

# *Bayesian Network of Targeting Interactions in Chromatin: Reanalyzed*

Ricky Lim

`Ricky.lim@student.uva.nl`

Universiteit van Amsterdam  
Faculty of Science  
Biomedical Sciences

June 26, 2012

Interdisciplinary Seminar on Biological Networks

# Outline

## 1 Introduction

Chromatin:Revisited

Targeting Interactions in Chromatin

Bayesian Network(BN):Revisited

# Outline

## 1 Introduction

Chromatin:Revisited

Targeting Interactions in Chromatin

Bayesian Network(BN):Revisited

## 2 Main Results

# Outline

## 1 Introduction

Chromatin:Revisited

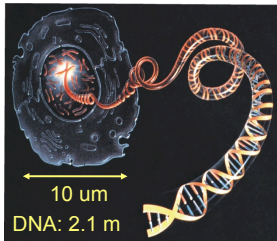
Targeting Interactions in Chromatin

Bayesian Network(BN):Revisited

## 2 Main Results

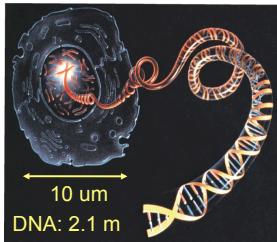
## 3 Conclusion

# Chromatin:Revisited

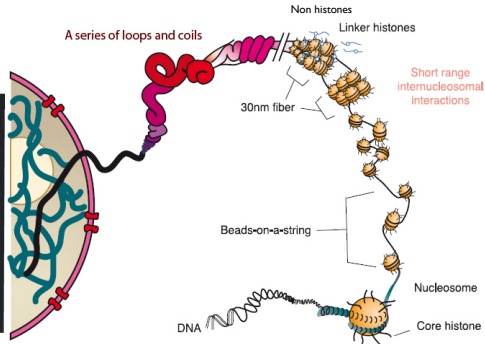


W, de laet, 2010  
(Lecture)

# Chromatin:Revisited

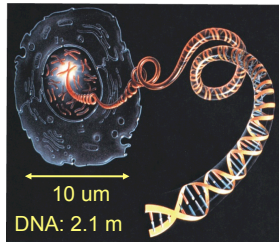


W, de laat, 2010  
(Lecture)

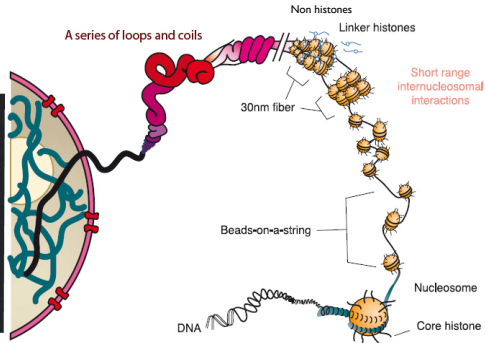


Horn 2002

# Chromatin:Revisited



W, de laet, 2010  
(Lecture)



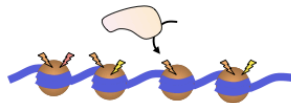
Horn 2002

**Chromatin** is a complex consisting of DNA, histones, and non-histones proteins(Chromatin Components)

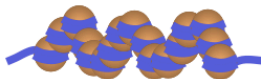
# Main Players in Chromatin



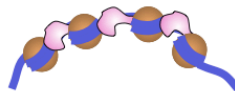
Remodelling complexes:  
repositioning



histone modifiers:  
covalent modifications



Heterochromatin proteins:  
packaging

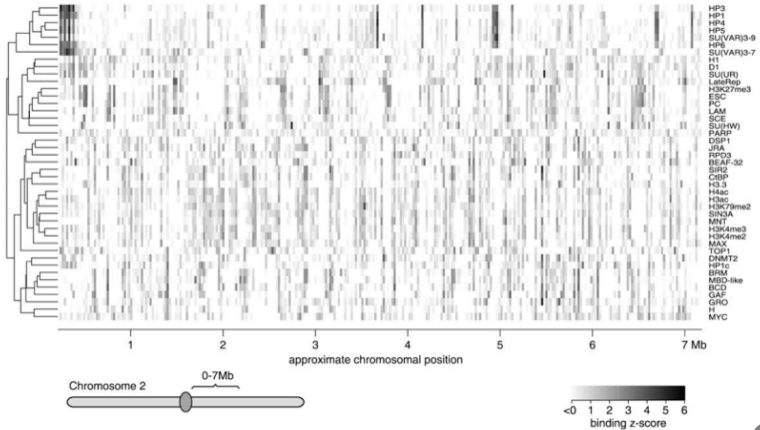


linker histones,  
HMG proteins:  
bending/folding/spacing

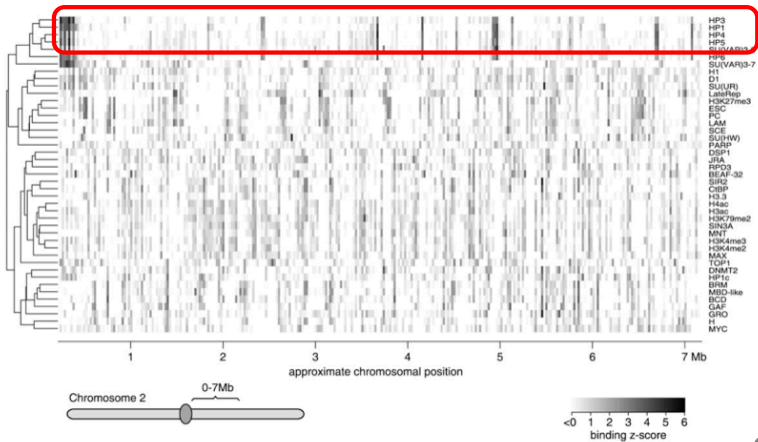


# Where Are These Players located ?

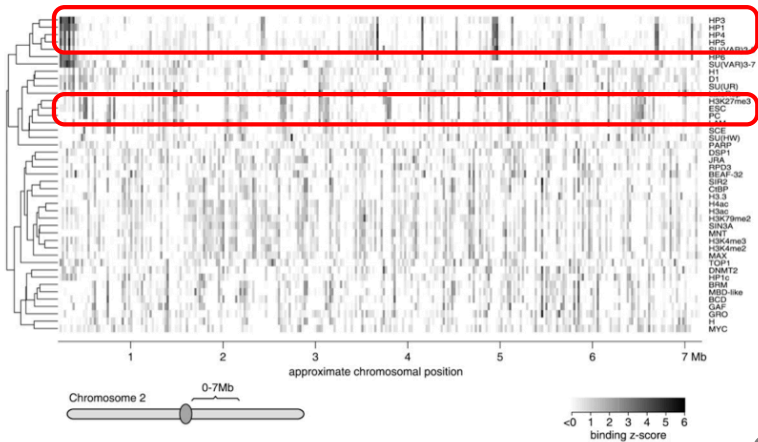
# Where Are These Players located ?



# Where Are These Players located ?



# Where Are These Players located ?



DamID

[illegible]

## Genome-wide specific binding patterns of components.

# What determines the specific locations ?

## **Hypothesis:**

By the interactions among the chromatin components

# What determines the specific locations ?

## **Hypothesis:**

By the interactions among the chromatin components



Network of Targeting Interactions

# What determines the specific locations ?

## Hypothesis:

By the interactions among the chromatin components



Network of Targeting Interactions



1. Comparing the Binding Maps
2. Systematic Search using Bayesian Network Analysis



# Bayesian Network(BN):Revisited

## Background:

- a model of dependencies between multiple components of a system

Peer, 2005

# Bayesian Network(BN):Revisited

## Background:

- a model of dependencies between multiple components of a system
- **Assumption: Most components are (conditionally) independent of each other**

Peer, 2005

# Bayesian Network(BN):Revisited

## Background:

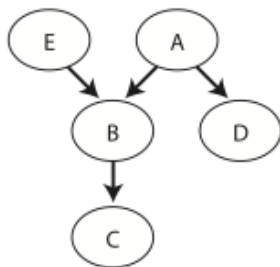
- a model of dependencies between multiple components of a system
- **Assumption: Most components are (conditionally) independent of each other**

Peer, 2005

# Bayesian Network(BN):Revisited

## Background:

- a model of dependencies between multiple components of a system
- **Assumption: Most components are (conditionally) independent of each other**



**Node** : Observed variables

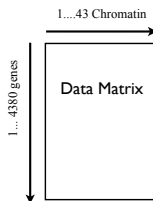
**Edge** : dependencies between the nodes

Peer, 2005

# Learn BN of Targeting Interactions from Scratch

## Problem Statement:

- Given a dataset  $D = \{\mathbf{x}^1, \dots, \mathbf{x}^{4380}\}$  of independent instances of  $\chi$  assuming that each genomic location (i.e, gene) as an independent sample.
- Find a network  $B = \langle G, \theta \rangle$  that best matches  $D$ .



# Learn BN of Targeting Interactions from Scratch

The bootstrapping procedure (adapted from Friedman,2000)

- For  $i = 1, \dots, 1000$ ,

# Learn BN of Targeting Interactions from Scratch

The bootstrapping procedure (adapted from Friedman,2000)

- For  $i = 1, \dots, 1000$ ,
- Construct a slightly perturbed input data,  $D_i$ , generating 1000 input data, each input data matrix has a dimension of 4380 rows each selected by random sampling with replacement from  $D$  (original dataset) and 43 columns (chromatin components)

# Learn BN of Targeting Interactions from Scratch

The bootstrapping procedure (adapted from Friedman,2000)

- For  $i = 1, \dots, 1000$ ,
- Construct a slightly perturbed input data,  $D_i$ , generating 1000 input data, each input data matrix has a dimension of 4380 rows each selected by random sampling with replacement from  $D$  (original dataset) and 43 columns (chromatin components)
- Apply the learning procedure using Banjo search on  $D_i$ , generating  $(G_1, \dots, G_{1000})$  in total of 1000 networks.



# Learn BN of Targeting Interactions from Scratch

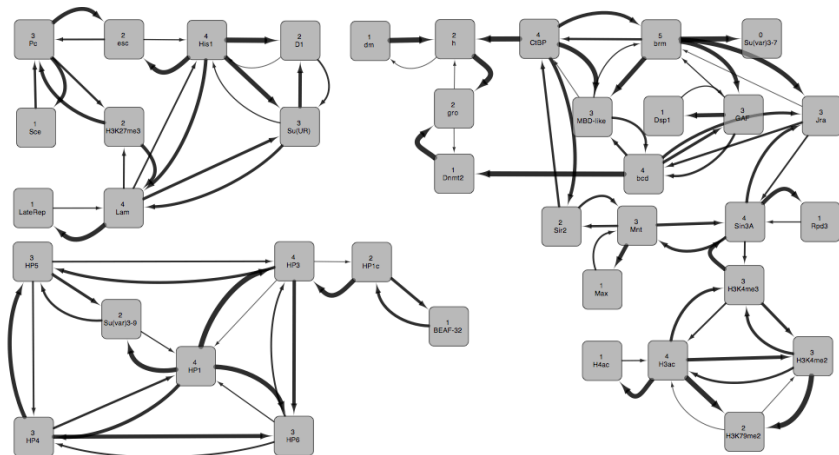
The bootstrapping procedure (adapted from Friedman,2000)

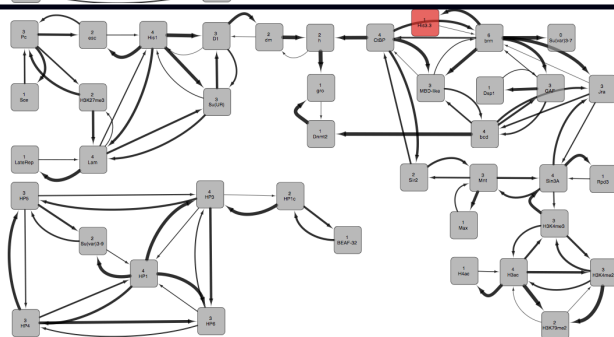
- For  $i = 1, \dots, 1000$ ,
- Construct a slightly perturbed input data,  $D_i$ , generating 1000 input data, each input data matrix has a dimension of 4380 rows each selected by random sampling with replacement from  $D$  (original dataset) and 43 columns (chromatin components)
- Apply the learning procedure using Banjo search on  $D_i$ , generating  $(G_1, \dots, G_{1000})$  in total of 1000 networks.
- For each feature of interest compute:

$$\text{Conf}(f) = \frac{1}{m} \sum_{i=1}^m f(G_i)$$

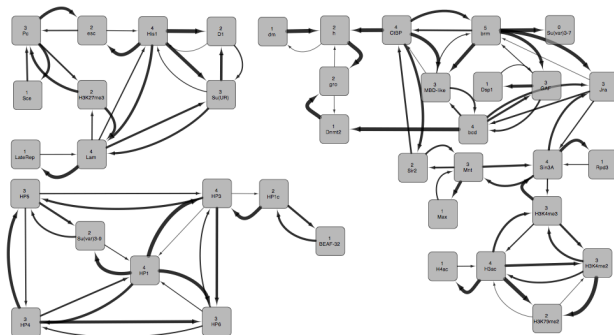
$$f(G) = \left\{ \begin{array}{ll} 1 & \text{if } f \text{ is a feature in } G \\ 0 & \text{if } f \text{ is absent} \end{array} \right\}$$

# Reconstructed BN80

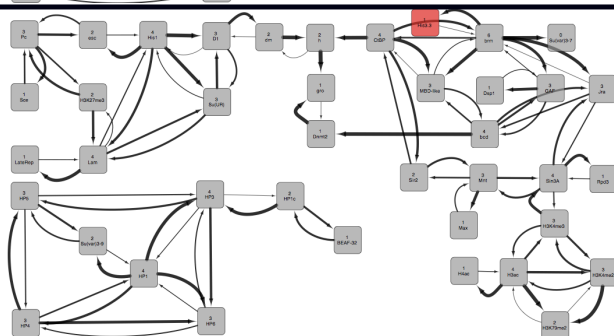




van Steensel et al. 2010

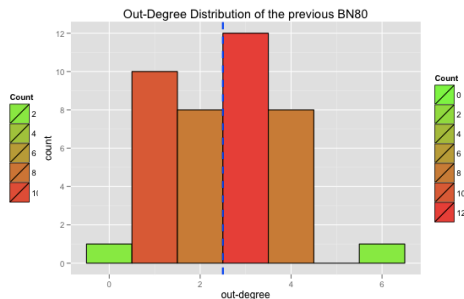
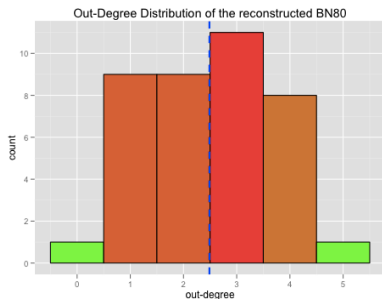


50 pairs out of 52 previously reported pairs are recovered. Thus the recovery rate is 96.15%



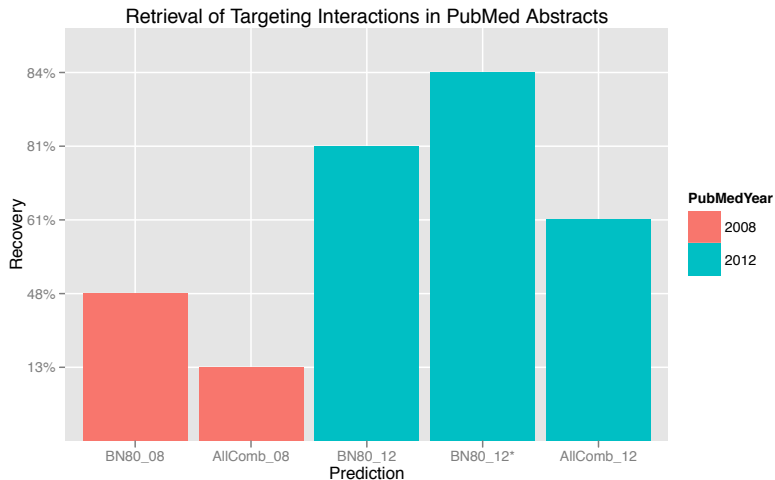
**Top:** Reconstructed BN80  
**Below:** Previous BN80

# BN80 Topology



----- : average out-degree

# PubMed Validation



# Conclusion

- 1 Successfully reconstructed BN80 with the recovery rate of 96.15%
- 2 The reconstructed BN80 shares similar network topology with the average out-degree of 2.5
- 3 84% of the predicted interactions from BN80 recovered back in PubMed abstracts

# Conclusion

- 1 Successfully reconstructed BN80 with the recovery rate of 96.15%
- 2 The reconstructed BN80 shares similar network topology with the average out-degree of 2.5
- 3 84% of the predicted interactions from BN80 recovered back in PubMed abstracts



# Conclusion

- 1 Successfully reconstructed BN80 with the recovery rate of 96.15%
- 2 The reconstructed BN80 shares similar network topology with the average out-degree of 2.5
- 3 84% of the predicted interactions from BN80 recovered back in PubMed abstracts

# Pitfalls and Future

- Learning the causal pattern from BN is assumed that no hidden variables(i.e no Hidden Common Cause)
- Static BN, No feedback loops hence the temporal features are desired for the future to model dynamic BN

# Pitfalls and Future

- Learning the causal pattern from BN is assumed that no hidden variables(i.e no Hidden Common Cause)
- Static BN, No feedback loops hence the temporal features are desired for the future to model dynamic BN

# Acknowledgment

- Gunnar W. Klau
- Alexander Schoenhuth
- Guillaume Fillion
- Tycho Bismeijer

# References I

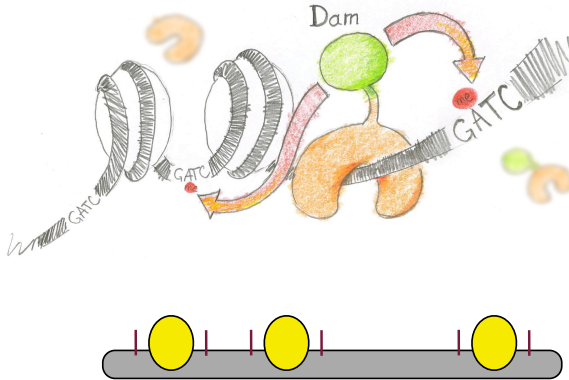
1. van Steensel B, Braunschweig U, Filion GJ, Chen M, van Bemmelen JG, et al. (2010) Bayesian network analysis of targeting interactions in chromatin. *Genome Research* 20: 190200. doi:10.1101/gr.098822.109.
2. Horn PJ (2002) MOLECULAR BIOLOGY: Chromatin Higher Order Folding—Wrapping up Transcription. *Science* 297: 18241827. doi:10.1126/science.1074200.
3. Rando OJ, Chang HY (2009) Genome-wide views of chromatin structure. *Annu Rev Biochem* 78: 245271. doi:10.1146/annurev.biochem.78.071107.134639.
4. Pe'er DD (2005) Bayesian network analysis of signaling networks: a primer. *CORD Conference Proceedings 2005*: pl4pl4.
5. Friedman N, Linial M, Nachman I, Pe'er D (2000) Using Bayesian networks to analyze expression data. *J Comput Biol* 7: 601620. doi:10.1089/106652700750050961.

## References II

6. Heckerman D (n.d.) A tutorial on learning with bayesian networks. researchmicrosoftcom.  
Available:<http://research.microsoft.com/pubs/69588/tr-95-06.pdf>.  
Accessed 21 February 2012.
7. Oufir EA (n.d.) Suitability of Bayesian Networks for the Inference and Completion of Genetic Networks using Microarray Data. Voorbraak F, Visser A, editors Amsterdam: science.uva.nl. pp.
8. Vesterström J (n.d.) Heuristic Algorithms in Bioinformatics. daimi.au.dk. pp.
9. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, et al. (2007) Integration of biological networks and gene expression data using Cytoscape. Nat Protoc 2: 23662382.  
doi:10.1038/nprot.2007.324.

10. Chavan SS, Bauer MA, Scutari M, Nagarajan R (2009)  
NATbox: a network analysis toolbox in R. BMC Bioinformatics 10  
Suppl 11: S14. doi:10.1186/1471-2105-10-S11-S14.

# DAM-ID: Mapping Method



Cartoon by Filion,G  
<http://en.wikipedia.org/wiki/DamID>



# DAM-ID: Mapping Method

**Experimental:**

**DAM FUSION PROTEIN**



*express in vivo*

isolate  
genomic DNA



PCR amplify  
methylated  
fragments



label  
fragments



Cy3

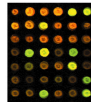


**Reference:**

**DAM ONLY**



Cy5



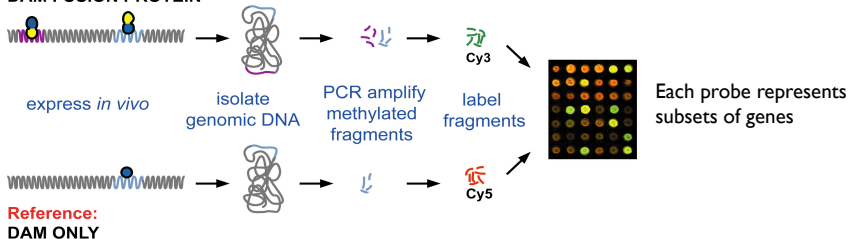
Each probe represents  
subsets of genes

van Steensel, et al., 2000

# DAM-ID: Mapping Method

## Experimental:

### DAM FUSION PROTEIN



van Steensel, et al., 2000

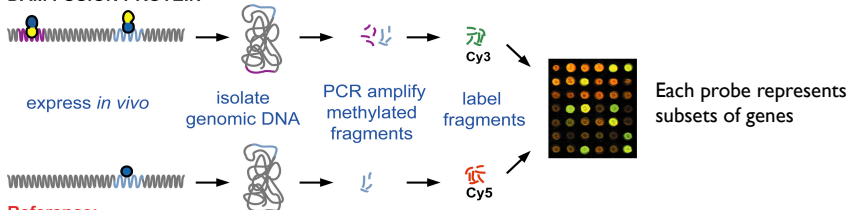
## Example of data matrix

	ID	HP1c
1	GH01001	-0.2725644
2	GH01039	-0.9612644
3	GH01043	-0.3777644
4	GH01053	-0.7831644
5	GH01059	-0.9795644

# DAM-ID: Mapping Method

## Experimental:

### DAM FUSION PROTEIN



## Reference:

### DAM ONLY

van Steensel, et al., 2000

## Example of data matrix Visualization of binding data

	ID	HP1c
1	GH01001	-0.2725644
2	GH01039	-0.9612644
3	GH01043	-0.3777644
4	GH01053	-0.7831644
5	GH01059	-0.9795644

