- Colors of edges relates to confidence in cooperativity
- Concentration in cell cycle
Discussion

- Combination of different data sources
- Lack of overlapping ORFs
  - Ways to increase accuracy
- Further research to characterize potentially cooperative TFs
- Useful in visualizing cross talk between functional groups
  - Allowing uploading of gene expression and binding data
- Errors
My Research Project

- Transcription factor binding site (TFBS) modeling
- Binding site sequences are varying for each TF
- Current work considers each position on sequence to contribute independently (of other positions) to binding
- Create model that considers dependency of one position in model upon another
  - Mutual Information Content
- Test model’s ability to identify known TFBSs, limit false negatives
References

Acknowledgements...

- Julie Blackwood and Tristan Richards
- Takis Benos
- Rajan Munshi
- BBSI Program n Peeps
- Pitt and Duquesne, and PSC and CMU
- Caterers Deanna Nachreiner and Carby Torok

THANKS!!!