

Co-occupancy networks for histone modifications and chromatin associated proteins

Martin Vingron

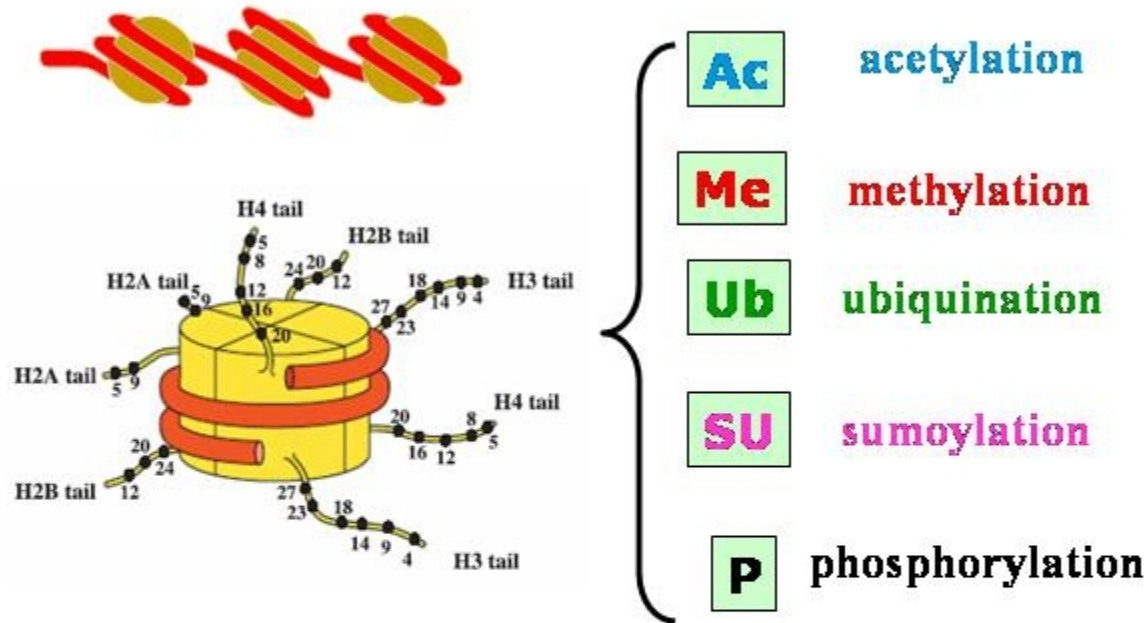
MPI for Molecular Genetics

Acknowledgements: Ho-Ryun Chung, Rosa Karlic, Julia Lasserre, Juliane Perner

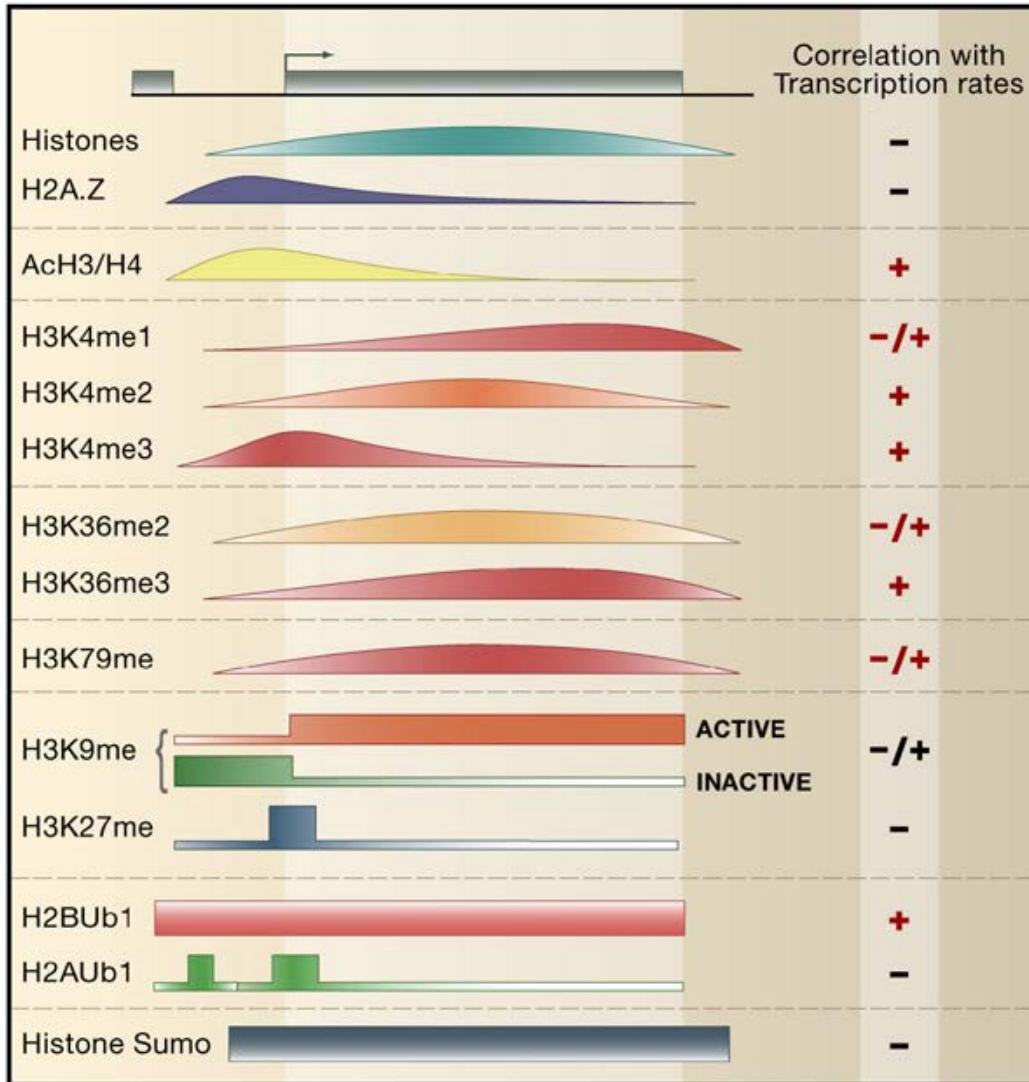
Co-occupancy Networks

- Can we predict gene expression? Predict from what? Histone modifications?
- Do „things“ occupy DNA together?
- Histone modification networks
- Partial correlation, Gaussian Graphical Models
- Histone modifications plus chromatin modifiers
- Compare to: Gene expression networks, e.g. BNs, $p \gg n$ problem

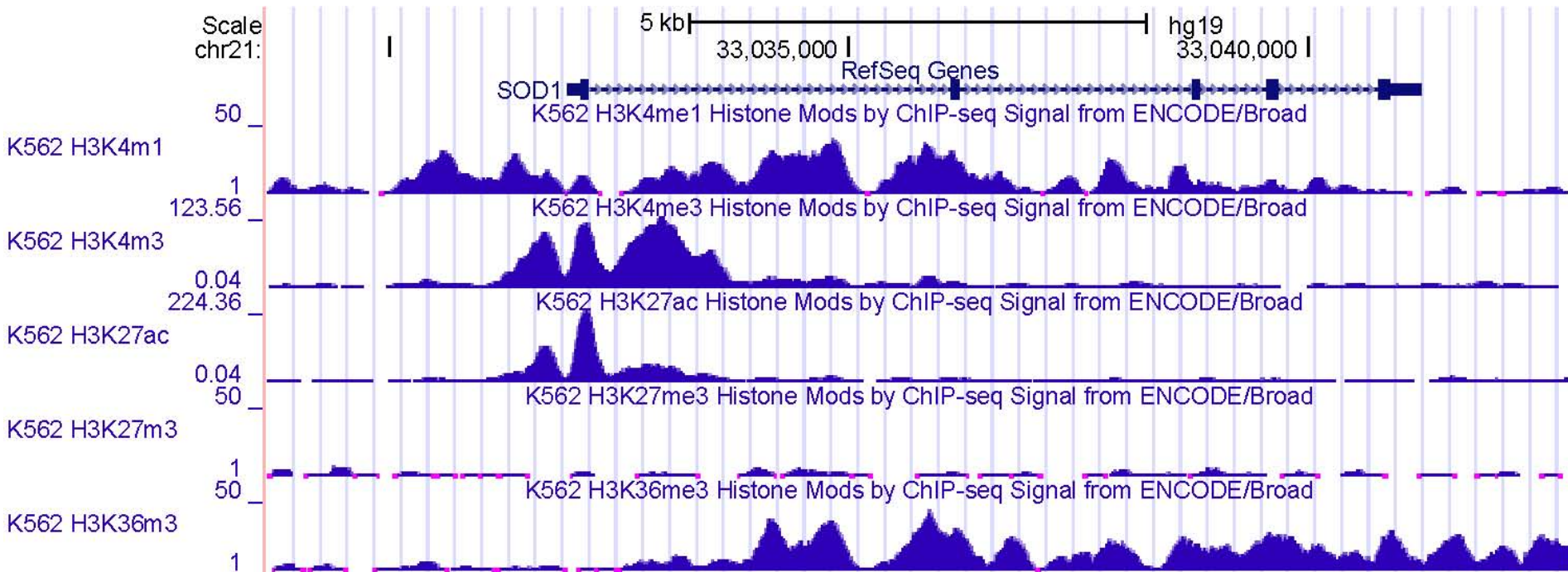
Histone modifications



The figure illustrates nucleosome models and major posttranslational modifications which play essential roles in gene expression regulation and disease processes



From: Li, Carey, Workman (2007) Cell 128:707-719



High-Resolution Profiling of Histone Methylations in the Human Genome

Artem Barski,^{1,3} Suresh Cuddapah,^{1,3} Kairong Cui,^{1,3} Tae-Young Roh,^{1,3} Dustin E. Schones,^{1,3} Zhibin Wang,^{1,3} Gang Wei,^{1,3} Iouri Chepelev,² and Keji Zhao^{1,*}

18 Histone acetylations

Combinatorial patterns of histone acetylations and methylations in the human genome

Zhibin Wang^{1,5}, Chongzhi Zang^{2,5}, Jeffrey A Rosenfeld³⁻⁵, Dustin E Schones¹, Artem Barski¹, Suresh Cuddapah¹, Kairong Cui¹, Tae-Young Roh¹, Weiqun Peng², Michael Q Zhang³ & Keji Zhao¹

Gene expression data

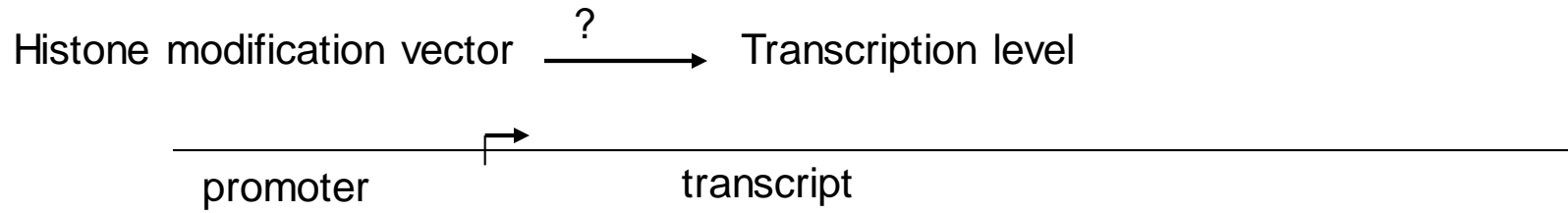
Dynamic Regulation of Nucleosome Positioning in the Human Genome

Dustin E. Schones,^{1,2} Kairong Cui,^{1,2} Suresh Cuddapah,¹ Tae-Young Roh,¹ Artem Barski,¹ Zhibin Wang,¹ Gang Wei,¹ and Keji Zhao^{1,*}

All data from a single cell type: human CD4+ T-cells

Control: CD4+ goat IgG and CD4+ rabbit IgG (Wang et al., 2009)

Histone modifications and transcription level



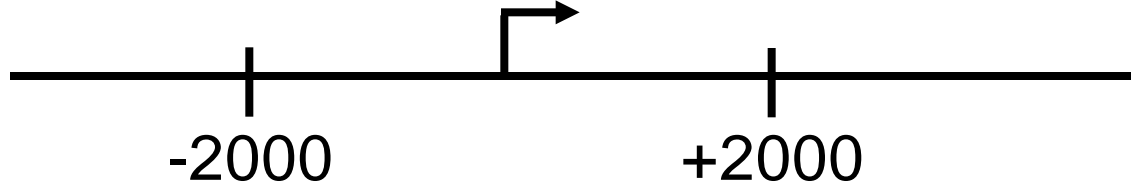
Focus on

- Sum of tags in promoter
- For many histone modifications
- in one and the same cell line.

Data from Keiji Zhao lab.

Results in a data matrix size #promoters x #histone modifications.

Model



Determine the number N_i of tags mapping to ± 2000 base pairs from the TSS



Transform N_i by
 $X_i = \log(N_i + \alpha_i)$

α_i (pseudocount) is chosen such that the correlation with the expression value is maximal

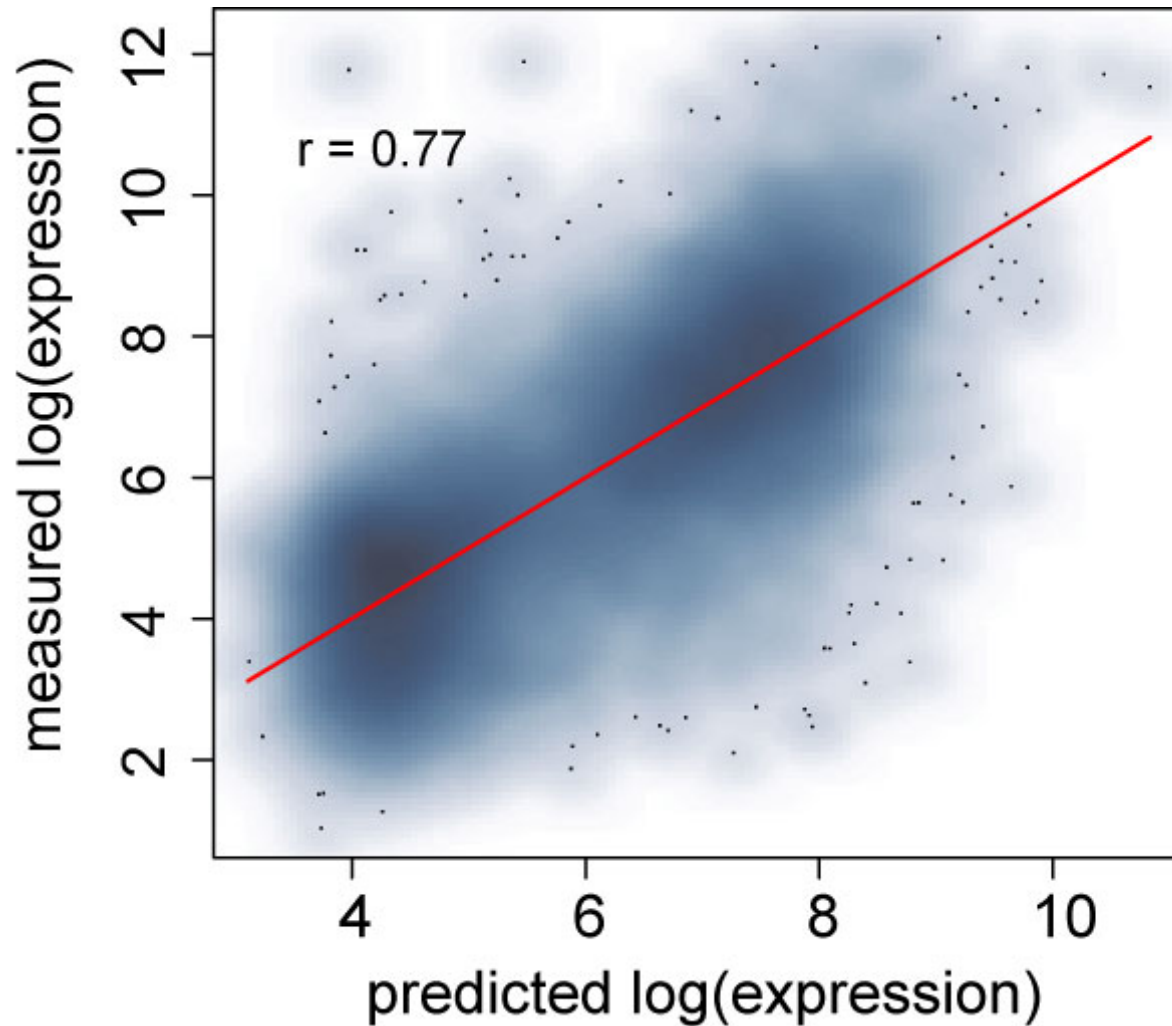


Standard linear regression

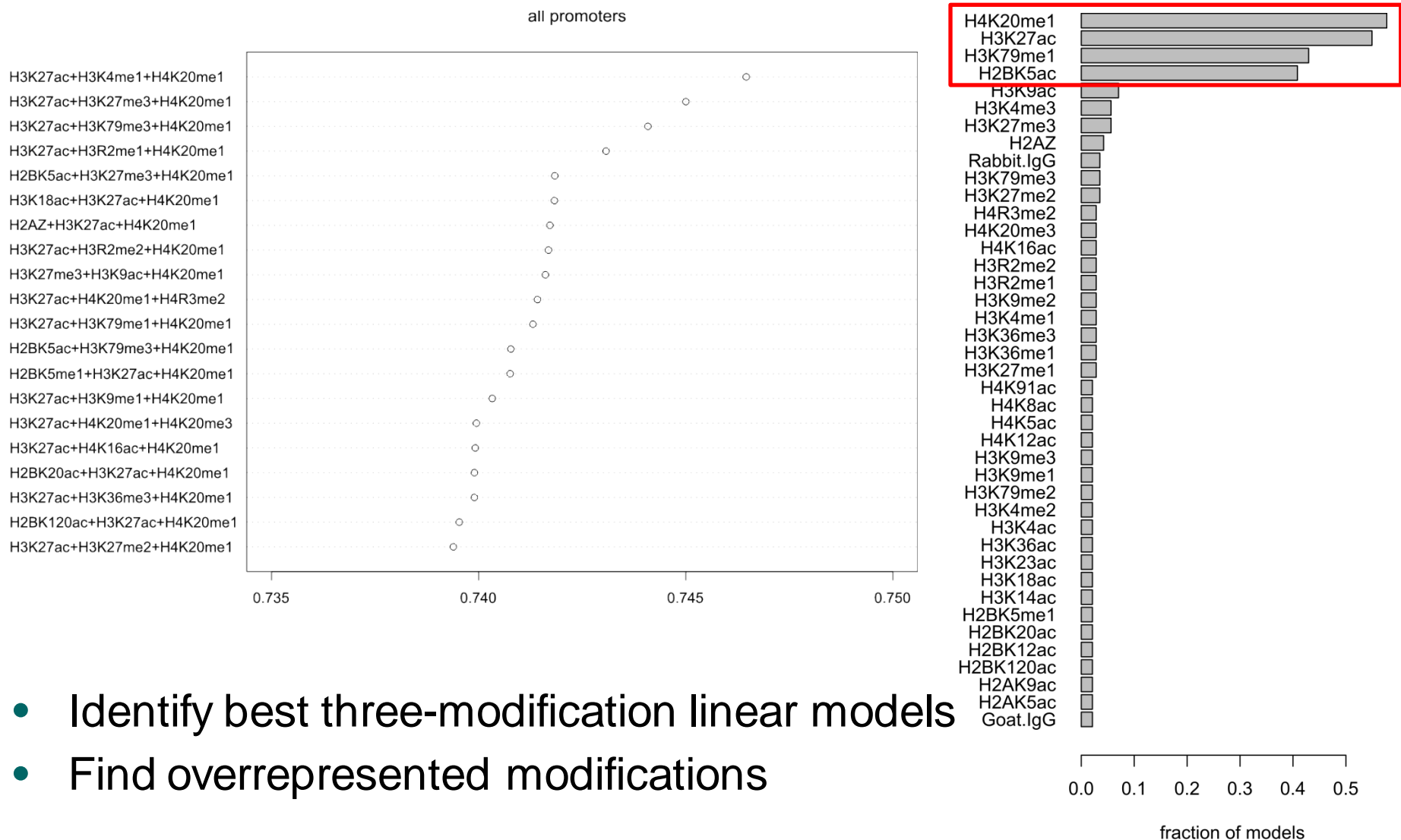
$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i,$$

$i=1, \dots, n$

Histone modifications and transcription level

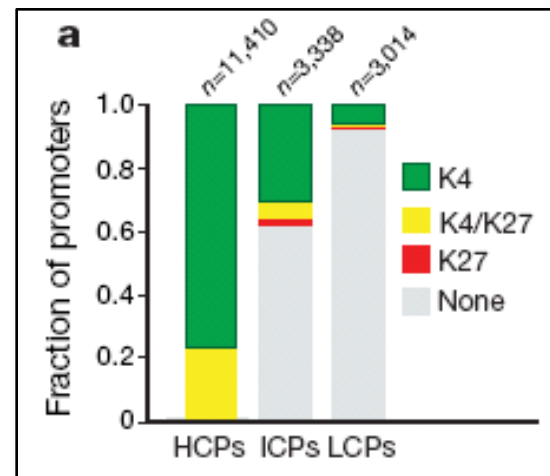
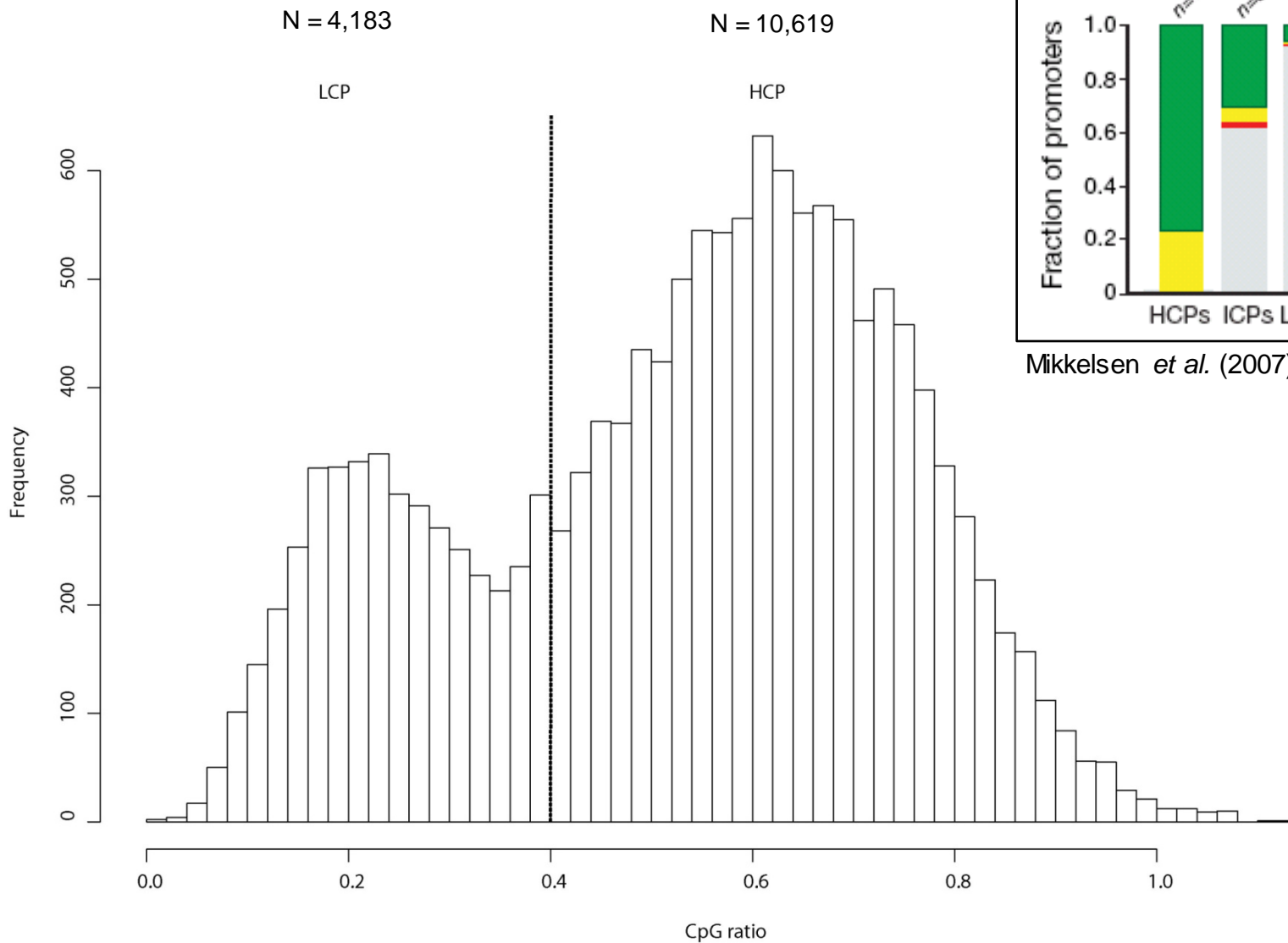


Feature Selection



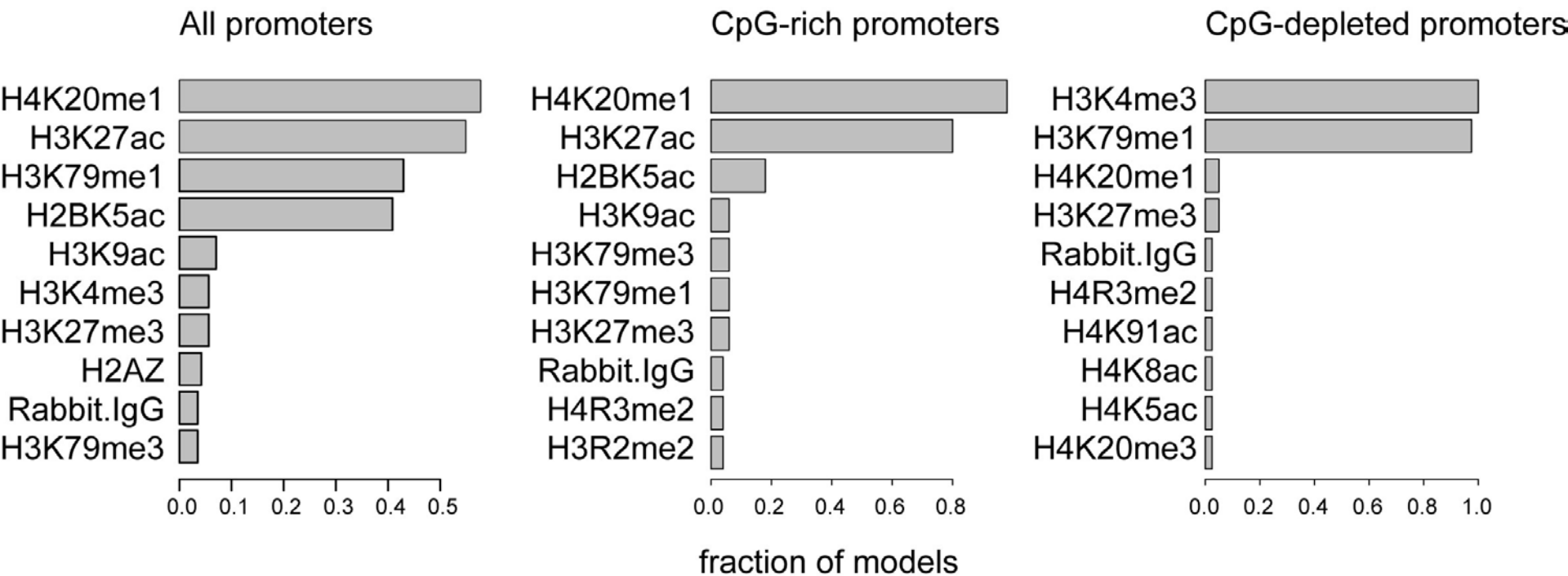
- Identify best three-modification linear models
- Find overrepresented modifications

Human Promoter Classes



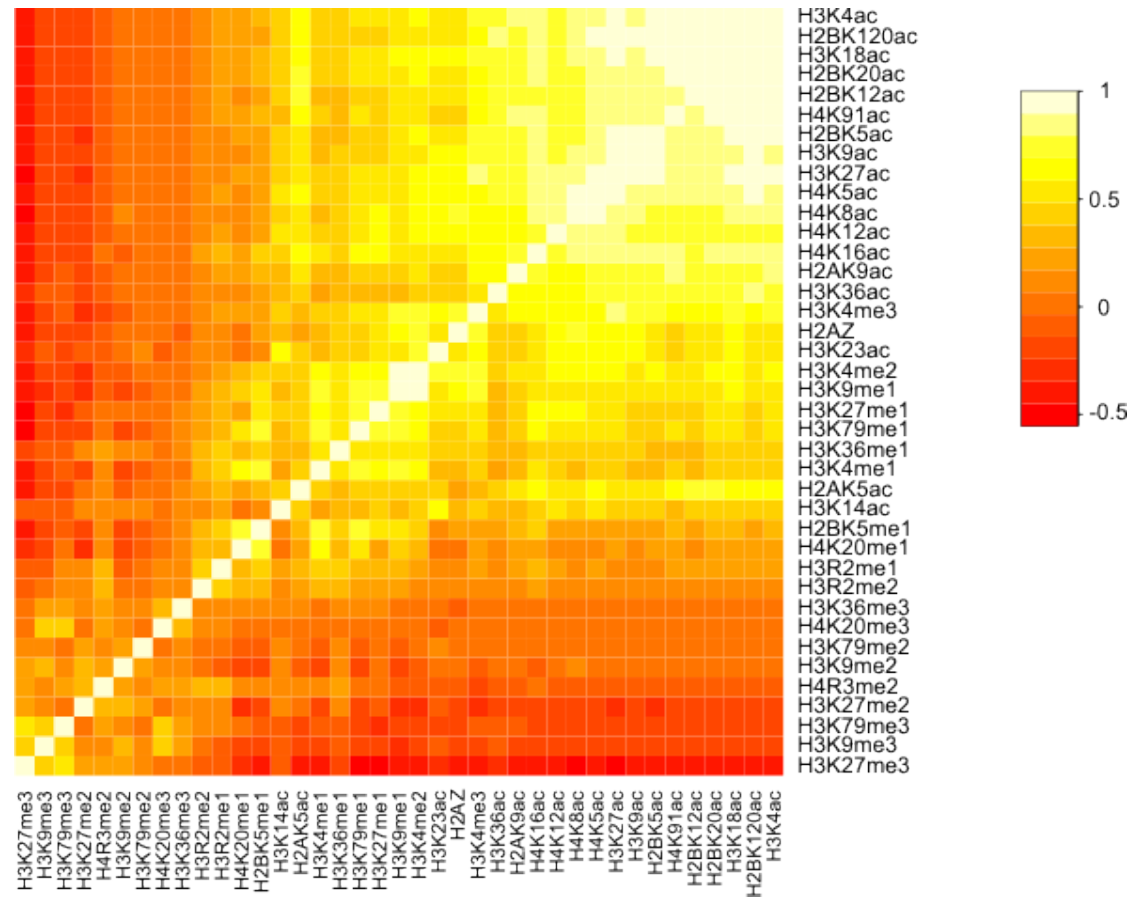
Mikkelsen *et al.* (2007) *Nature* **448**, 548

Informative modifications stratified by CpG contents



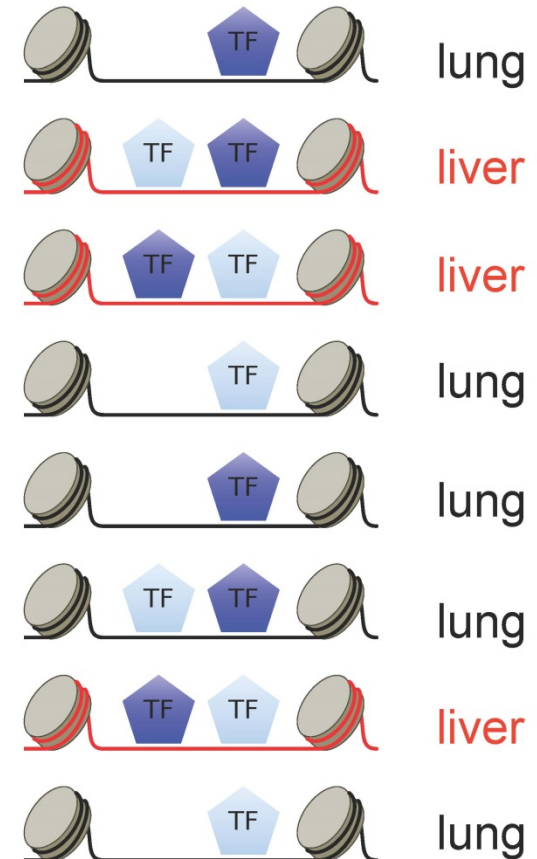
Correlations among modifications

Correlation between different histone modifications in the promoter regions of 14,802 human genes

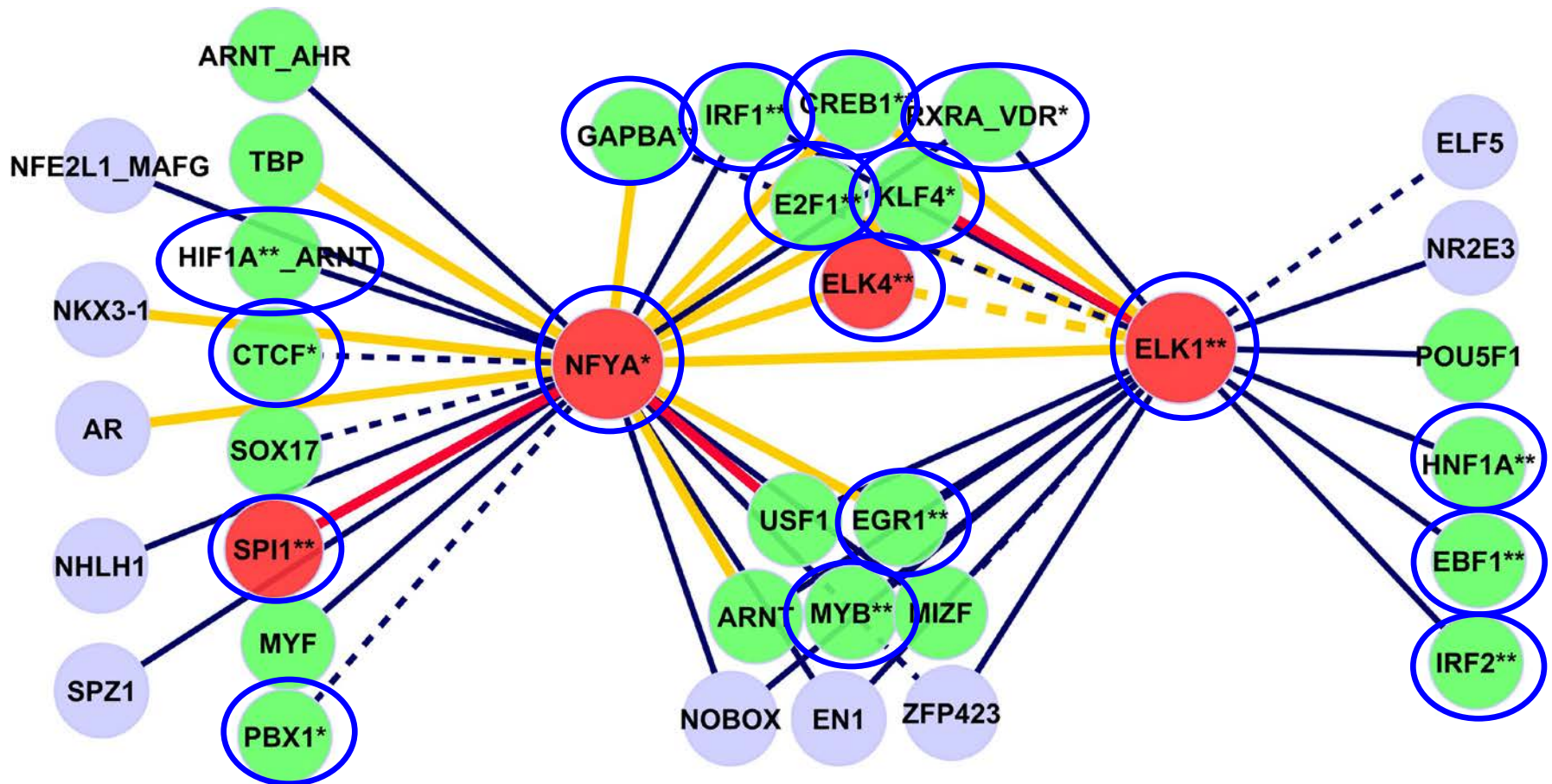


Detection of co-regulating TFs

- non-random binding of a TF pair on hypersensitive sites
- cell-type/tissue specific
- **TFs**: represented as ranked list of hypersensitive sites ordered by binding affinity
- **idea**: *TF pairs with similar top-ranked lists might co-regulate the transcription of their target genes*



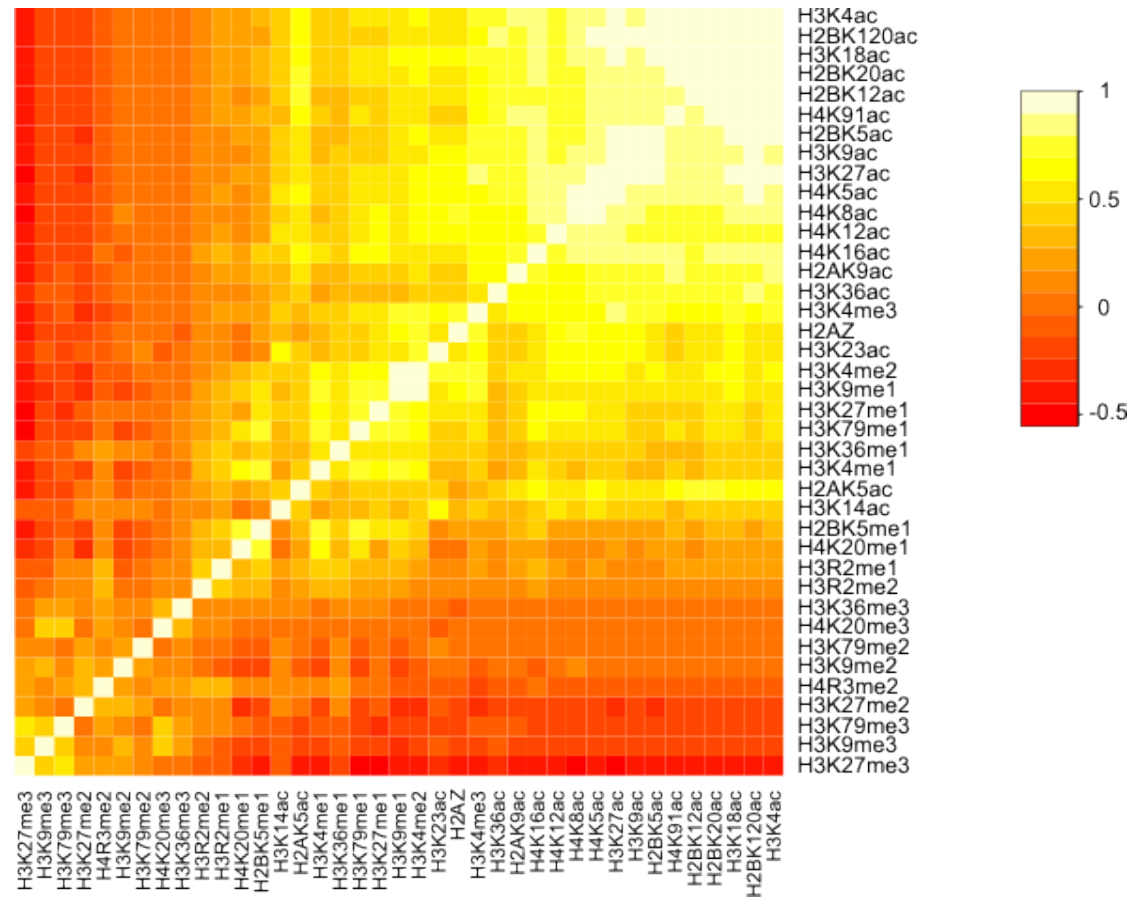
Predicted interactions in hematop. stem cells



Hematopoiesis

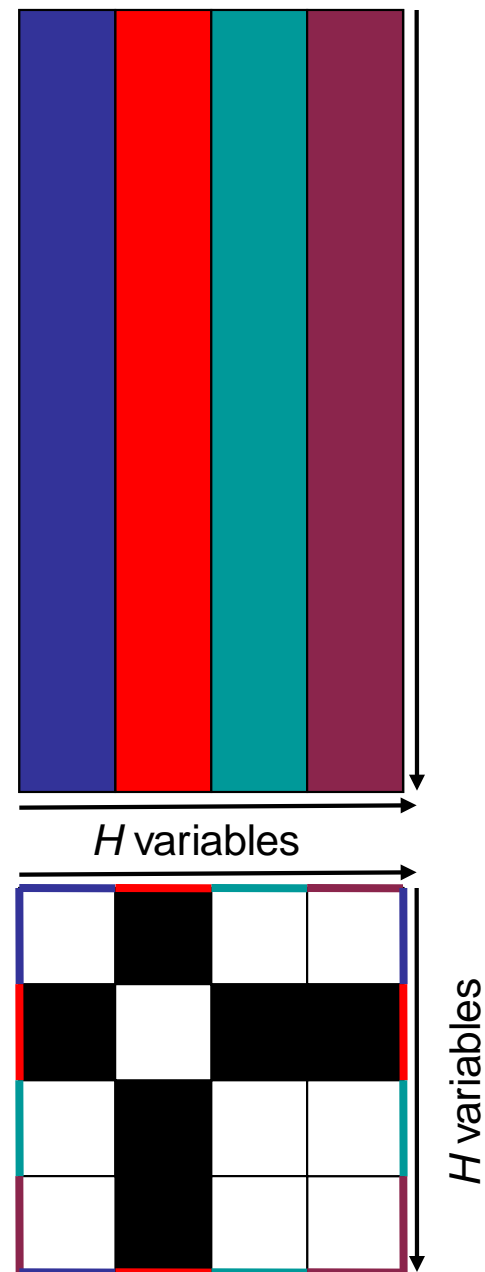
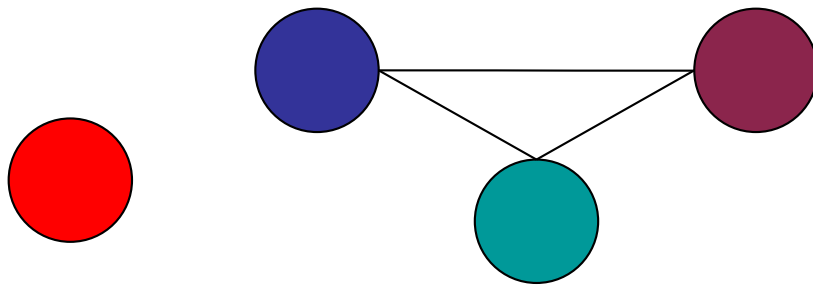
Correlations among modifications

Correlation between different histone modifications in the promoter regions of 14,802 human genes



From correlations to partial correlations (Gaussian Graphical Models)

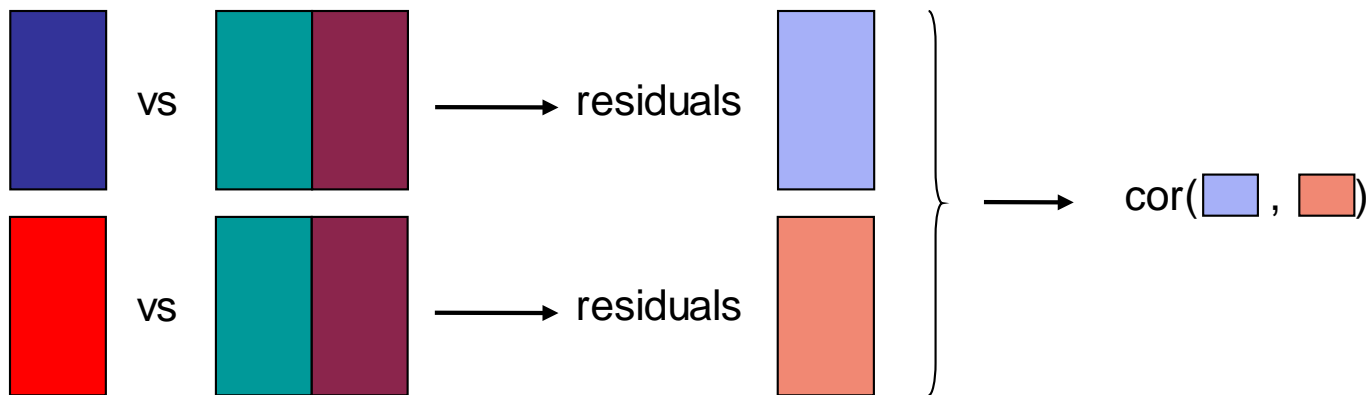
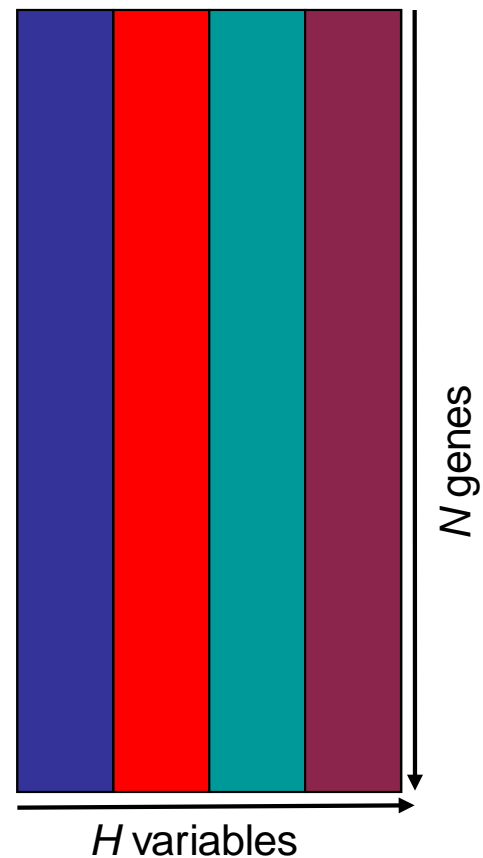
- Data matrix of dimension $N \times H$
 - N : number of genes
 - H : number of variables
- For each pair (h_i, h_j) of variables compute the correlation c_{ij} between r_i and r_j

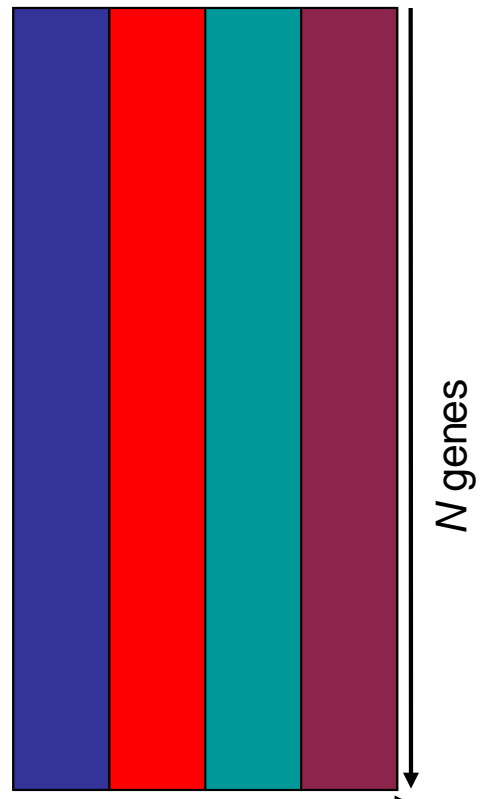
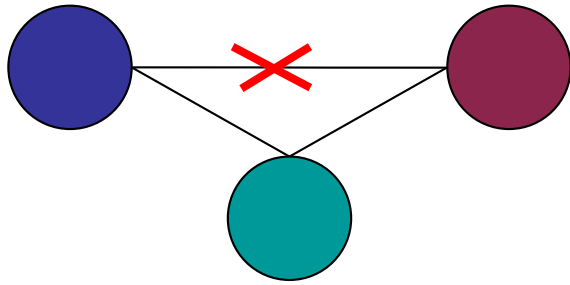
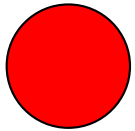


Partial correlation coefficient for variables i and j :

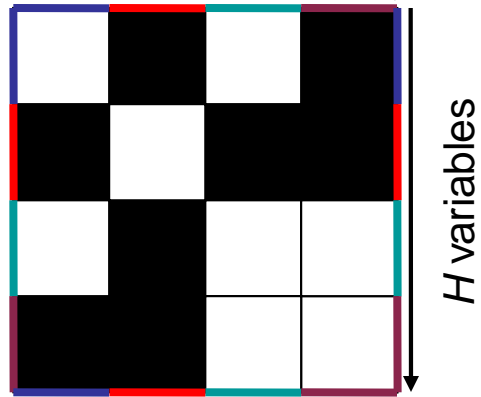
Regress i and j on the remaining variables.
Determine the two sets of residuals after explaining i and j .

Partial correlation between i and j is defined as the correlation between the two vectors of residuals.





H variables



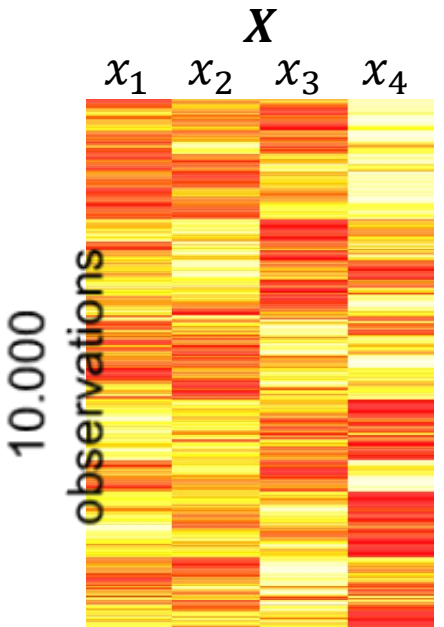
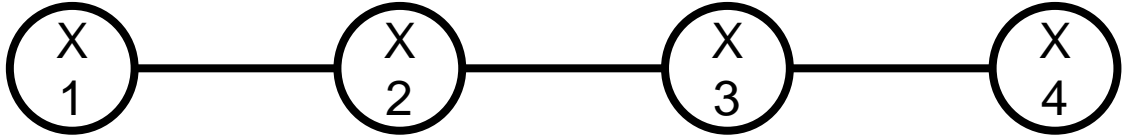
Theorem on how to compute partial correlations:

Partial correlation coefficients can be computed from the entries of the inverse of the variance-covariance matrix.

Theorem on the meaning:

Under the assumption that the variables are multivariate Gaussian, the partial correlation $\rho_{XY.Z}$ is zero if and only if X is conditionally independent from Y given \mathbf{Z} .

Example



$Cov(X) = \Sigma$

	X_1	X_2	X_3	X_4
X_1	1.0	2.0	6.0	3.0
X_2	2.0	5.0	6.0	6.0
X_3	6.0	6.0	1.0	1.0
X_4	3.0	7.6	23.1	12.6

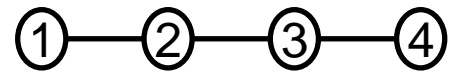
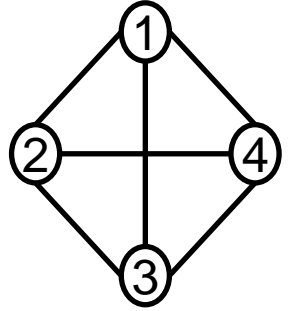
$\Sigma_{ij} = 0$
 \leftrightarrow
 Independence

$Cov(X)^{-1} = \Sigma^{-1}$

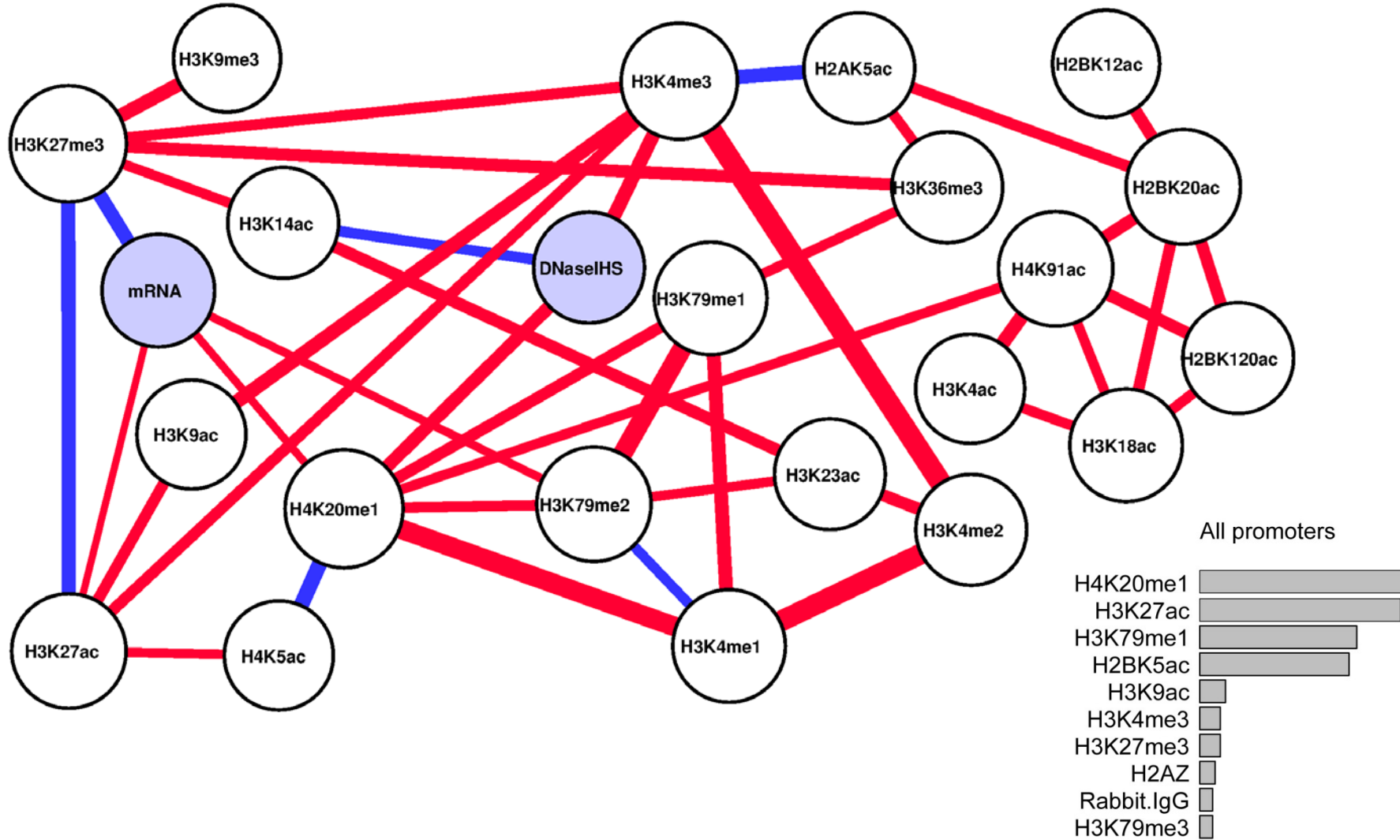
	X_1	X_2	X_3	X_4
X_1	5.0	-2.0	0.0	0.0
X_2	-2.0	5.0	0.0	0.0
X_3	0.0	0.0	-0.5	1.0
X_4	0.0	0.0	-0.5	1.0

$\Sigma_{ij}^{-1} = 0$
 \leftrightarrow
 Conditional independence

$X_1 \sim N(0, 1)$
 $X_2 \sim N(2X_1 + 1, 1)$
 $X_3 \sim N(3X_2 - 0.5, 1)$
 $X_4 \sim N(0.5X_3, 1)$



CD4+ network



Odds and Ends

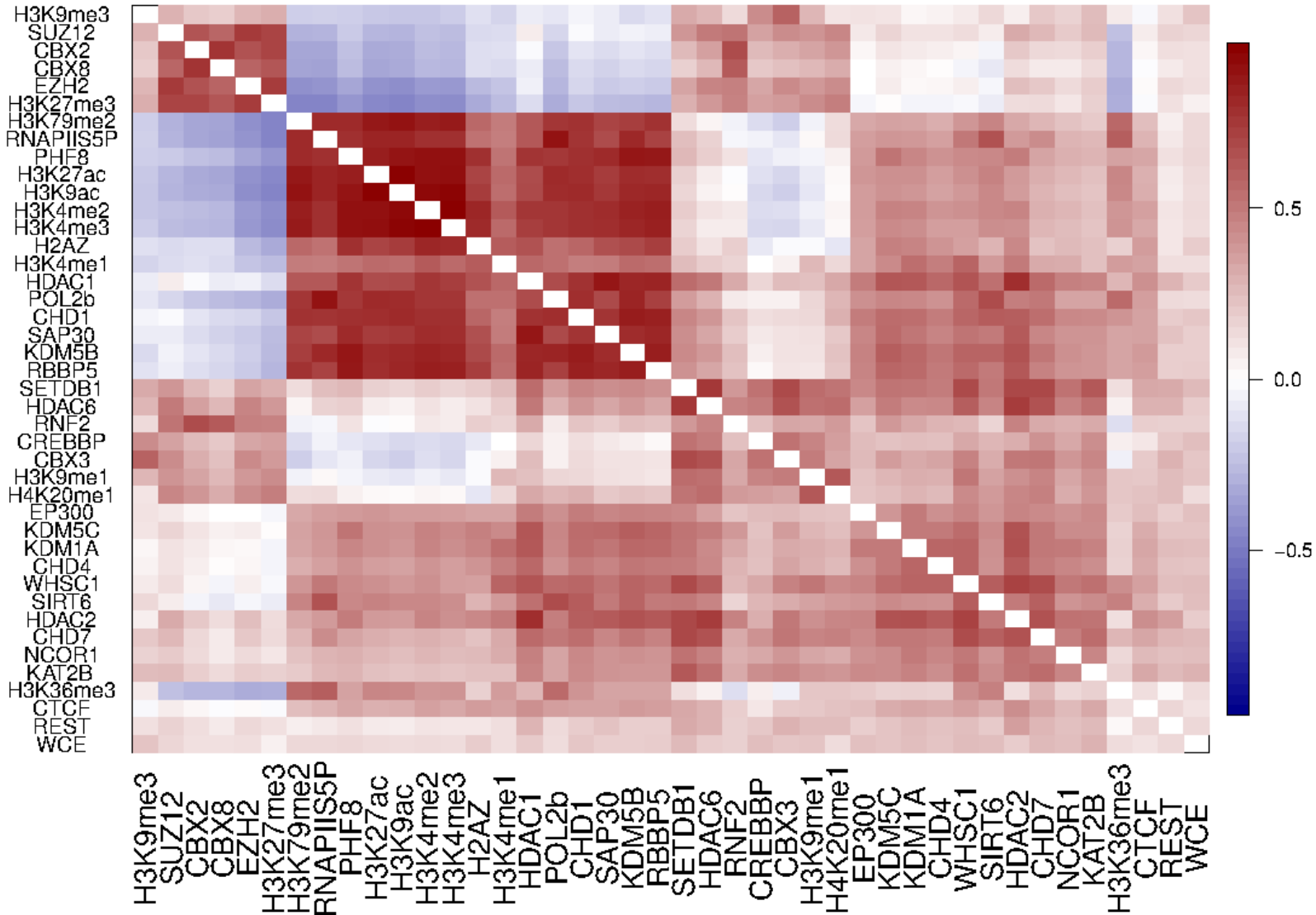
- We use rank-sorted data (corresponding to a rank correlation coefficient)
- There is huge number of entries in the vectors for which we compute partial correlation coefficients. Therefore we get tiny p-values.
- Remedy: Choose a p-value cutoff and resample from the promoters.
- Accept edges with more than, say, 70% bootstrap support.

Enter chromatin modifiers

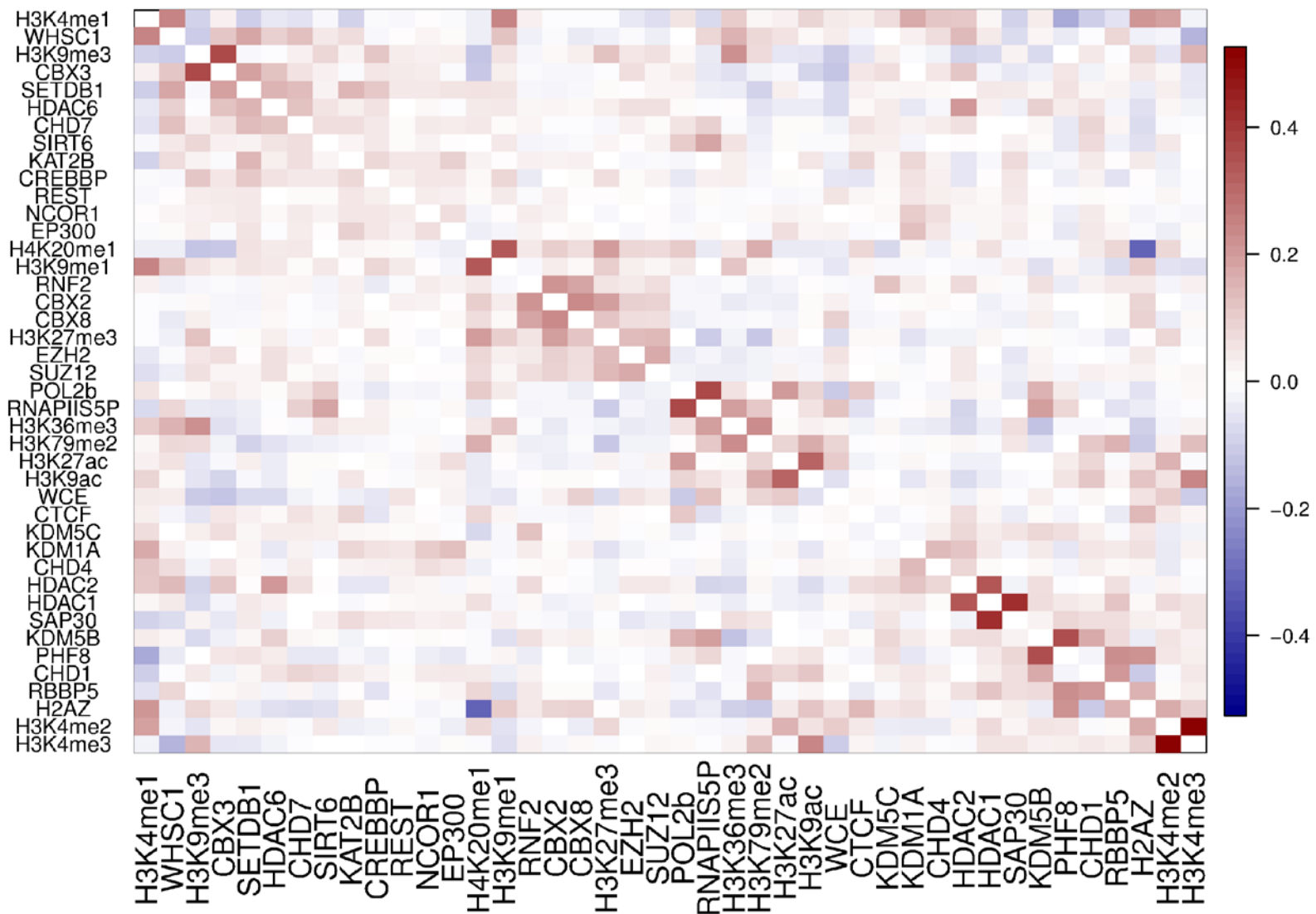
Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells.

Oren Ram, Alon Goren, Ido Amit, Noam Shores, Nir Yosef, Jason Ernst, Manolis Kellis, Melissa Gymrek, Robbyn Issner, Michael Coyne, Timothy Durham, Xiaolan Zhang, Julie Donaghey, Charles B. Epstein, Aviv Regev, Bradley E. Bernstein
Cell, Vol. 147, No. 7. (23 December 2011), pp. 1628-1639

Correlations among Histone Modifications plus Chromatin Modifiers

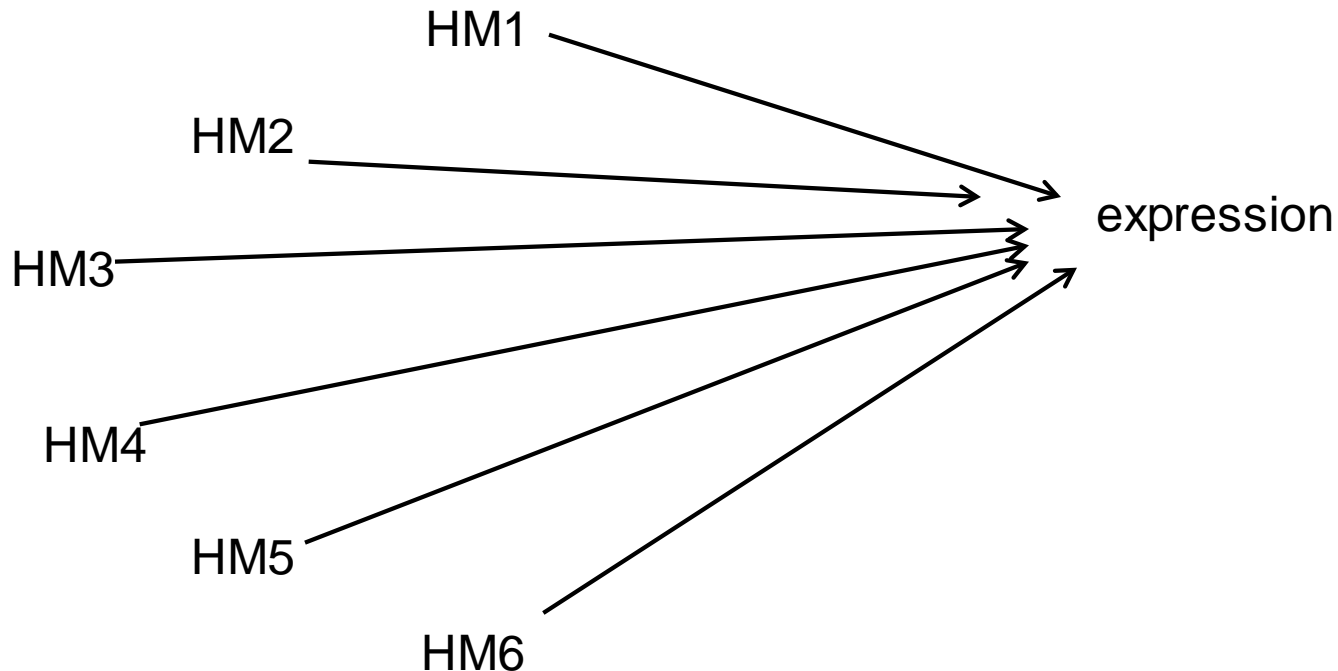


Partial Correlations among Histone Modifications plus Chromatin Modifiers



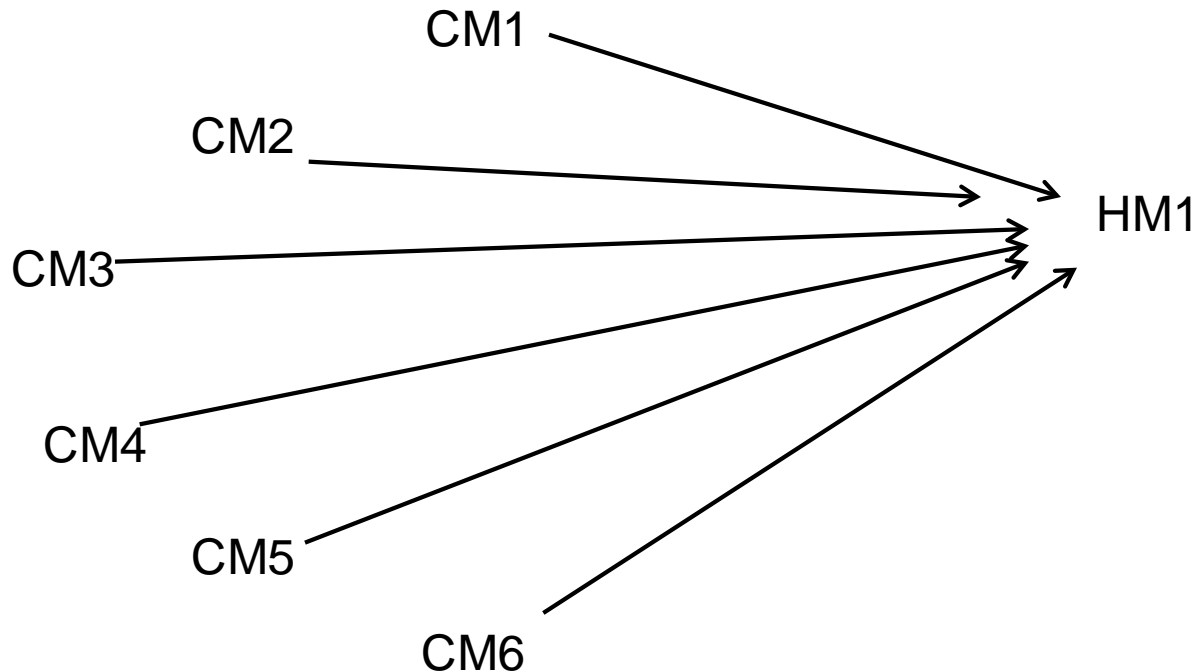
More network construction ...

- We modeled expression from HMs, with subsequent feature selection:



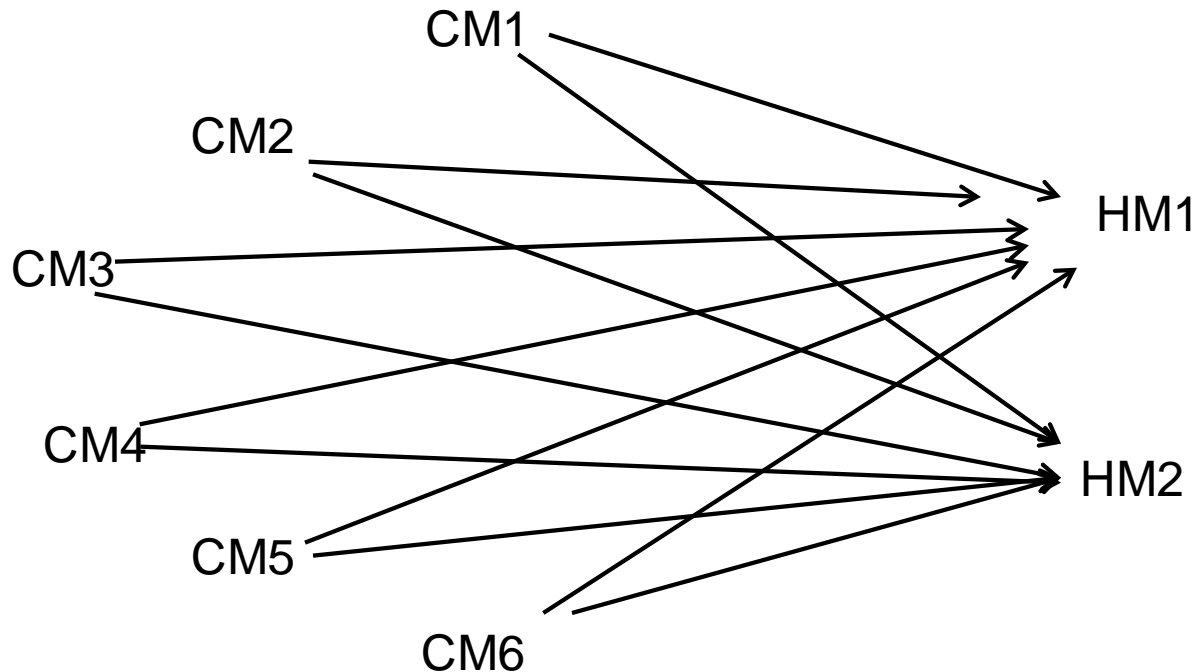
More network construction: HM-> CM

- Why not model a HMs from CMs?



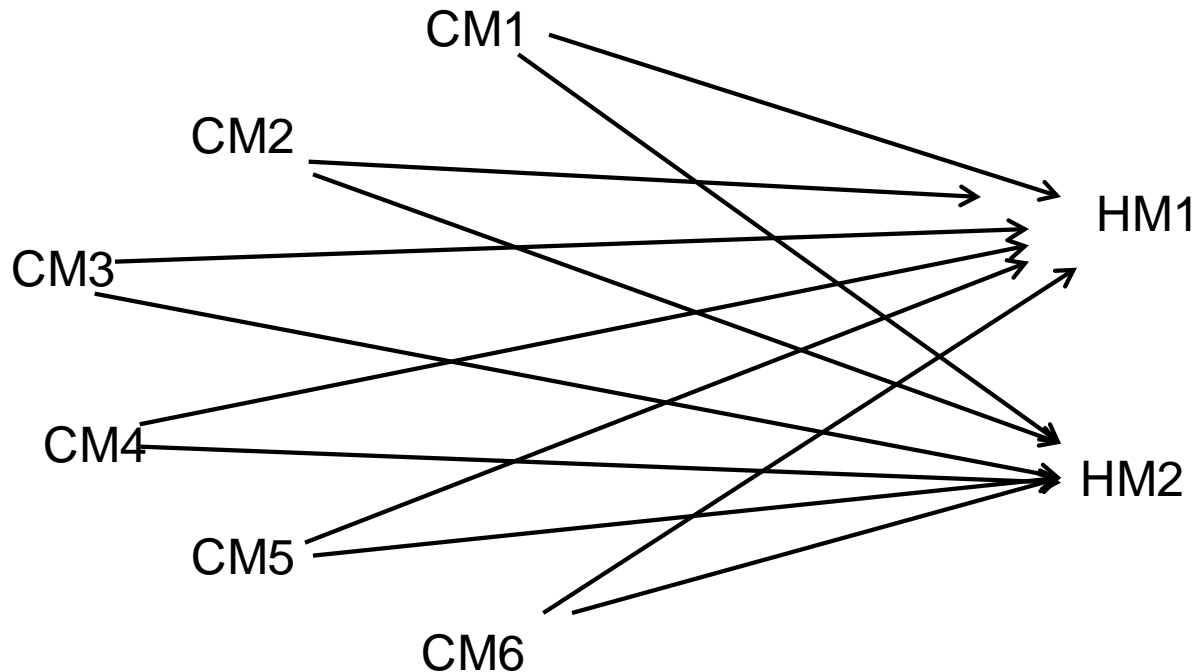
More network construction: HM-> CM

- Why not model **all** HMs from CMs?

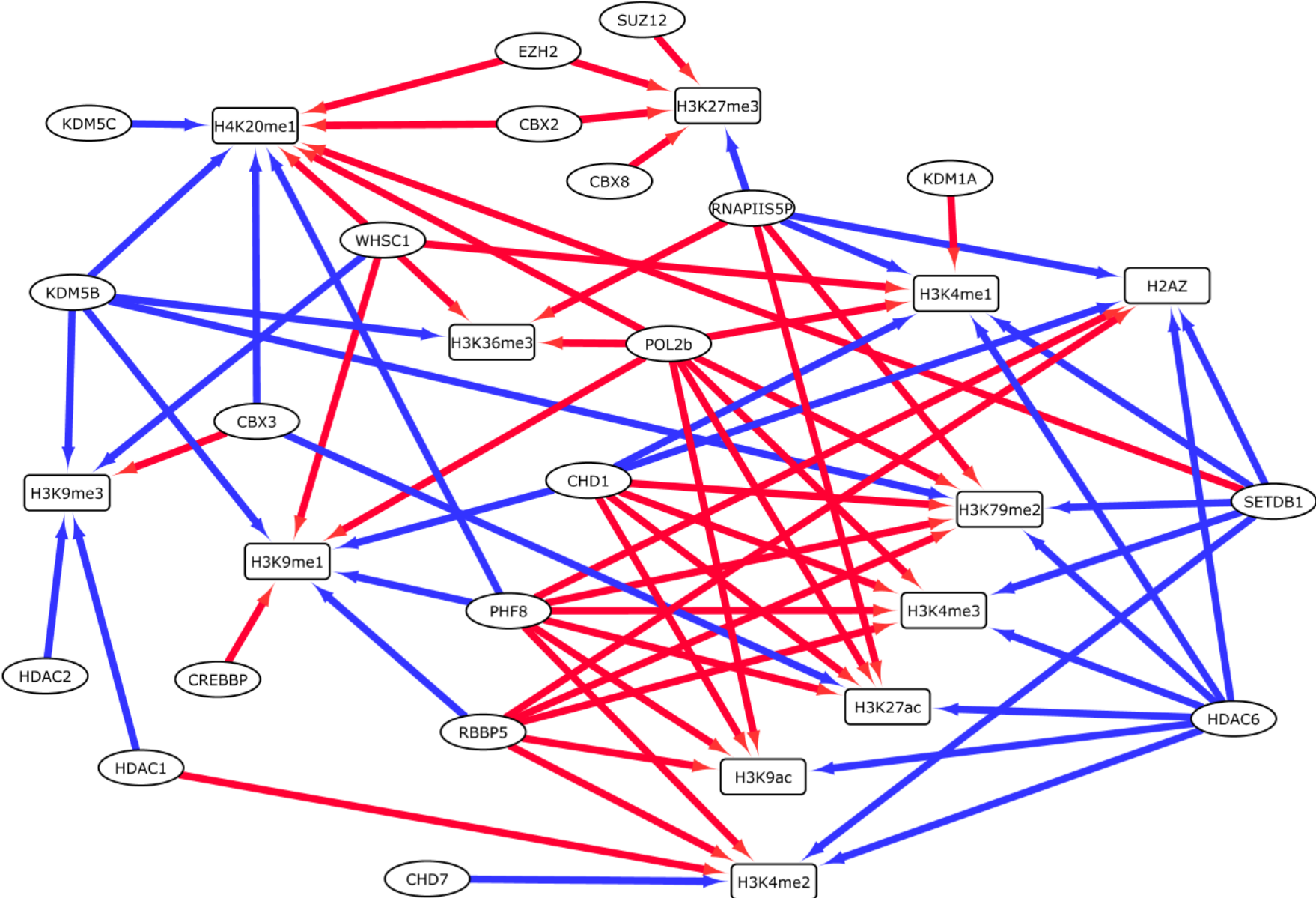


More network construction: sparse linear regression, elastic net

- Sparse regression replaces feature selection

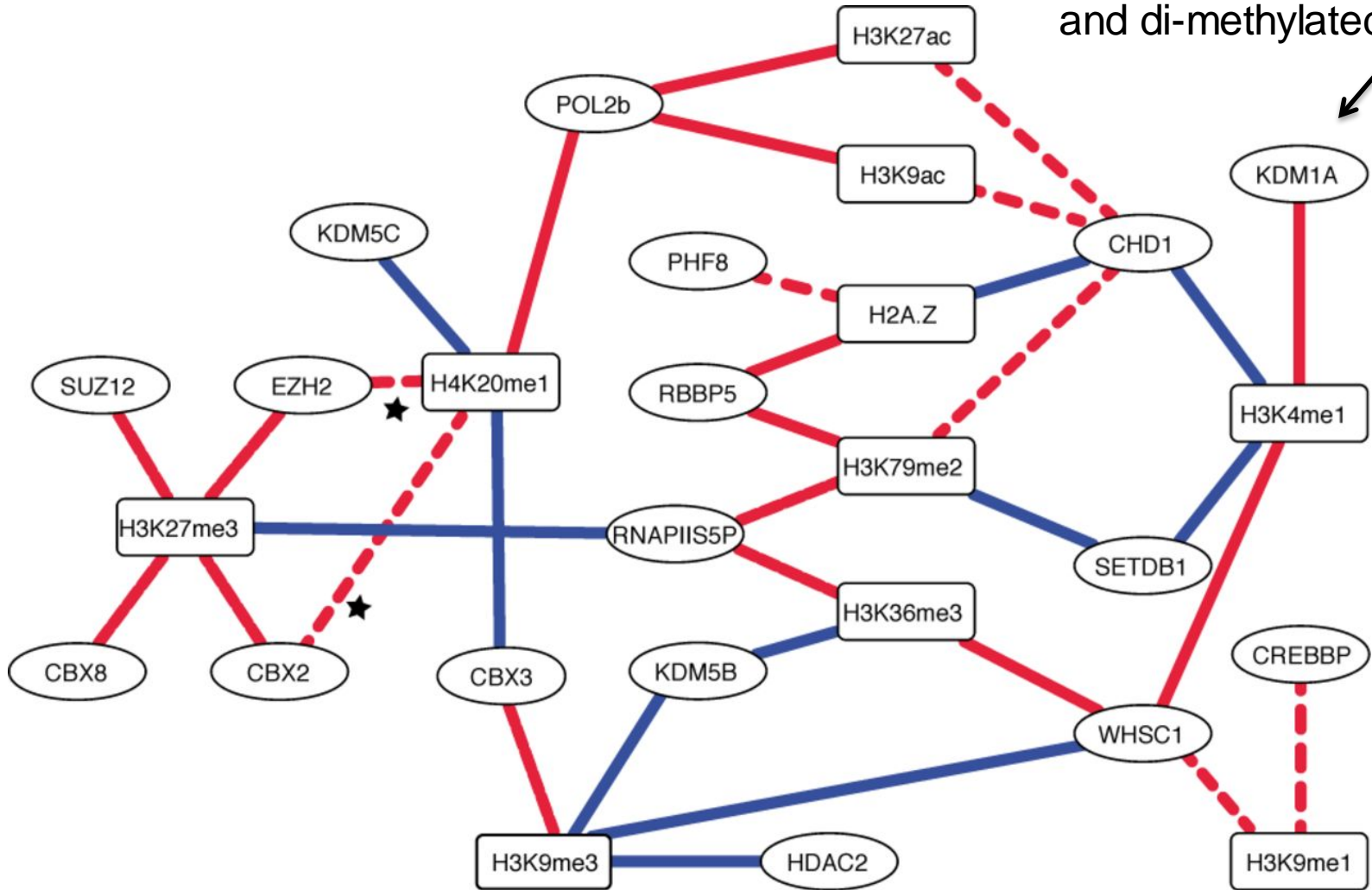


Sparse linear model (elastic net) explaining Histone Modifications from Chromatin Modifiers



Chromatin-signalling network.

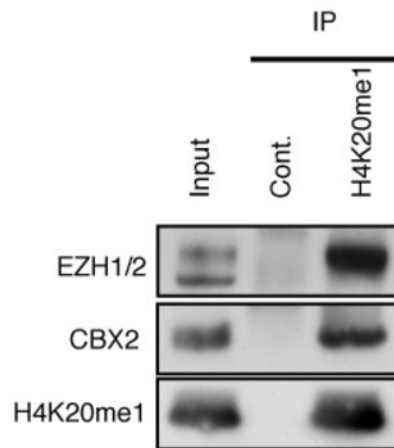
aka LSD1=lysine specific histone demethylase, demethylates mono- and di-methylated H3K4



Perner J et al. Nucl. Acids Res. 2014;42:13689-13695

Verification of two predicted interactions links H4K20me1 to Polycomb-mediated repression.

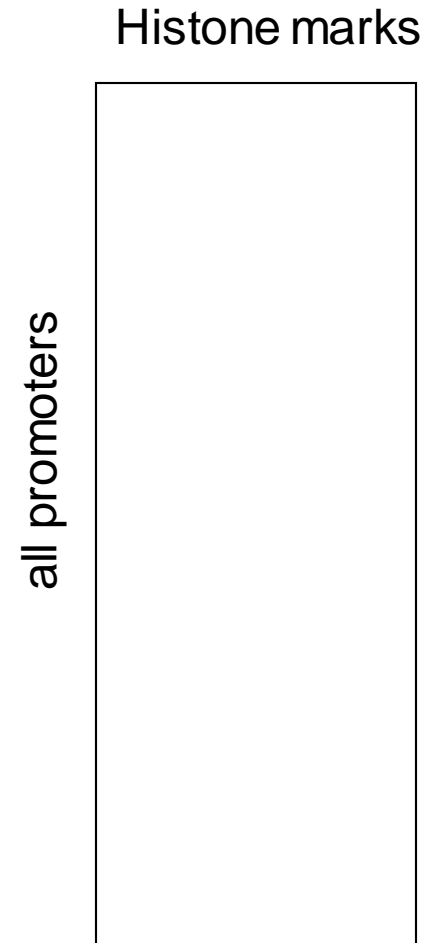
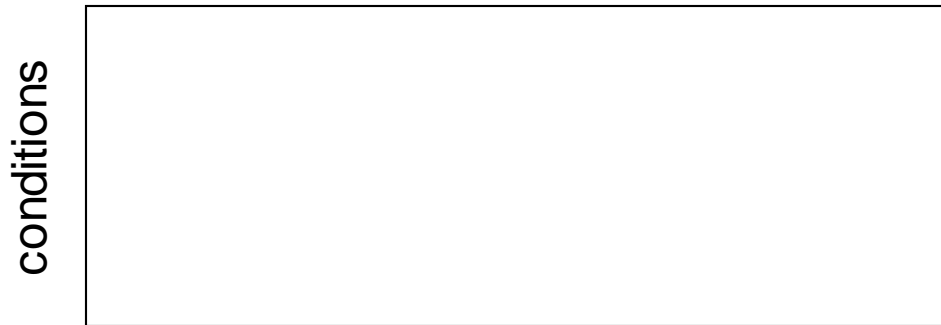
A



Perner J et al. Nucl. Acids Res. 2014;42:13689-13695

#nodes >> #conditions?

Not with histone modifications!



Acknowledgements

Ho-Ryun Chung, Rosa Karlic ... Linear models, HMs
Julia Lasserre Gaussian Graphical Models
Juliane Perner ... HM-CM networks
Sarah Kinkley ... validation experiments

