

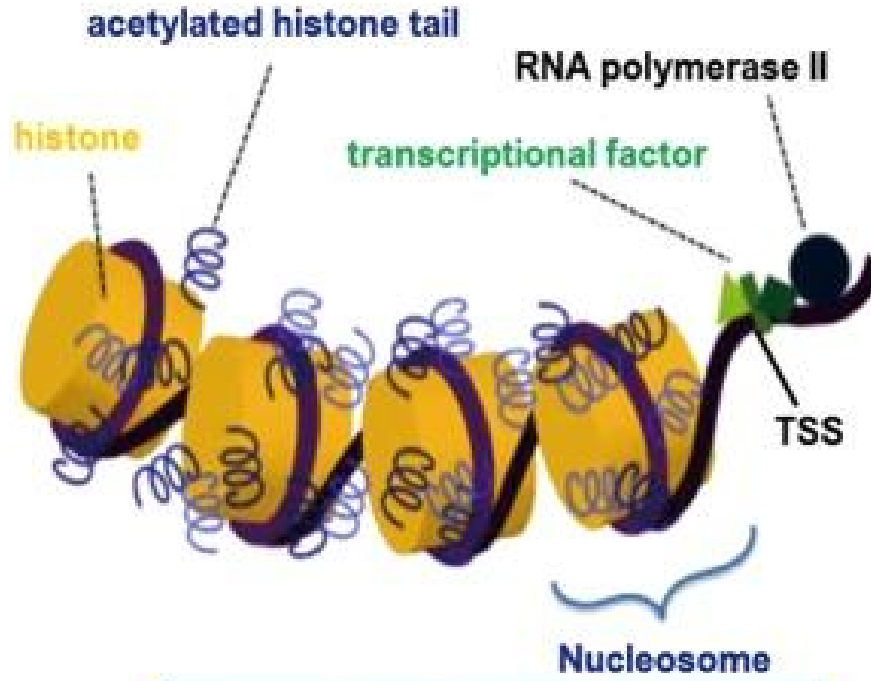
# Chromatin states at boundary elements

---

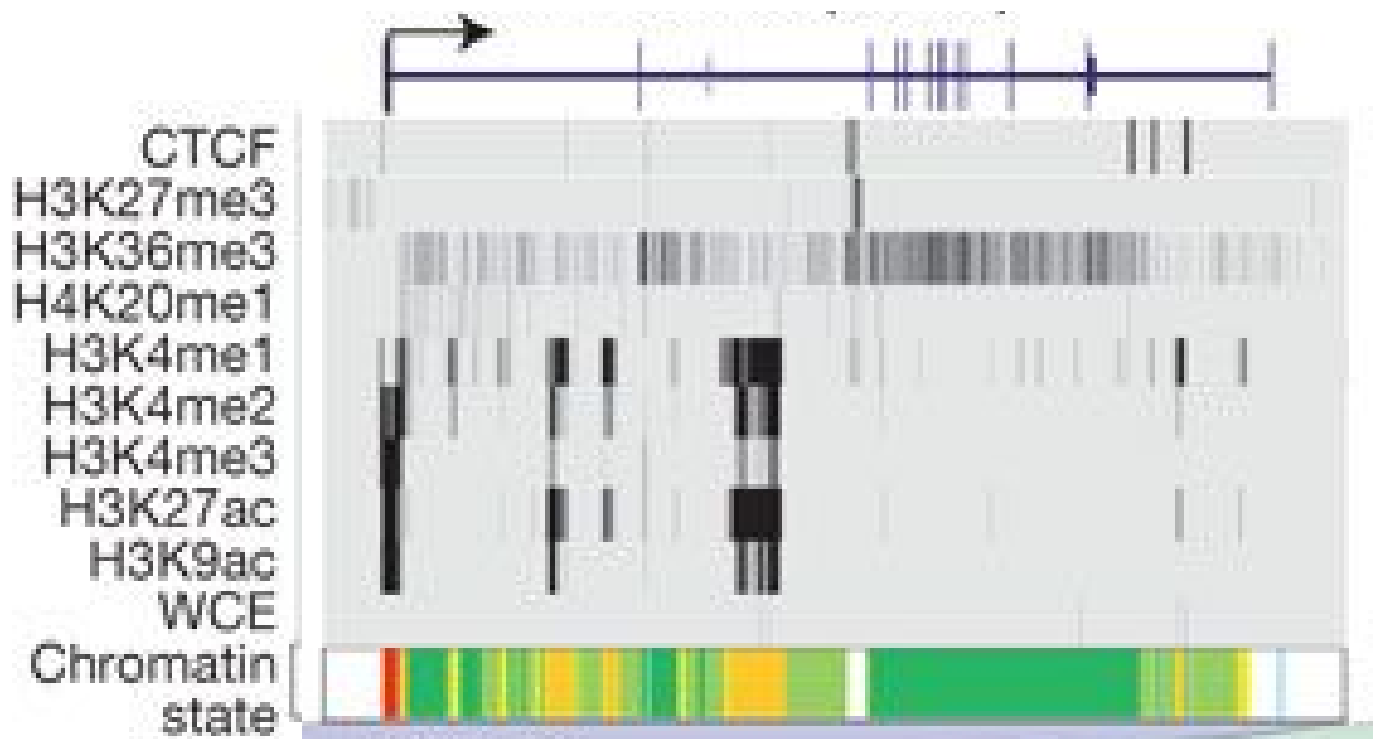
Betsy Pu

Background

# Histones & Modifications



# ChromHMM



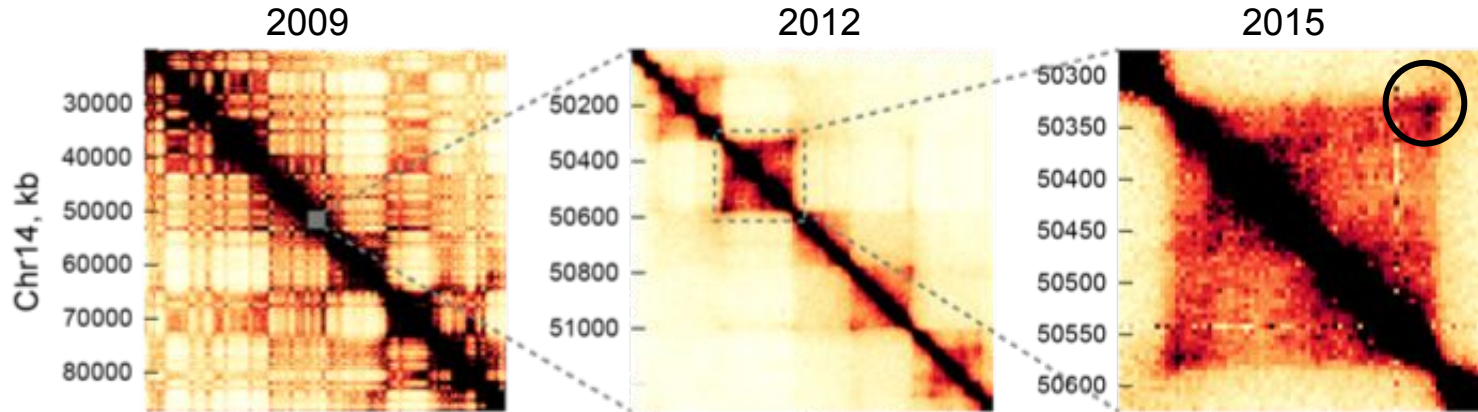
# Chromatin States



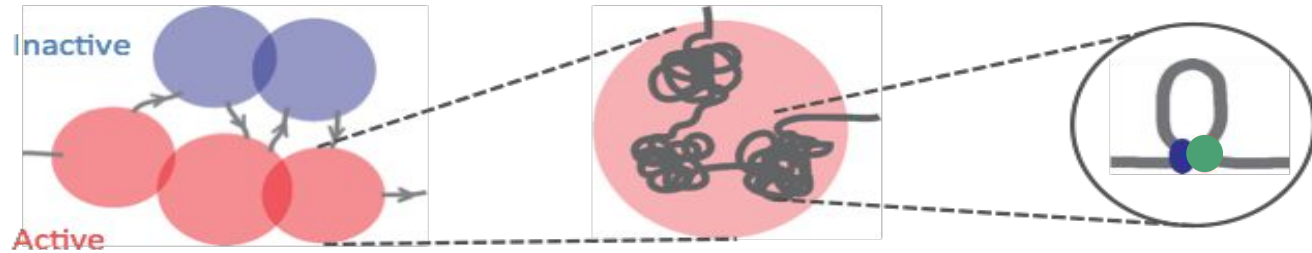
- |                            |                                     |
|----------------------------|-------------------------------------|
| 1 Active Promoter          | 8 Insulator (CTCF)                  |
| 2 Weak Promoter            | 9 Transcriptional transition        |
| 3 Inactive/poised Promoter | 10 Transcriptional elongation       |
| 4 Strong Enhancer          | 11 Weak transcribed                 |
| 5 Strong Enhancer          | 12 Polycomb-repressed               |
| 6 Weak/poised Enhancer     | 13 Heterochromatin; low signal      |
| 7 Weak/poised Enhancer     | 14 Repetitive/Copy Number Variation |
|                            | 15 Repetitive/Copy Number Variation |

# Methods and Results

# Loops and domains



A/B Compartments

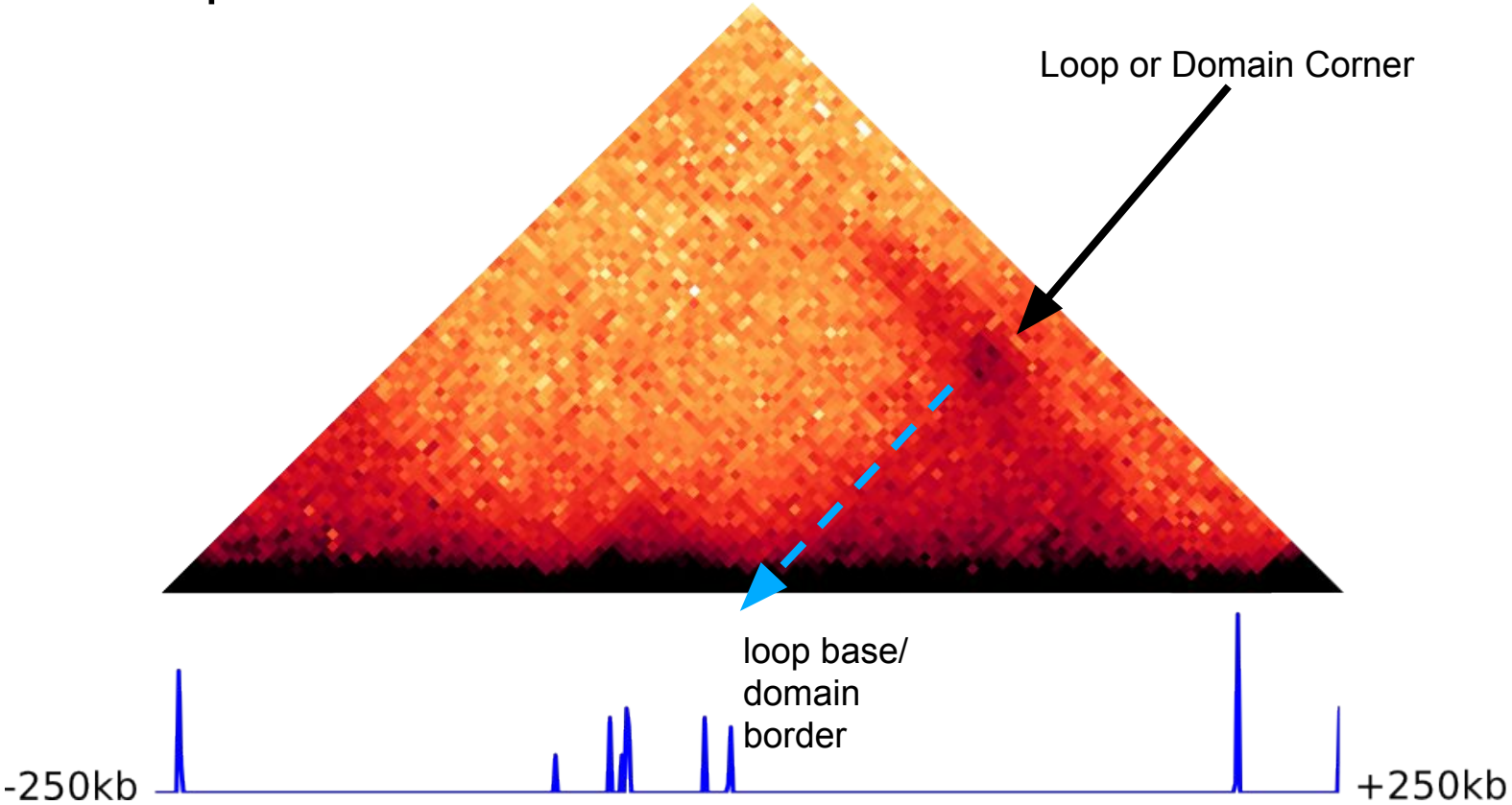


Lieberman Aiden, et al. Science 2009

Nora et al. Nature 2012

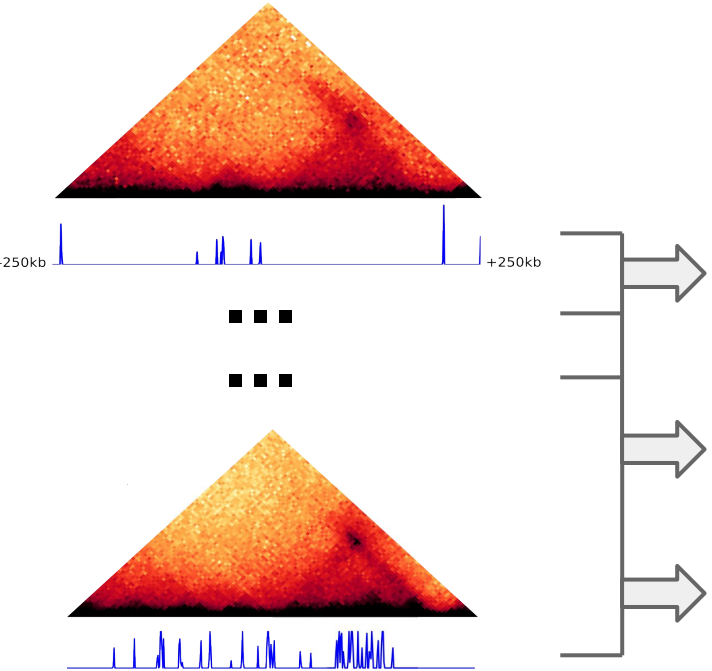
Rao et al. Cell 2014

# Method: aggregate analysis of genomic features around loops or domains

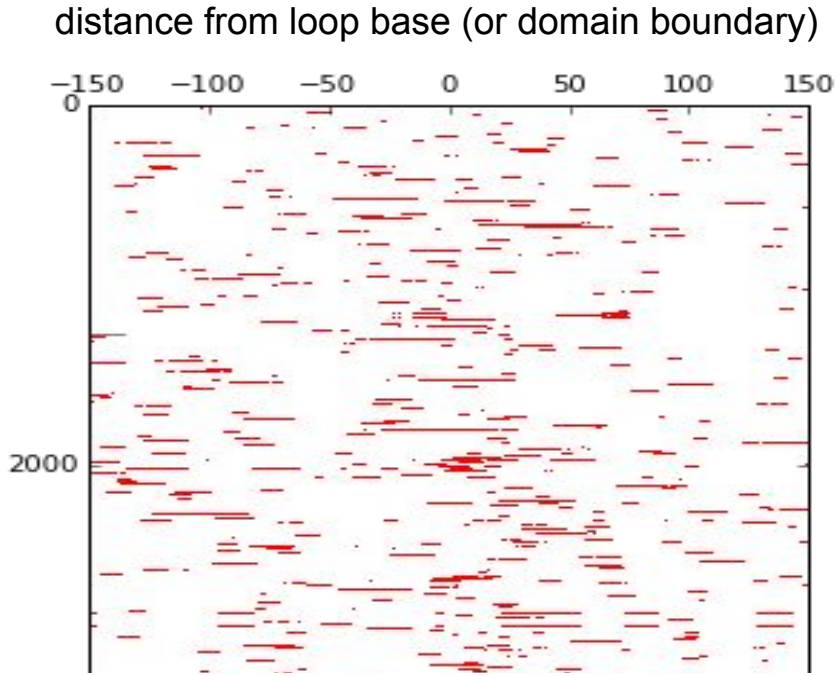




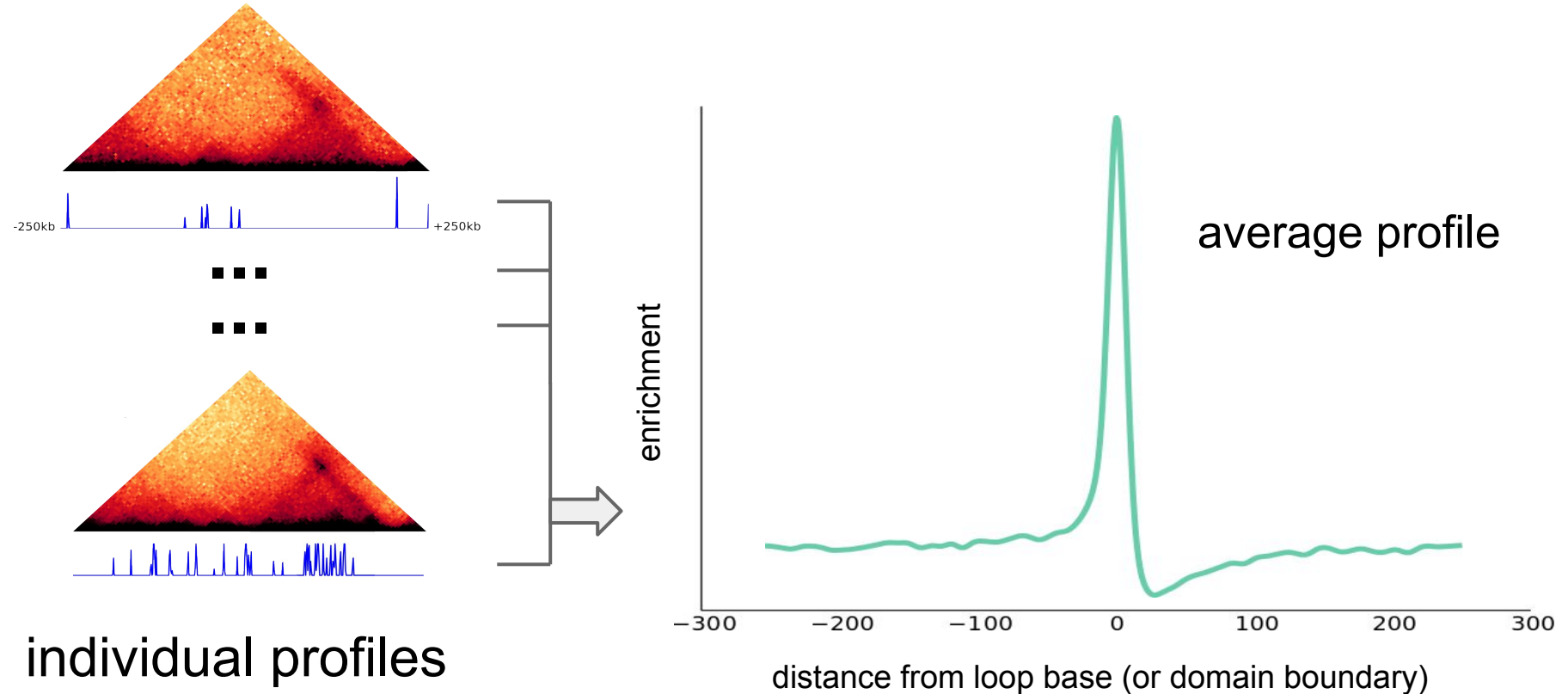
# Method: aggregate analysis of genomic features around loops or domains



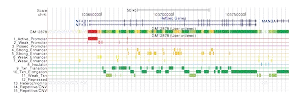
individual profiles



# Method: aggregate analysis of genomic features around loops or domains

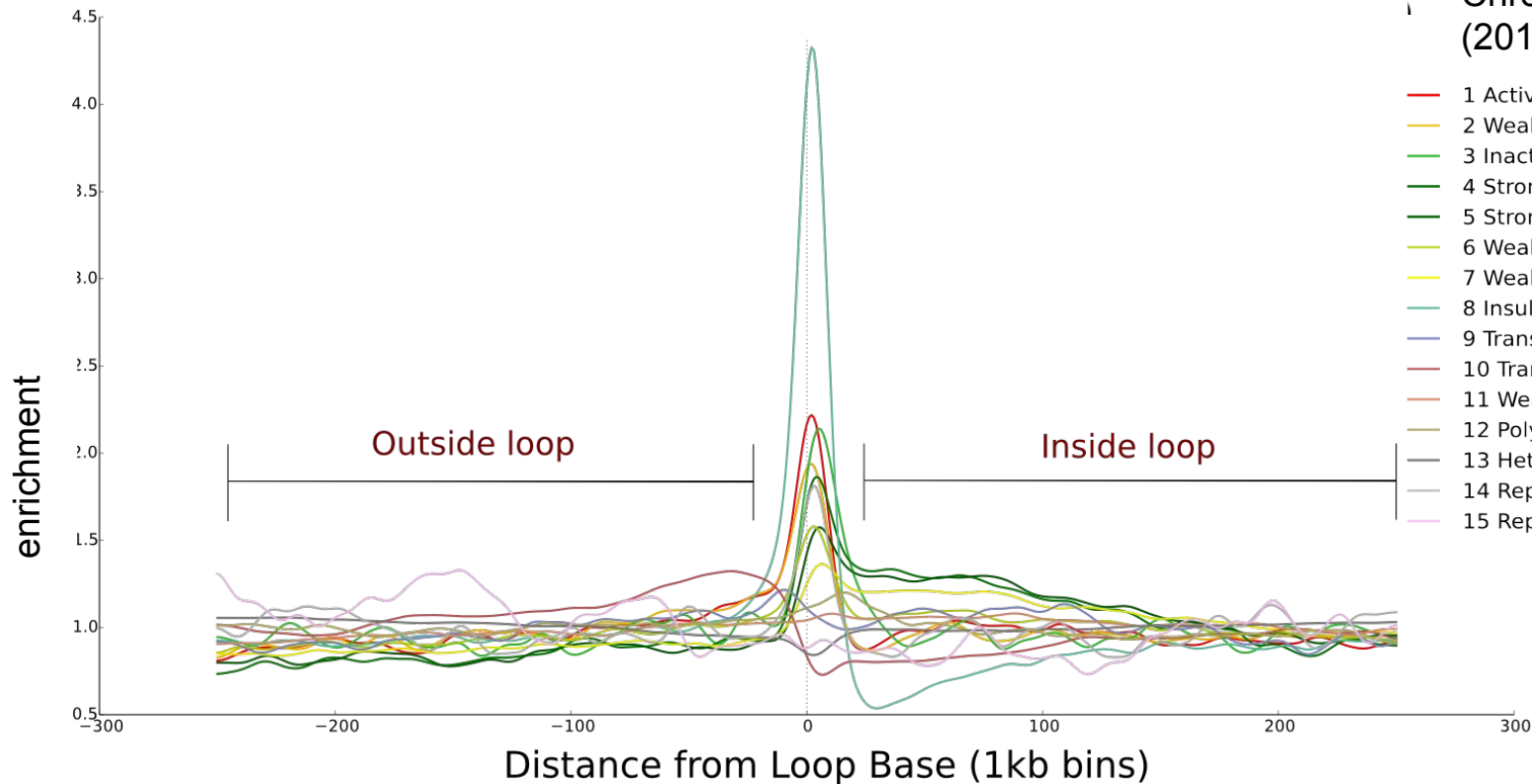


# Chromatin states: Diverse behaviors around loop bases

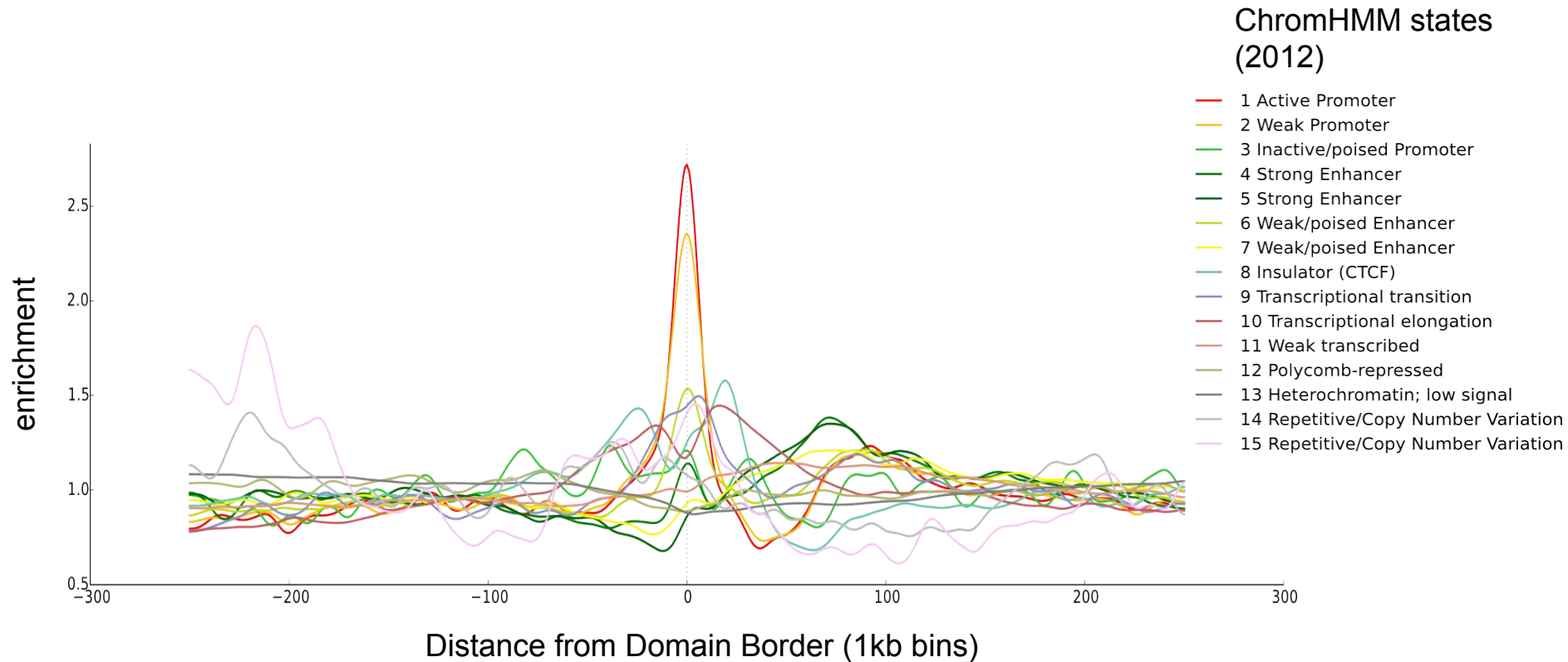
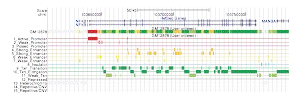


## ChromHMM states (2012)

- 1 Active Promoter
- 2 Weak Promoter
- 3 Inactive/poised Promoter
- 4 Strong Enhancer
- 5 Strong Enhancer
- 6 Weak/poised Enhancer
- 7 Weak/poised Enhancer
- 8 Insulator (CTCF)
- 9 Transcriptional transition
- 10 Transcriptional elongation
- 11 Weak transcribed
- 12 Polycomb-repressed
- 13 Heterochromatin; low signal
- 14 Repetitive/Copy Number Variation
- 15 Repetitive/Copy Number Variation

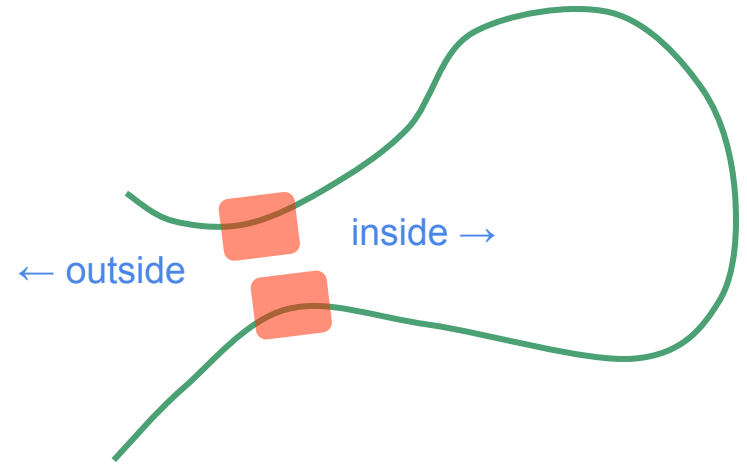
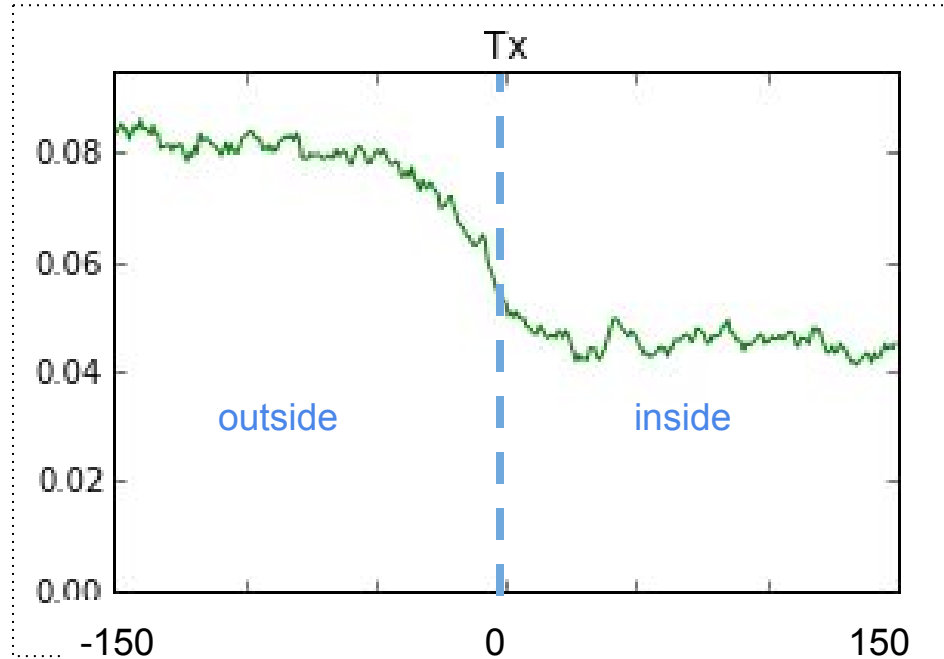


# Chromatin states: Diverse and different behaviors around domain boundaries

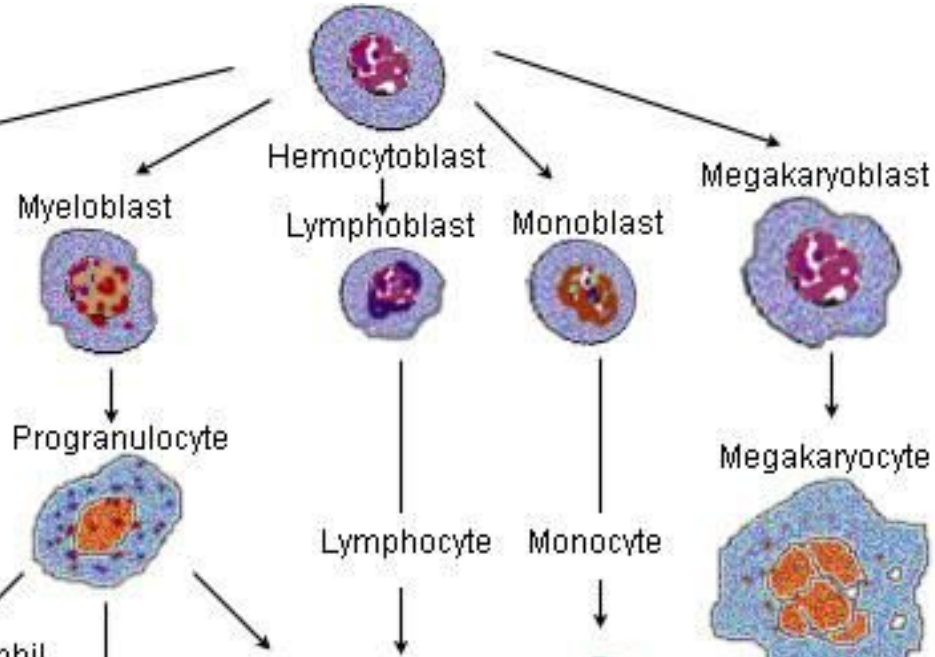


# Asymmetry of chromstates around loci

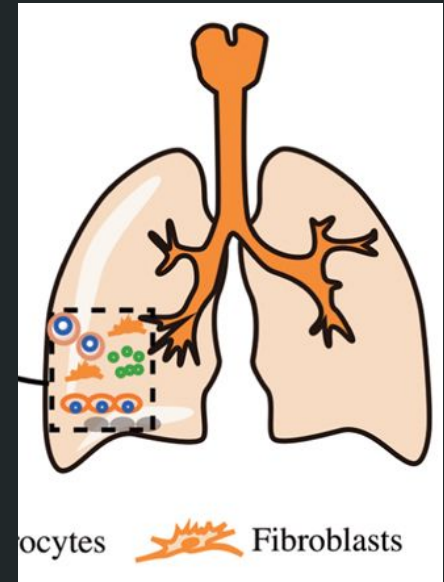
asymmetry of transcription elongation around loops



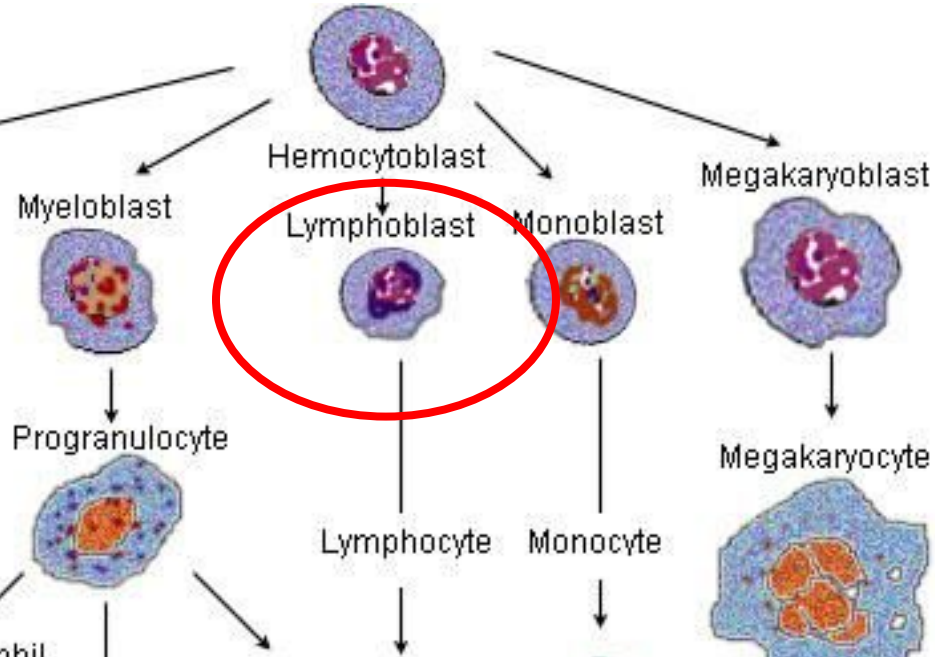
# Cross-cell-type analysis



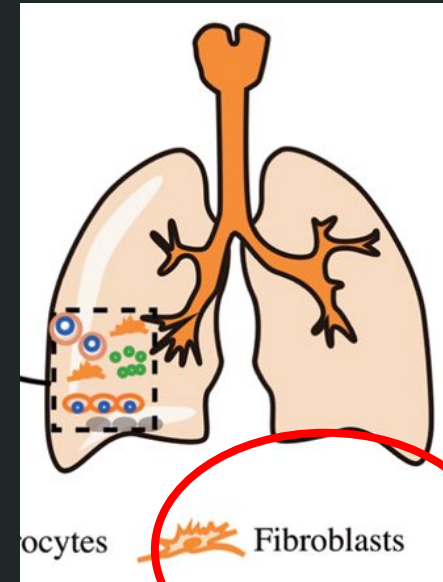
We compared chromatin state distributions between cell types GM12878 (lymphoblast cell line) and IMR90 (a lung fibroblast line)



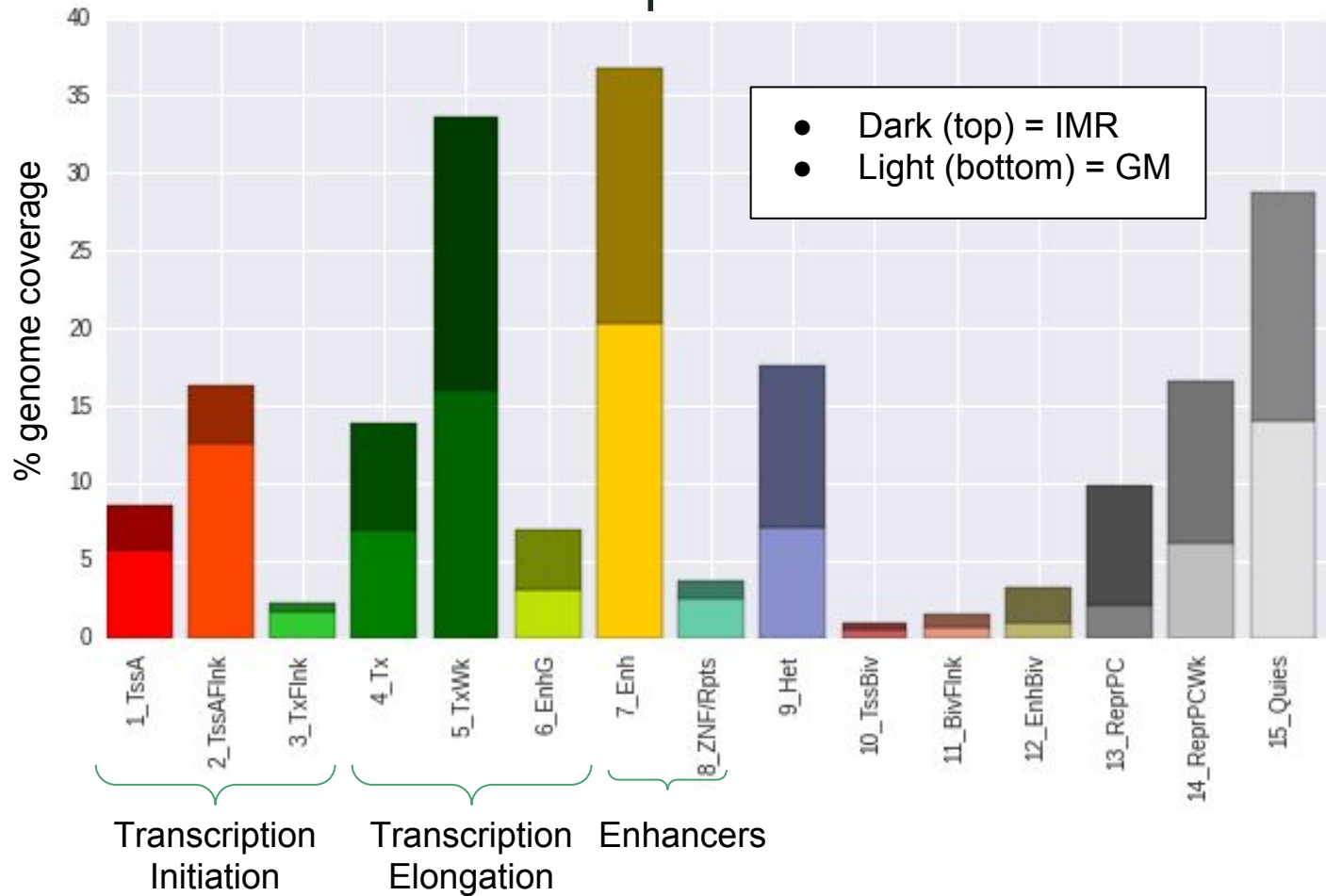
# Cross-cell-type analysis



We compared chromatin state distributions between cell types GM12878 (lymphoblast cell line) and IMR90 (a lung fibroblast line)

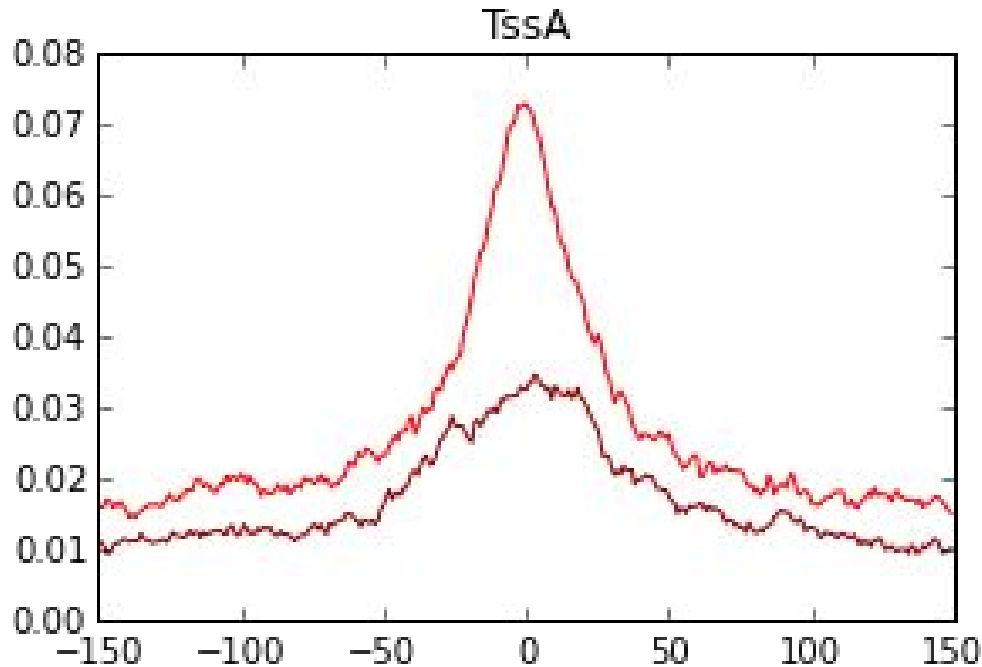


# GM vs IMR90 state composition



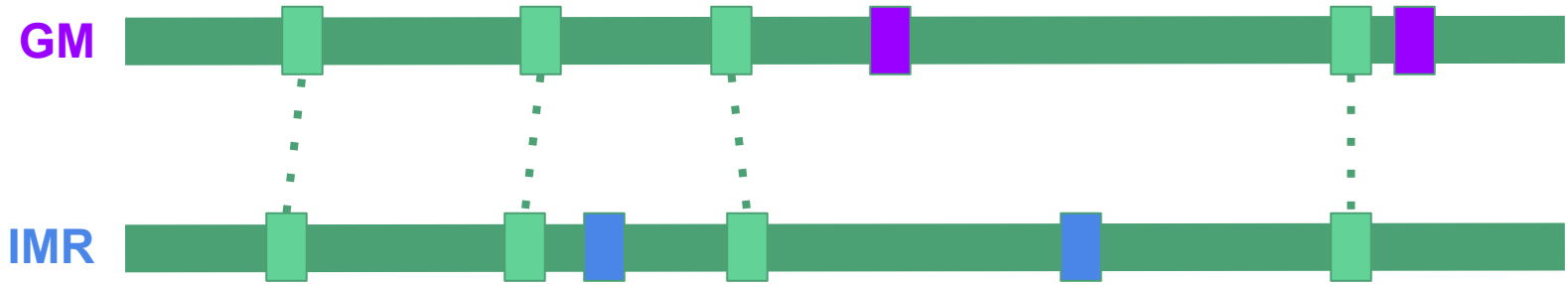


# Chromstate distributions differ between cell types



- GM domain boundaries on GM chromatin states
- IMR domain boundaries on IMR chromatin states

# Cell type specificity of boundary elements



A boundary alignment algorithm was used to pair up domain boundaries / loop bases between GM and IMR to find which are **shared** between the two cell types, and which are not shared (**cell-type specific** loci).

# Sequence Alignment (Edit Distance)

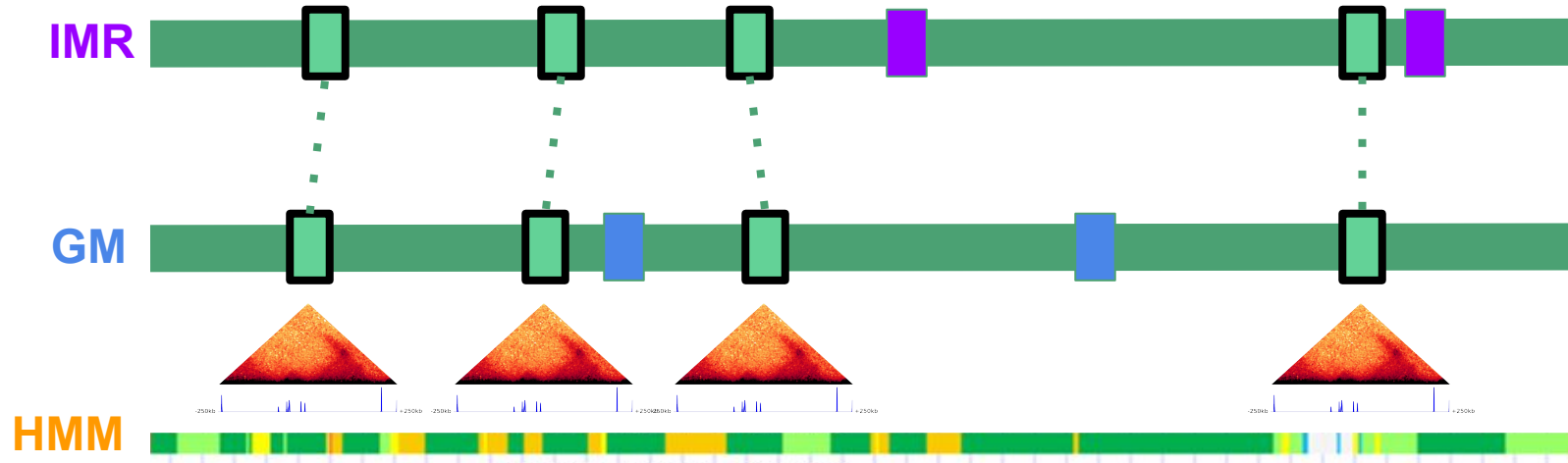
match = 1    mismatch = -1    gap = -1

		G	C	A	T	G	C	U
	0	-1	-2	-3	-4	-5	-6	-7
G	-1	1	0	-1	-2	-3	-4	-5
A	-2	0	0	1	0	-1	-2	-3
T	-3	-1	-1	0	2	1	0	-1
T	-4	-2	-2	-1	1	1	0	-1
A	-5	-3	-3	-1	0	0	0	-1
C	-6	-4	-2	-2	-1	-1	1	0
A	-7	-5	-3	-1	-2	-2	0	0

GCATG CU  
 G ATTACA

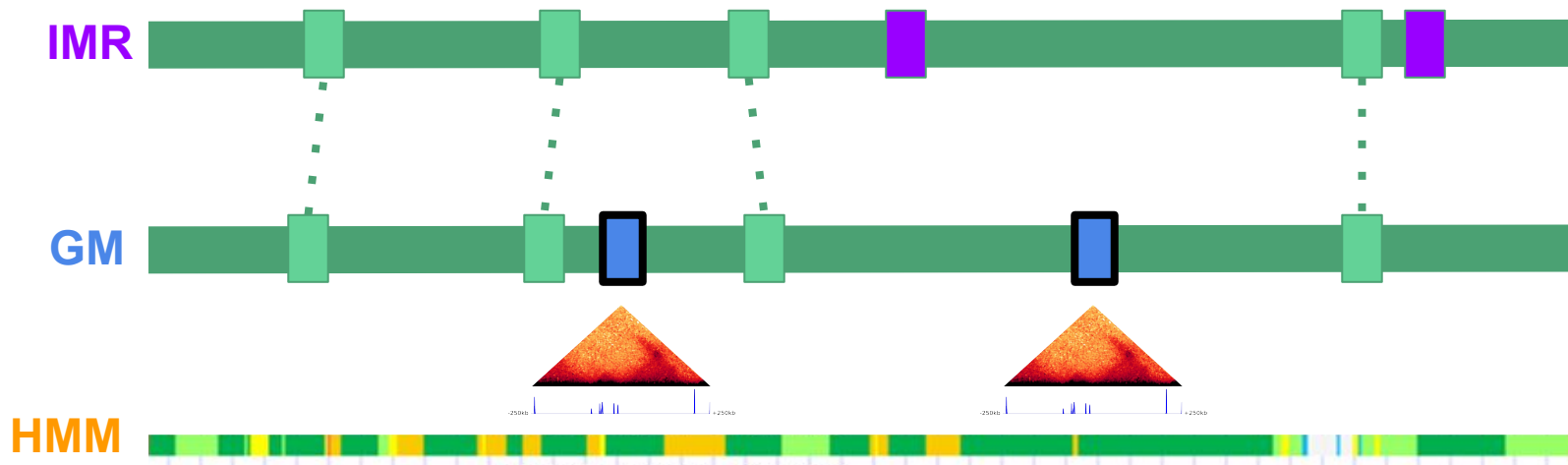
- Edit distance on boundary loci (domain boundaries, loop bases)
- mismatch cost = distance between loci
- gap cost = 50kb

# Shared loci



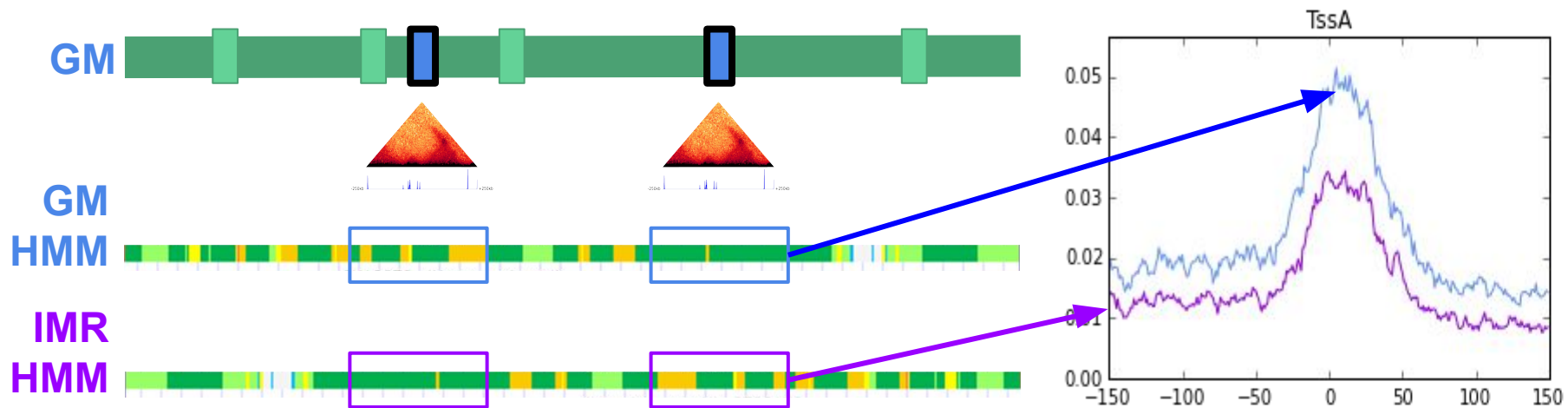
- H: Do cell-type specific loci reflect cell-type specific chromatin state profiles?
- P: We would expect profiles to differ more between cell types at specific loci than at conserved loci.

# Cell-type specific loci

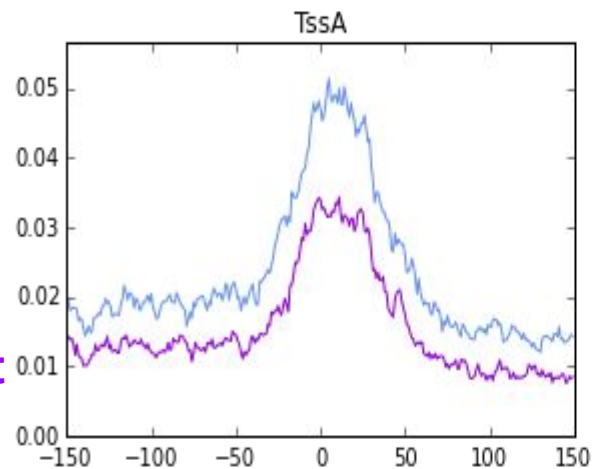
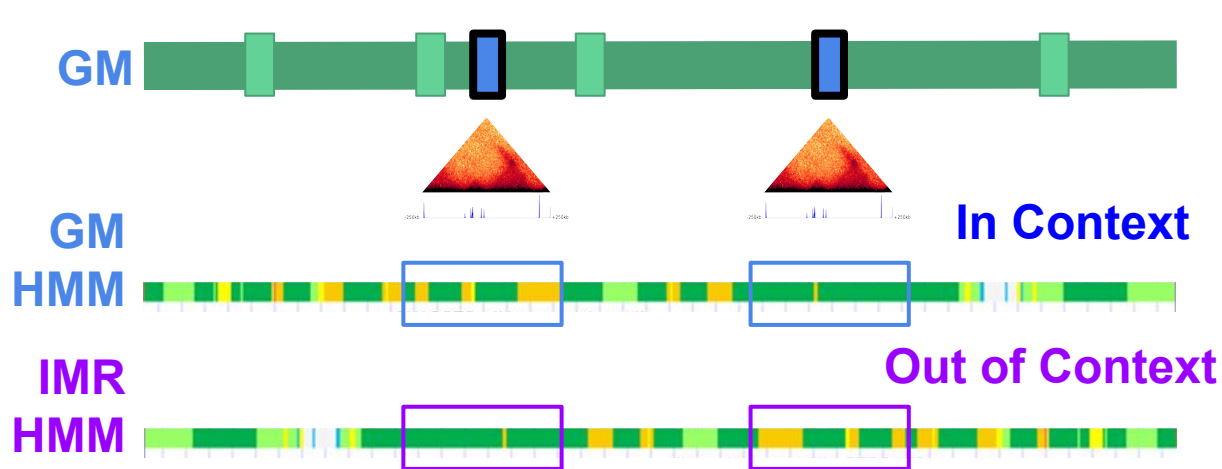


- H: Do cell-type specific loci reflect cell-type specific chromatin state profiles?
- P: We would expect profiles to differ more between cell types at specific loci than at conserved loci.

# Cell-type specific loci



# Cell-type specific loci

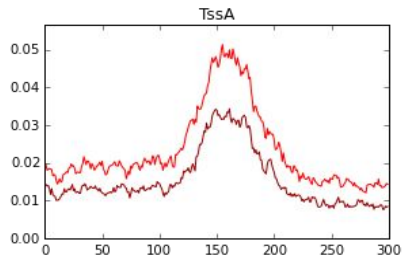


# Transcription Initiation States (Roadmap 1-3)

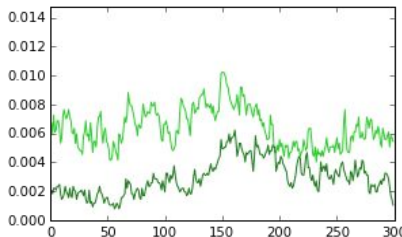
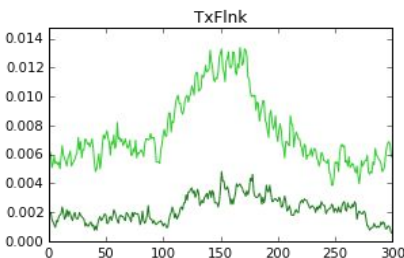
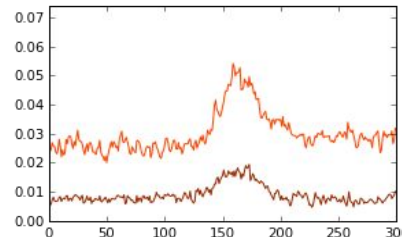
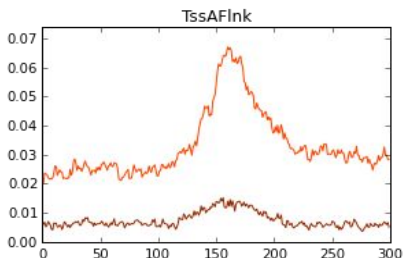
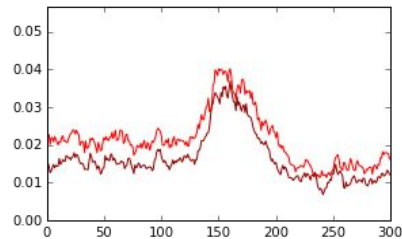
light: *in-context*

dark: *out-of-context*

GM-specific loop bases



Shared loop bases



Loci have cell context dependent profiles:

GM profiles tend to have higher frequencies of transcription initiation / active promoters than the IMR profiles.

GM-specific loop bases show more pronounced TSS activity in the GM context.

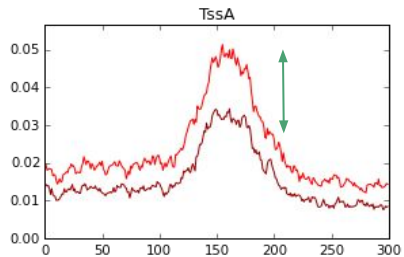


# Transcription Initiation States (Roadmap 1-3)

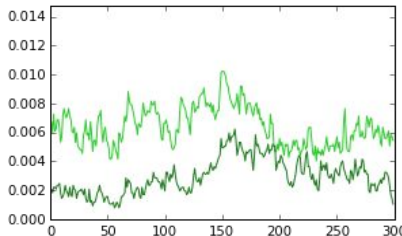
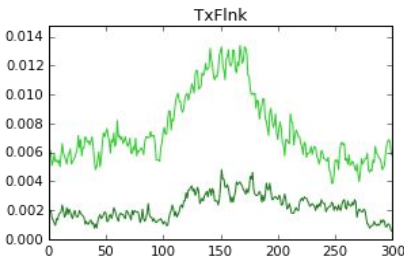
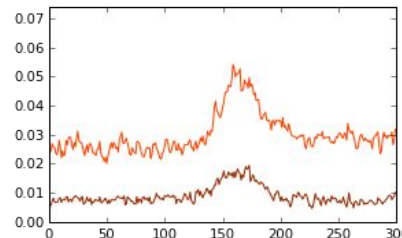
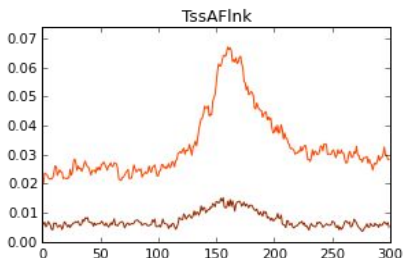
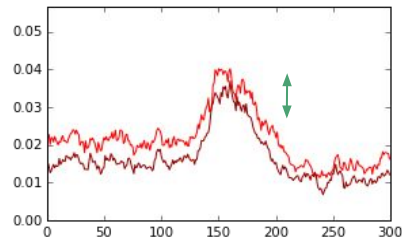
light: *in-context*

dark: *out-of-context*

GM-specific loop bases



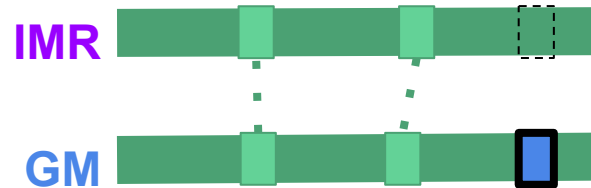
Shared loop bases



Loci have cell context dependent profiles:

GM profiles tend to have higher frequencies of transcription initiation / active promoters than the IMR profiles.

GM-specific loop bases show more pronounced TSS activity in the GM context.

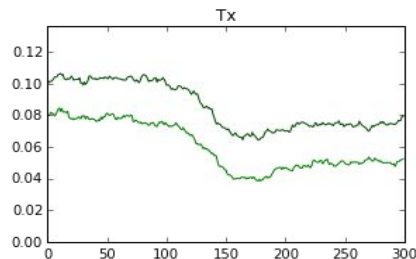


# Transcription Elongation States (Roadmap 4-6)

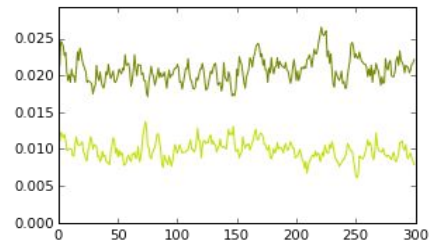
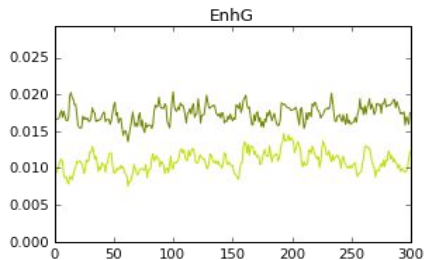
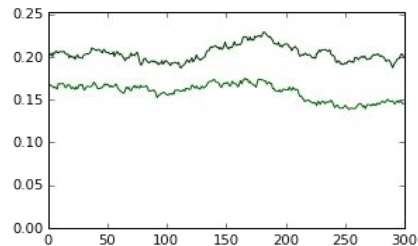
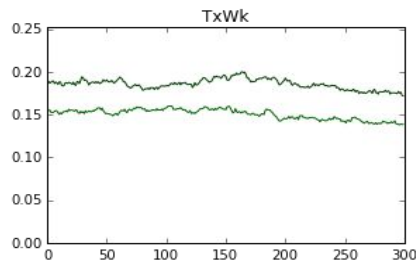
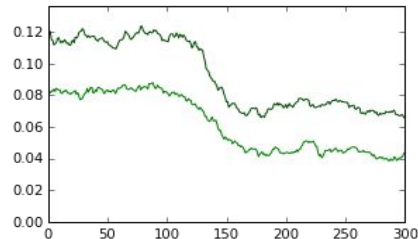
*light: in-context*

*dark: out-of-context*

GM-specific loop bases



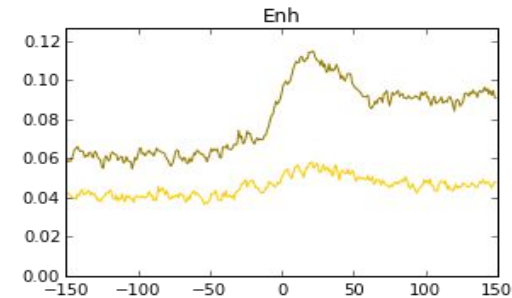
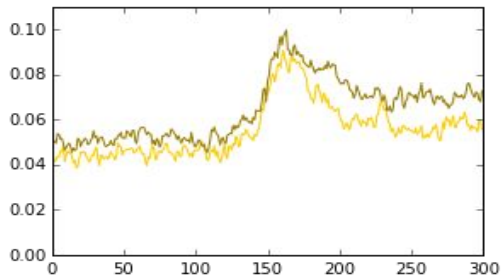
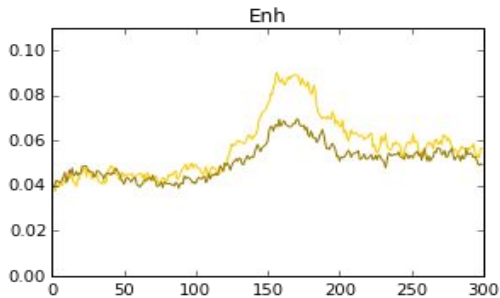
Shared loop bases



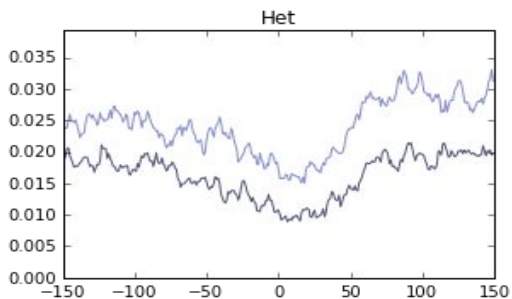
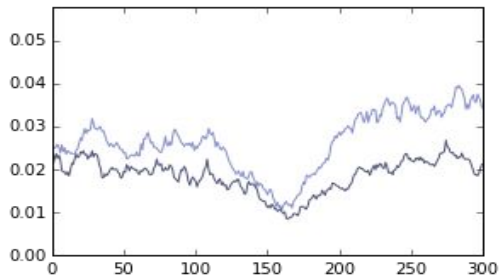
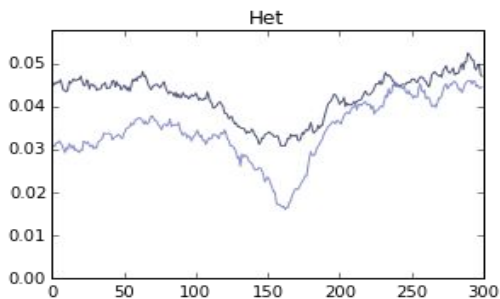
Difference between contexts is consistent between GM-specific and shared loop bases in transcription initiation states.

# Enhancers / Heterochrom highly context-dependent

Enhancers



Heterochromatin



GM-specific loops

shared loops

IMR-specific loops

Light = loci on GM chromHMM  
Dark = loci on IMR chromHMM

# CTCF-enriched loci

CTCF is an architectural protein that has been dubbed the “master weaver” of the genome

GM12878 domain boundaries enriched in the architectural protein CTCF tend to possess more transcription-related chromatin states at their borders.

---

# CTCF-enriched loci

## Domains:

enrichments: TssA, TssAFlnk, TxFlnk, Enh, TssBiv, BivFlnk, EnhBiv

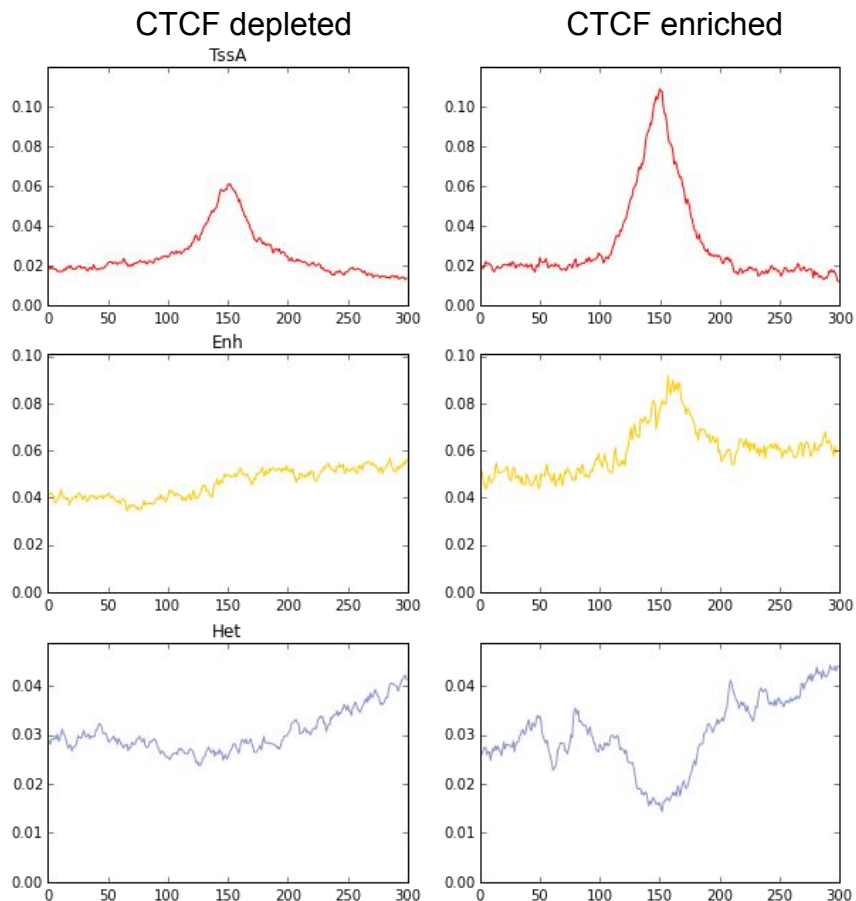
depletions: Tx, ZNF/Rpts, Het, Quies

## Loops:

enrichments: TssA, TssAFlnk, Enh, TssBiv, BivFlnk, EnhBiv

depletions: ZNF/Rpts, Het, ReprPC

## GM Domain Boundaries

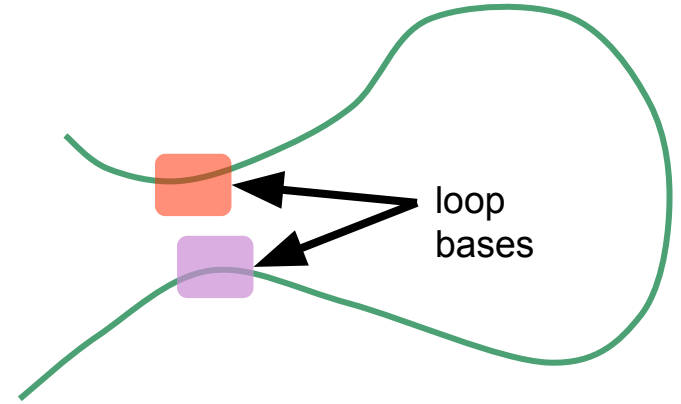
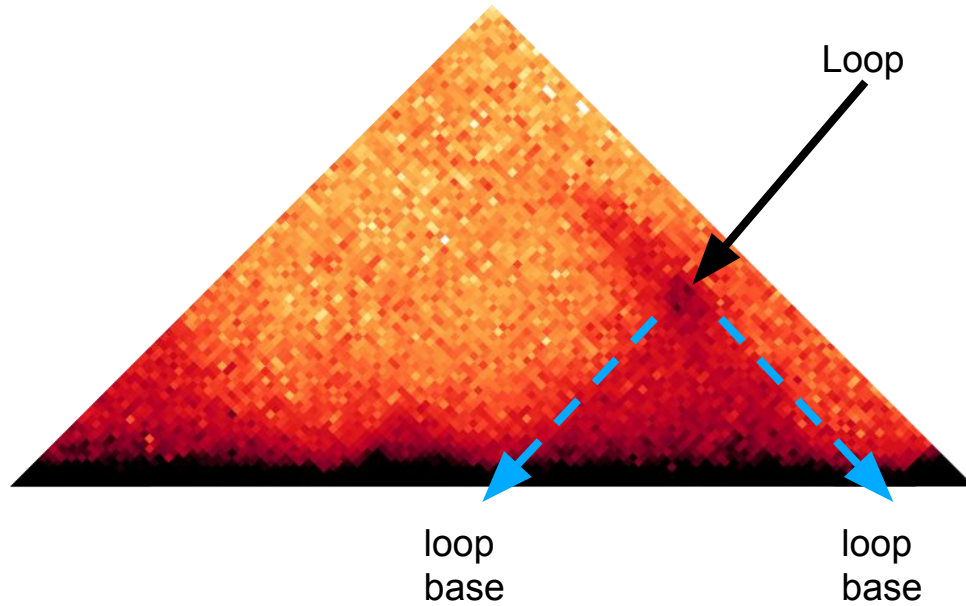


# Similarity of loop starts / ends

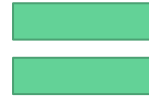
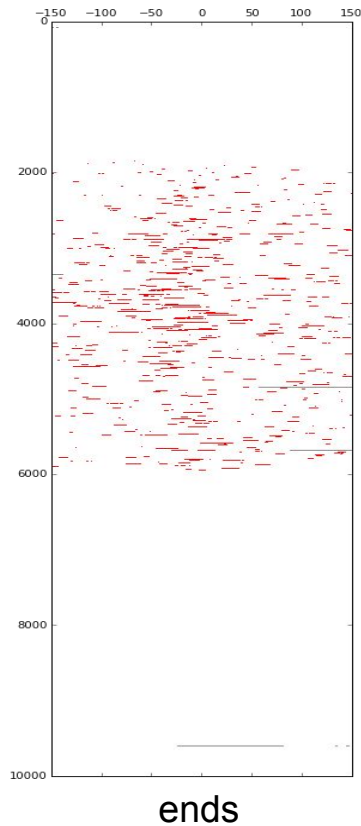
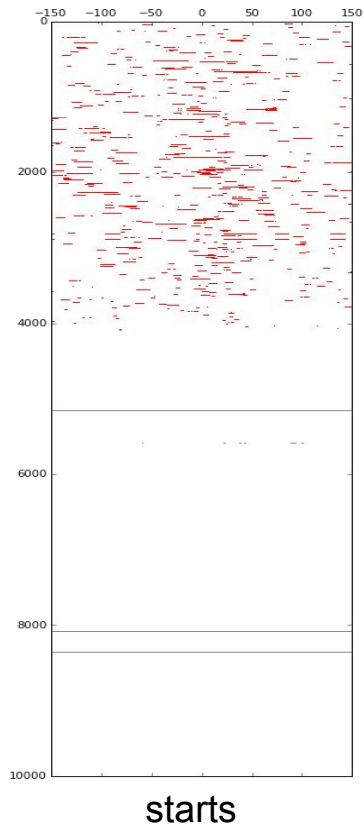
We checked to see if the chromatin state enrichments at the starts and ends of a loop are correlated.

---

# Two bases of a loop

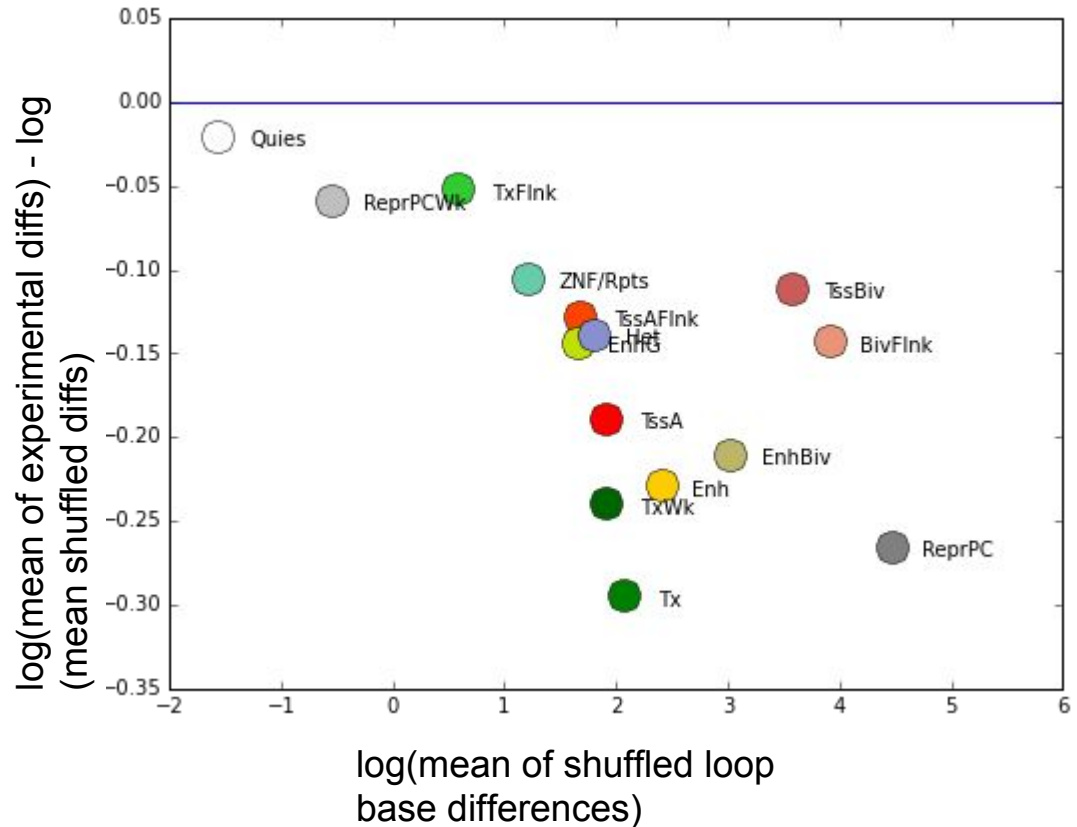
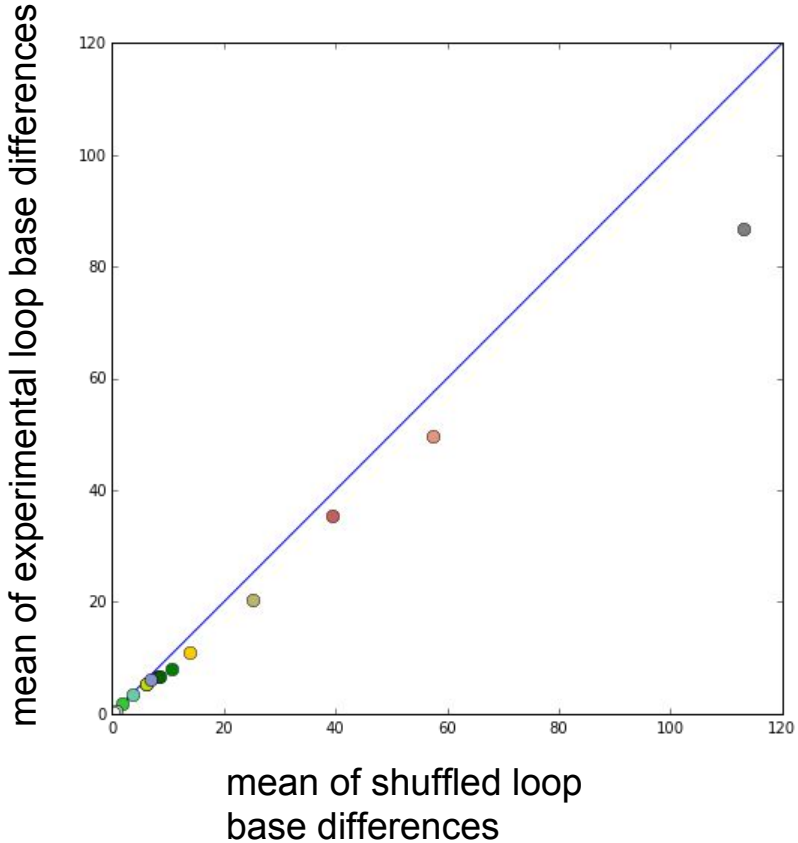


# Similarity of the two bases of a loop

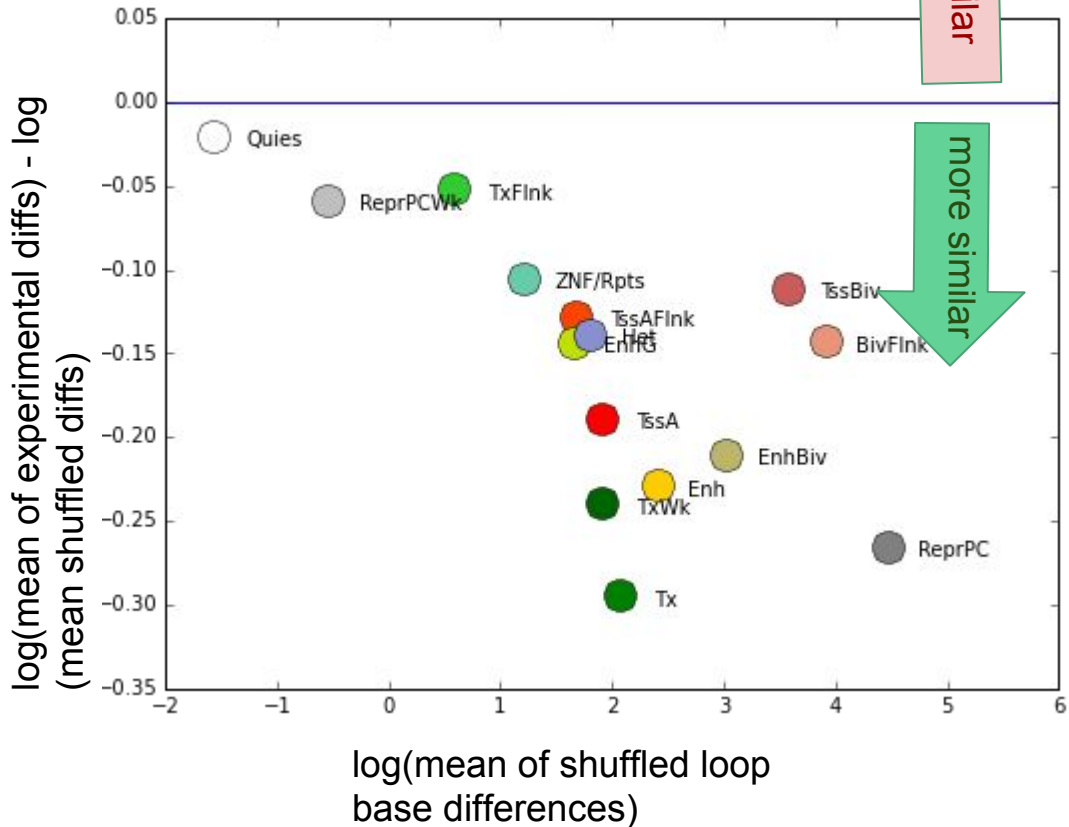
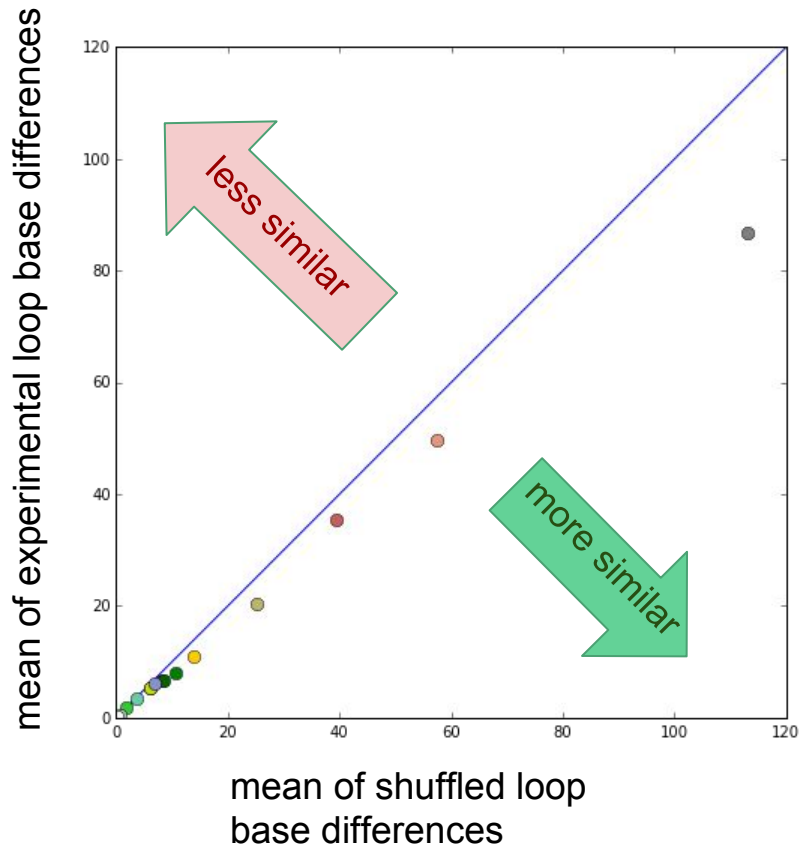




# Loop bases are more similar than shuffled counterparts



# Loop bases are more similar than shuffled counterparts



# Conclusion

- Transcription tends to proceed outside of loops

# Conclusion

- Transcription tends to proceed outside of loops
- GM-specific loop bases show more pronounced TSS activity in the GM context.

# Conclusion

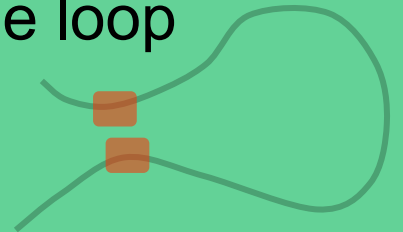
- Transcription tends to proceed outside of loops
- GM-specific loop bases show more pronounced TSS activity in the GM context.
- Transcription elongation at loop bases is always more enriched in IMR90 than GM

# Conclusion

- Transcription tends to proceed outside of loops
- GM-specific loop bases show more pronounced TSS activity in the GM context.
- Transcription elongation at loop bases is always more enriched in IMR90 than GM
- CTCF-enriched domains/loops are enriched in transcription & regulation - related chromatin states

# Conclusion

- Transcription tends to proceed outside of loops
- GM-specific loop bases show more pronounced TSS activity in the GM context.
- Transcription elongation at loop bases is always more enriched in IMR90 than GM
- CTCF-enriched domains/loops are enriched in transcription & regulation - related chromatin states
- Most chromatin states at bases of the same loop show some similarity



# Acknowledgements

## Mirny lab:

Prof. Leonid Mirny

Nezar Abdennur

Boryana Doyle

Dr. Geoff Fudenberg

Dr. Maxim Imakaev

Anton Goloborodko

The Espresso Machine

Dr. Slava Gerovitch and  
Dr. Pavel Etingof for  
dealing with our poor  
adherence to deadlines

And my parents for  
supporting me through  
this journey

