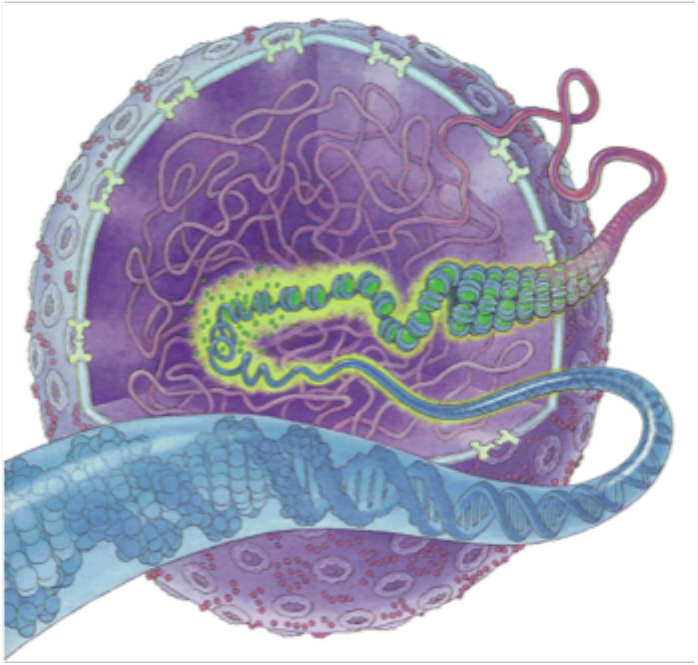
The chromosome folding problem:

How to organize a 2 meter genome into a 20 micron nucleus?



A chromosome consists of a DNA molecule packed together with proteins

The main component of the genome in most bacteria is one double-stranded, circular DNA molecule that is associated with a small amount of protein. Although we refer to this structure

ciated with a large amount of protein. In *E. coli*, the chromosomal DNA consists of about 4.6 million nucleotide pairs, representing about 4,400 genes. This is 100 times more DNA than is found in a typical virus, but only about one-thousandth as much DNA as in a human somatic cell. Still, that is a lot of DNA to be packaged in such a small container.

Stretched out, the DNA of an *E. coli* cell would measure about a millimeter in length, 500 times longer than the cell.

that it fills only part of the cell. Unlike the nucleus of a eukaryotic cell, this dense region of DNA in a bacterium, called the **nucleoid**, is not bounded by membrane (see Figure 6.5).

Eukaryotic chromosomes each contain a single linear DNA double helix that, in humans, averages about 1.5×10^8 nucleotide pairs. This is an enormous amount of DNA relative to a chromosome's condensed length. If completely stretched out, such a DNA molecule would be about 4 cm

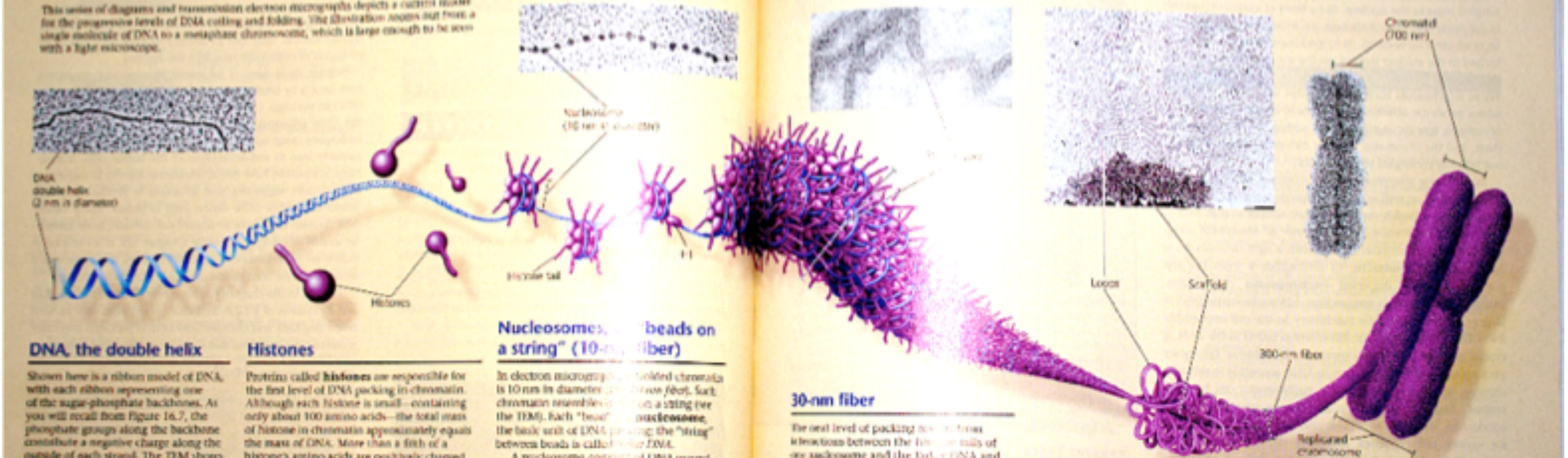
chromosome!

In the cell, eukaryotic DNA is precisely combined with a large amount of protein. Together, this complex of DNA and protein, called **chromatin**, fits into the nucleus through an elaborate, multilevel system of packing. Our current view of the successive levels of DNA packing in a chromosome is outlined in **Figure 16.22**. Study this figure carefully before reading further.

▼ **Figure 16.22**

Exploring Chromatin Packing in a Eukaryotic Chromosome

This series of diagrams and transmission electron micrographs depicts a current model for the progressive levels of DNA coiling and folding. The illustration zooms out from a single molecule of DNA to a metaphase chromosome, which is large enough to be seen with a light microscope.



DNA, the double helix

Shown here is a ribbon model of DNA, with each ribbon representing one of the sugar-phosphate backbones. As you will recall from Figure 16.7, the phosphate groups along the backbone contribute a negative charge along the outside of each strand. The TEM shows a molecule of naked DNA; the double helix alone is 2 nm across.

Histones

Proteins called **histones** are responsible for the first level of DNA packing in chromatin. Although each histone is small—containing only about 100 amino acids—the total mass of histone in chromatin approximately equals the mass of DNA. More than a fifth of a histone's amino acids are positively charged (lysine or arginine) and therefore bind tightly to the negatively charged DNA.

Four types of histones are most common in chromatin: H2A, H2B, H3, and H4. The histones are very similar among eukaryotes; for example, all but two of the amino acids in cow H4 are identical to those in pea H4. The apparent conservation of histone genes during evolution probably reflects the important role of histones in organizing DNA within cells.

The four main types of histones are critical to the next level of DNA packing. (A fifth type of histone, called H1, is involved in a further stage of packing.)

Nucleosomes, or "beads on a string" (10-nm fiber)

In electron micrographs, unfolded chromatin is 10 nm in diameter (the *10-nm fiber*). Such chromatin resembles beads on a string (see the TEM). Each "bead" is a **nucleosome**, the basic unit of DNA packing; the "string" between beads is called *linker DNA*.

A nucleosome consists of DNA wound twice around a protein core composed of two molecules each of the four main histone types. The amino end (N-terminus) of each histone (the *histone tail*) extends outward from the nucleosome.

In the cell cycle, the histones leave the DNA only briefly during DNA replication. Generally, they do the same during transcription, another process that requires access to the DNA by the cell's molecular machinery. Chapter 18 will discuss some recent findings about the role of histone tails and nucleosomes in the regulation of gene expression.

30-nm fiber

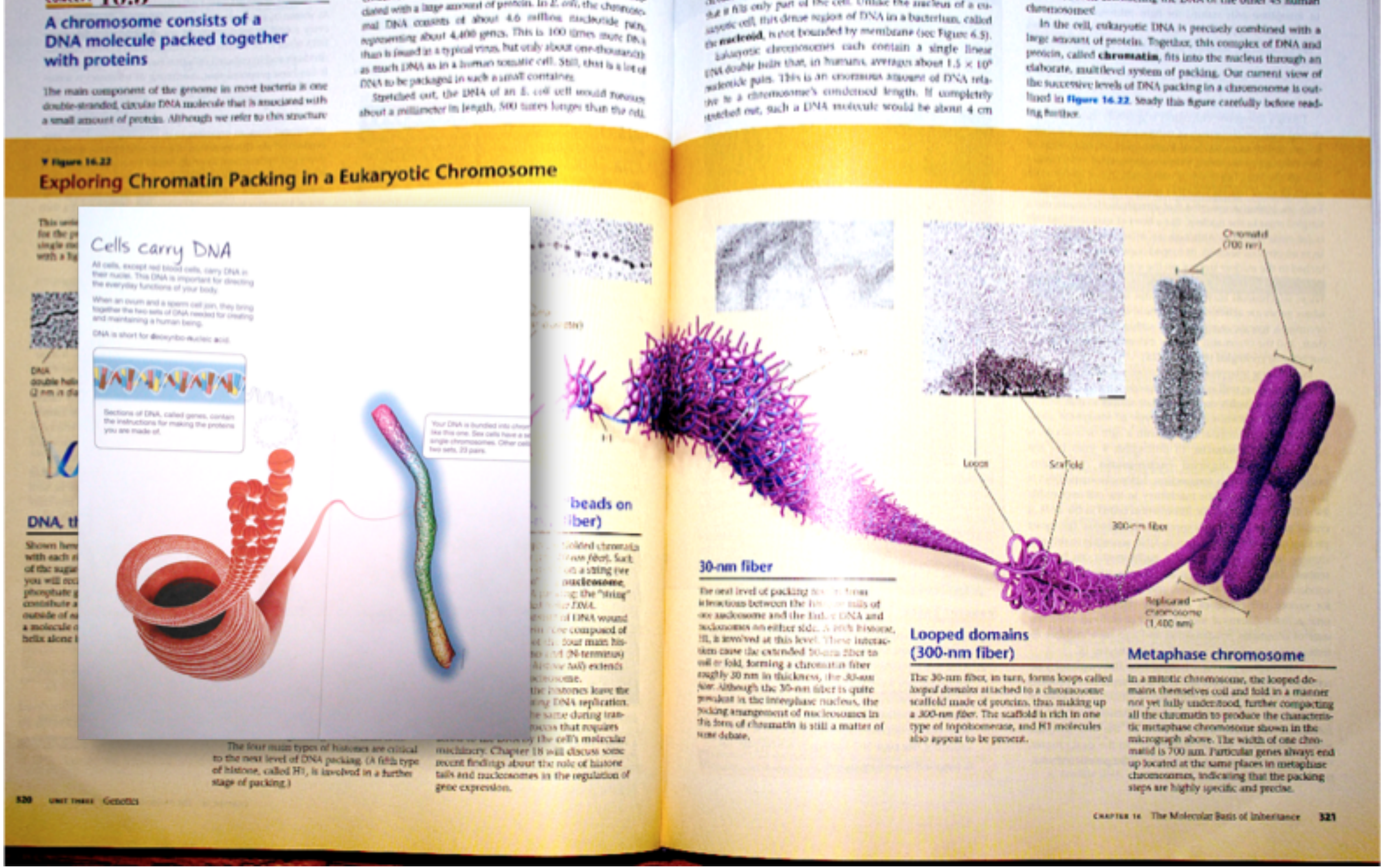
The next level of packing is due to interactions between the histone tails of one nucleosome and the linker DNA and nucleosomes on either side. A fifth histone, H1, is involved at this level. These interactions cause the extended 10-nm fiber to coil or fold, forming a chromatin fiber roughly 30 nm in thickness, the *30-nm fiber*. Although the 30-nm fiber is quite prevalent in the interphase nucleus, the packing arrangement of nucleosomes in this form of chromatin is still a matter of some debate.

Looped domains (300-nm fiber)

The 30-nm fiber, in turn, forms loops called *looped domains* attached to a chromosome scaffold made of proteins, thus making up a *300-nm fiber*. The scaffold is rich in one type of topoisomerase, and H1 molecules also appear to be present.

Metaphase chromosome

In a mitotic chromosome, the looped domains themselves coil and fold in a manner not yet fully understood, further compacting all the chromatin to produce the characteristic metaphase chromosome shown in the micrograph above. The width of one chromatid is 700 nm. Particular genes always end up located at the same places in metaphase chromosomes, indicating that the packing steps are highly specific and precise.



A chromosome consists of a DNA molecule packed together with proteins

The main component of the genome in most bacteria is one double-stranded, circular DNA molecule that is associated with a small amount of protein. Although we refer to this structure

…ciated with a large amount of protein. In *E. coli*, the chromosomal DNA consists of about 4.6 million nucleotide pairs, representing about 4,400 genes. This is 100 times more DNA than is found in a typical virus, but only about one-thousandth as much DNA as in a human somatic cell. Still, that is a lot of DNA to be packaged in such a small container.

Stretched out, the DNA of an *E. coli* cell would measure about a millimeter in length, 500 times longer than the cell.

…that it fills only part of the cell. Unlike the nucleus of a eukaryotic cell, this dense region of DNA in a bacterium, called the **nucleoid**, is not bounded by membrane (see Figure 6.5).

Eukaryotic chromosomes each contain a single linear DNA double helix that, in humans, averages about 1.5 × 10^8 nucleotide pairs. This is an enormous amount of DNA relative to a chromosome's condensed length. If completely stretched out, such a DNA molecule would be about 4 cm

…chromosome!

In the cell, eukaryotic DNA is precisely combined with a large amount of protein. Together, this complex of DNA and protein, called **chromatin**, fits into the nucleus through an elaborate, multilevel system of packing. Our current view of the successive levels of DNA packing in a chromosome is outlined in **Figure 16.22**. Study this figure carefully before reading further.

▼ Figure 16.22

Exploring Chromatin Packing in a Eukaryotic Chromosome

Cells carry DNA

All cells, except red blood cells, carry DNA in their nuclei. This DNA is important for directing the everyday functions of your body.

When an ovum and a sperm cell join, they bring together the two sets of DNA needed for creating and maintaining a human being.

DNA is short for deoxyribo-nucleic acid.

Sections of DNA, called genes, contain the instructions for making the proteins you are made of.

30-nm fiber

The next level of packing results from interactions between the histone tails of one nucleosome and the linker DNA and nucleosomes on either side. A fifth histone, H1, is involved at this level. These interactions cause the extended 10-nm fiber to coil or fold, forming a chromatin fiber roughly 30 nm in thickness, the *30-nm fiber*. Although the 30-nm fiber is quite prevalent in the interphase nucleus, the packing arrangement of nucleosomes in this form of chromatin is still a matter of some debate.

Looped domains (300-nm fiber)

The 30-nm fiber, in turn, forms loops called *looped domains* attached to a chromosome scaffold made of proteins, thus making up a *300-nm fiber*. The scaffold is rich in one type of topoisomerase, and H1 molecules also appear to be present.

Metaphase chromosome

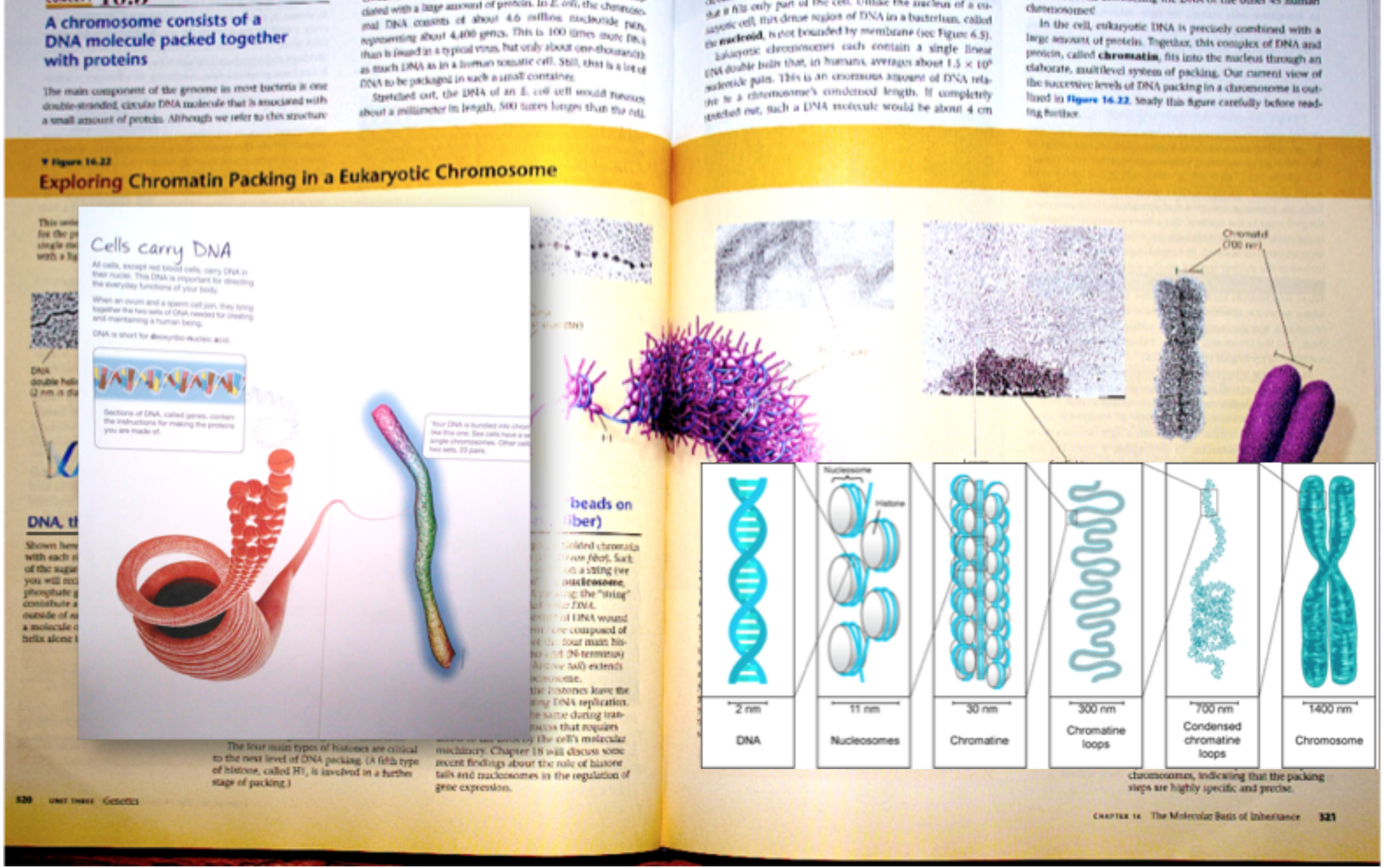
In a mitotic chromosome, the looped domains themselves coil and fold in a manner not yet fully understood, further compacting all the chromatin to produce the characteristic metaphase chromosome shown in the micrograph above. The width of one chromatid is 700 nm. Particular genes always end up located at the same places in metaphase chromosomes, indicating that the packing steps are highly specific and precise.

The four main types of histones are critical to the next level of DNA packing. (A fifth type of histone, called H1, is involved in a further stage of packing.)

…machinery. Chapter 18 will discuss some recent findings about the role of histone tails and nucleosomes in the regulation of gene expression.

520 UNIT THREE Genetics

CHAPTER 16 The Molecular Basis of Inheritance 321



A chromosome consists of a DNA molecule packed together with proteins

The main component of the genome in most bacteria is one double-stranded, circular DNA molecule that is associated with a small amount of protein. Although we refer to this structure

…ciated with a large amount of protein. In *E. coli*, the chromosomal DNA consists of about 4.6 million nucleotide pairs, representing about 4,400 genes. This is 100 times more DNA than is found in a typical virus, but only about one-thousandth as much DNA as in a human somatic cell. Still, that is a lot of DNA to be packaged in such a small container.

Stretched out, the DNA of an *E. coli* cell would measure about a millimeter in length, 500 times longer than the cell.

…that it fills only part of the cell. Unlike the nucleus of a eukaryotic cell, this dense region of DNA in a bacterium, called the **nucleoid**, is not bounded by membrane (see Figure 6.5).

Eukaryotic chromosomes each contain a single linear DNA double helix that, in humans, averages about 1.5×10^8 nucleotide pairs. This is an enormous amount of DNA relative to a chromosome's condensed length. If completely stretched out, such a DNA molecule would be about 4 cm

…chromosomes!

In the cell, eukaryotic DNA is precisely combined with a large amount of protein. Together, this complex of DNA and protein, called **chromatin**, fits into the nucleus through an elaborate, multilevel system of packing. Our current view of the successive levels of DNA packing in a chromosome is outlined in **Figure 16.22**. Study this figure carefully before reading further.

▼ Figure 16.22

Exploring Chromatin Packing in a Eukaryotic Chromosome

Cells carry DNA

All cells, except red blood cells, carry DNA in their nuclei. This DNA is important for directing the everyday functions of your body.

When an ovum and a sperm cell join, they bring together the two sets of DNA needed for creating and maintaining a human being.

DNA is short for deoxyribo-nucleic acid.

Sections of DNA, called genes, contain the instructions for making the proteins you are made of.

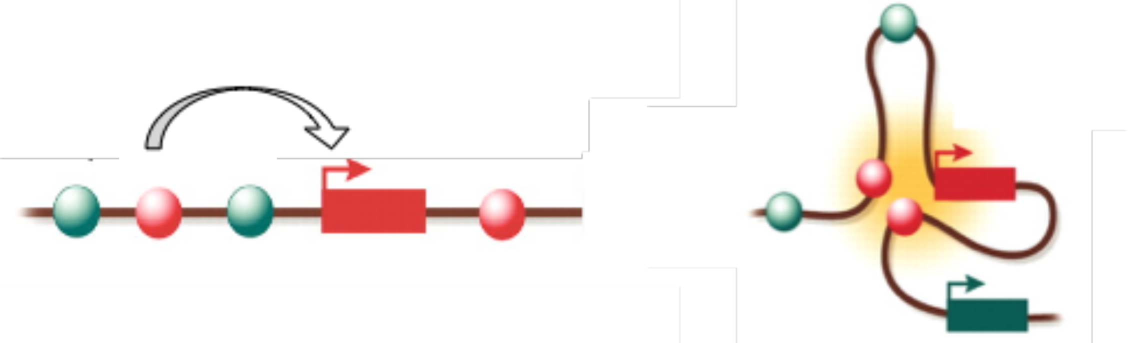
Your DNA is bundled into chrom… like this one. Sex cells have a s… single chromosomes. Other cell… two sets, 23 pairs.

520 UNIT THREE Genetics

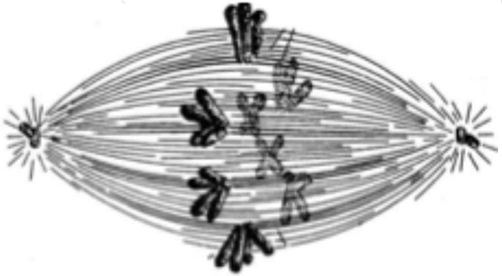
CHAPTER 16 The Molecular Basis of Inheritance 521

Why study chromosome organization?

Read genetic information
(gene regulation)



Transmit genetic information
(DNA replication & cell division)

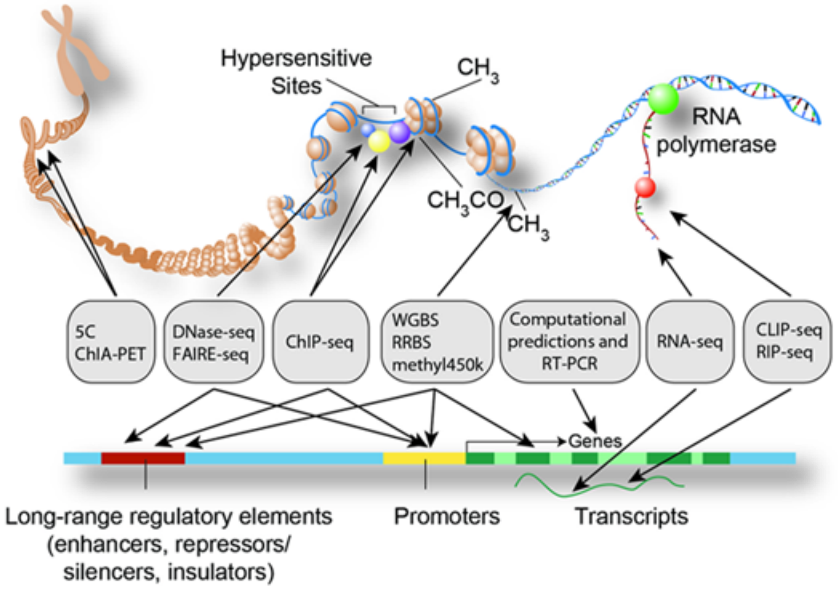


Human genome project-- now that we know our ATCGs, what do they mean?

...ACCTCTCCTACAAGGACAGGAAGGGGTGGTGGTTCTGGATGA
AGGGCGTGAAGCTGGACTCGGTGCCGAGACGGGTCCAATTCCCGG
TGCCGTTCTTCCACCGGCTGCGCGACTGGTTGCGGCACCGCGTGC
ACCTGCTGTACGGGTTGCGCGACAGGCGGGACTCGCTGGACGTGC
GCGTGTTCGAAGCCCACCTGGGCCAGTTGAAGTTCGAGGATTCGG
TGACGGACGACCACTGGGACCGGCGGGTGGAGGGGCGGCTCAAGT
GGGGACGCCACGTGCGGAGGGACCTGTTCAAGGACCGAAGACACT
CGTGGCACGACTGGAGGTTTATGGCAATTCGACCTCGGAGCCATC
GTCAAGGAGGACGGTCTACCCGGAGGGTTGCCCGGGAGGAGGGGA
GGAACGTGGCCGGGAAGGACCAGAAACTTATTTCAGACTCACCCG
CCG...

One 10-millionth part of the human genome

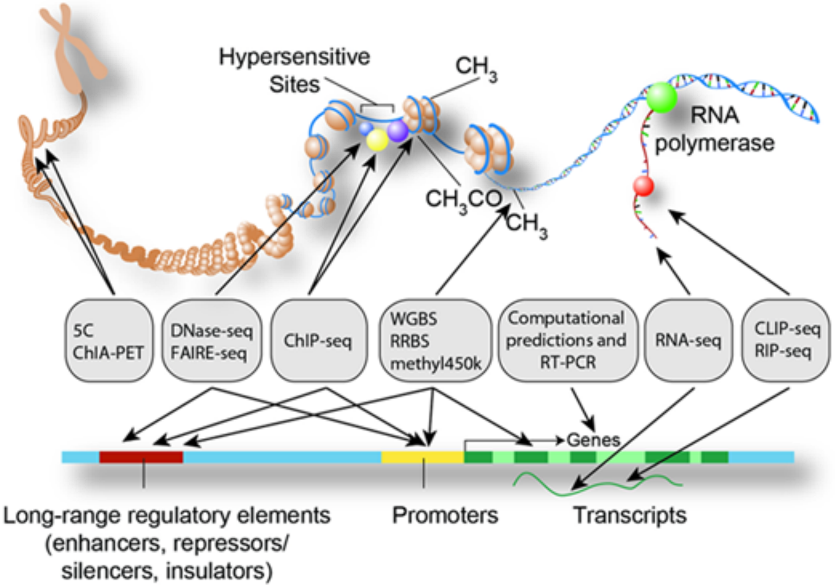
ENCODE: mapping functional states along the 1D genome

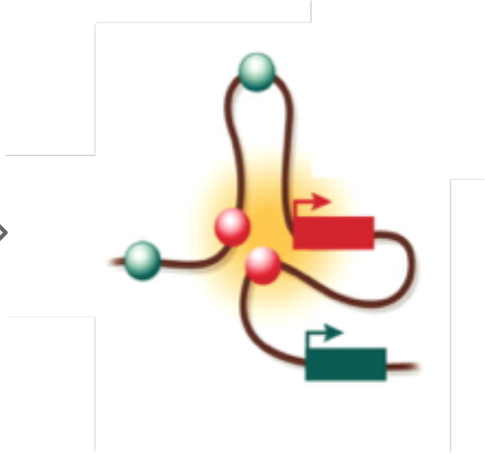


Goal: Develop a parts list of **functional elements** in the human genome

https://www.encodeproject.org/

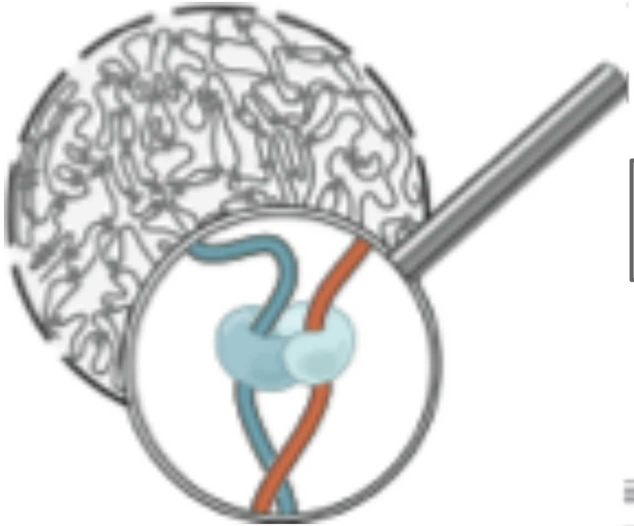
How is 1D functional organization related to 3D spatial organization?



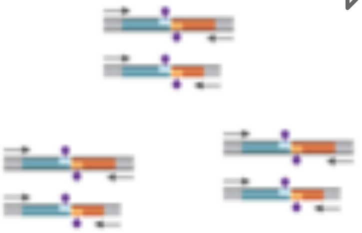


Hi-C is a ‘molecular microscope’ for studying the 3D genome

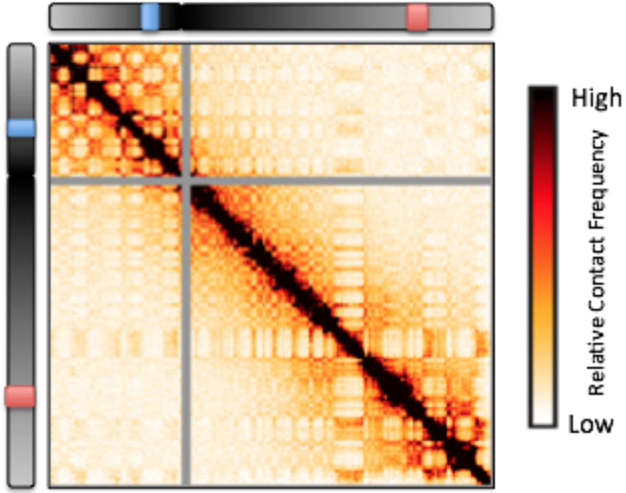
Crosslinking to `freeze’ chromosomes in place



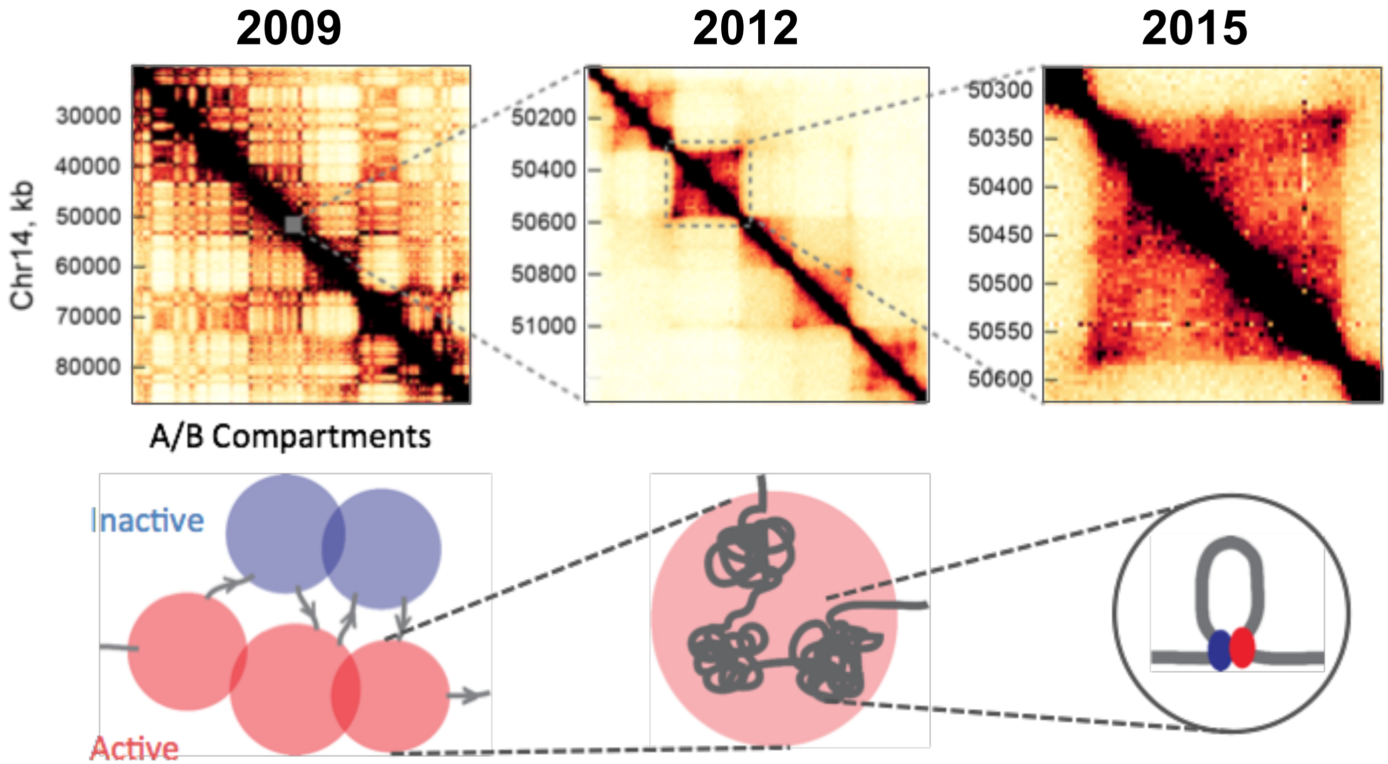
High-throughput sequencing



Chromosome contact maps



Hi-C reveals multiple hierarchical levels of chromosome organization

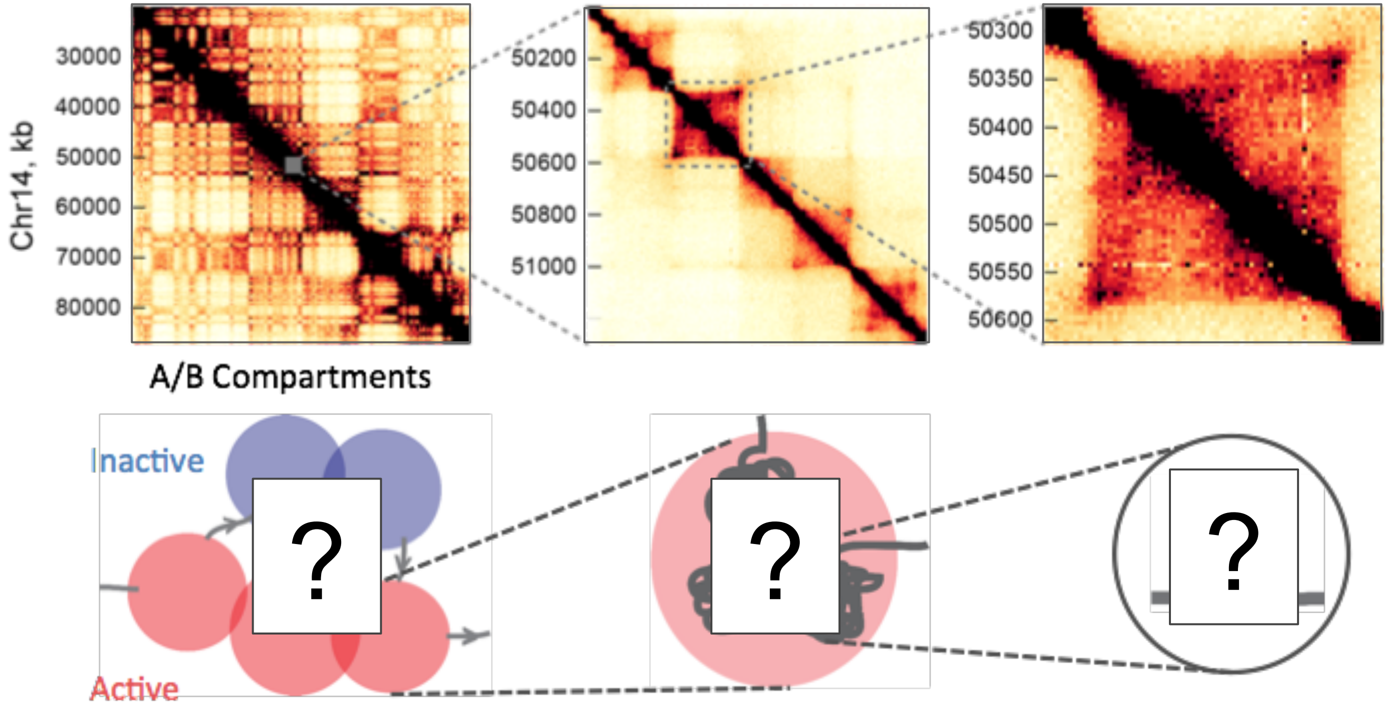


Lieberman Aiden, et al. Science 2009

Nora et al. Nature 2012

Rao et al. Cell 2014

Hi-C reveals multiple hierarchical levels of chromosome organization


2009

2012

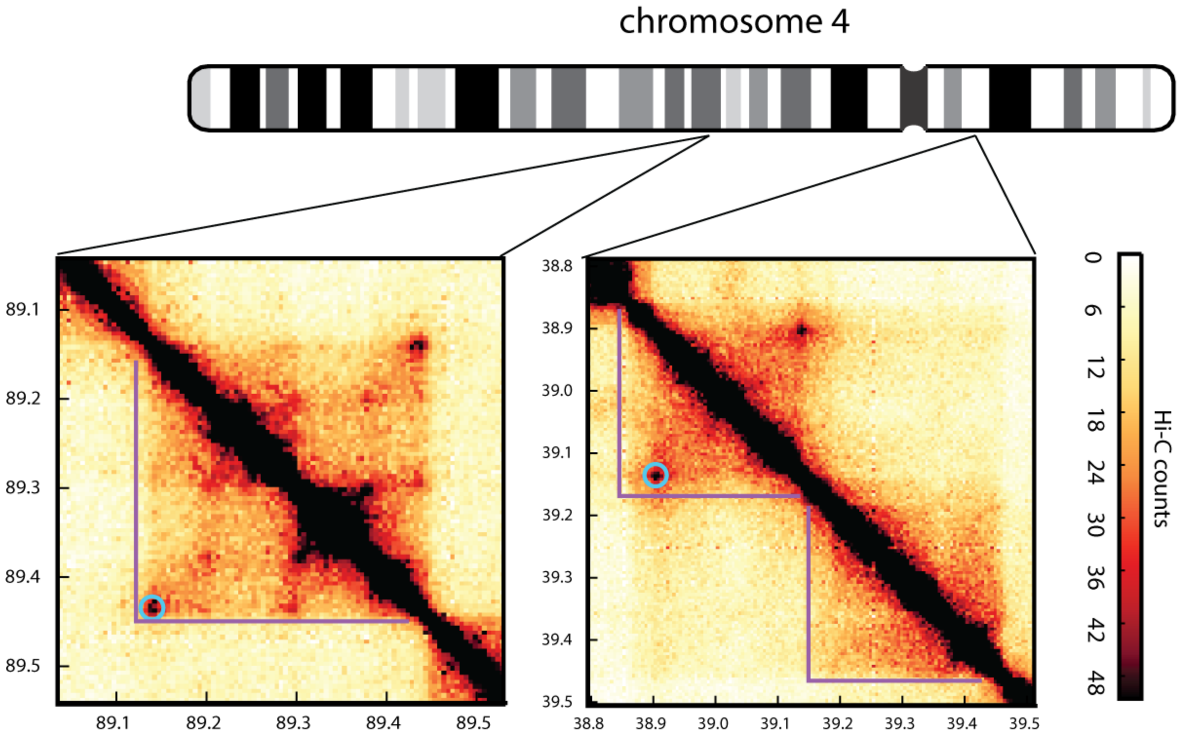
2015

Lieberman Aiden, et al. Science 2009

Nora et al. Nature 2012

Rao et al. Cell 2014

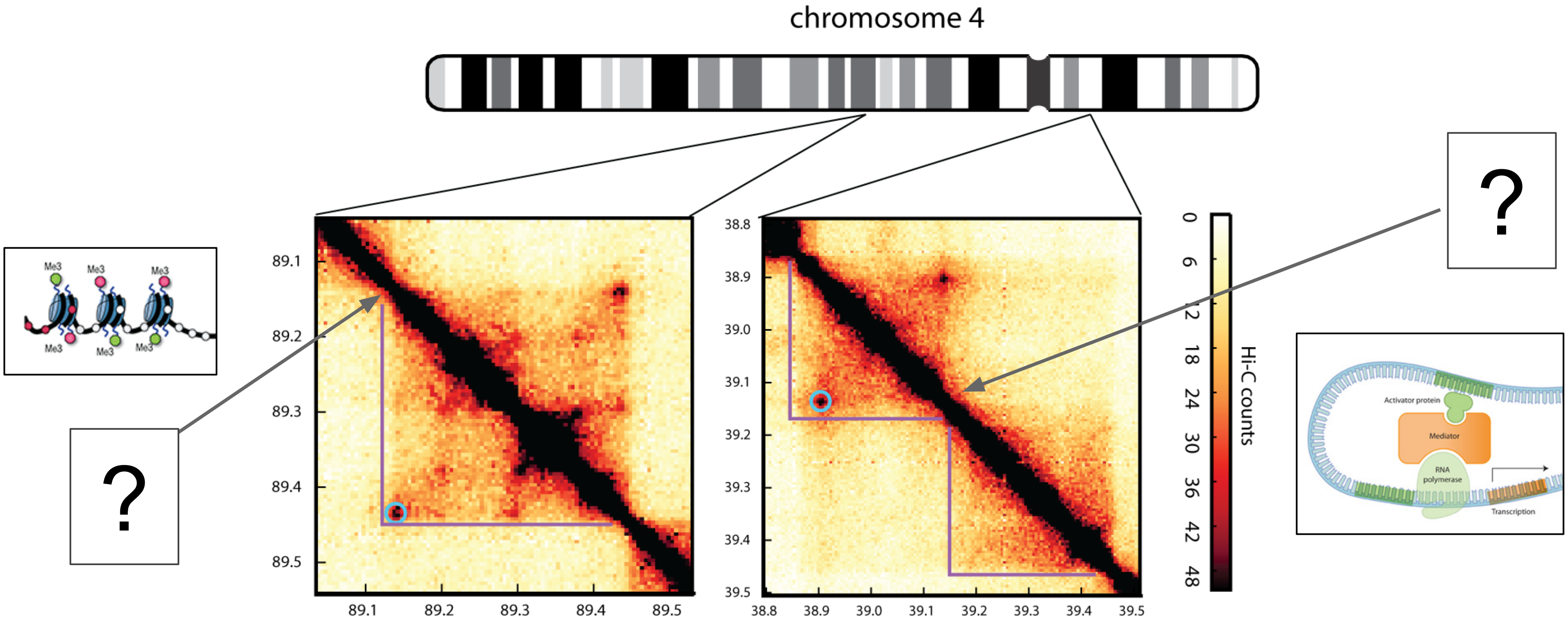
High-resolution Hi-C reveals a diversity of domain organization



Rao et al., 2014

Research Question:

How is the genome functionally organized around loops and domains?

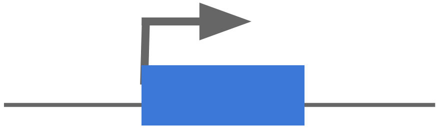


Rao et al., 2014

Functional datasets examined:

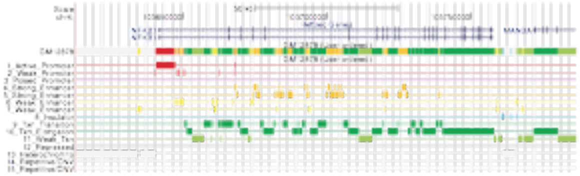
-- Transcription start sites

(TSS)



-- ENCODE & Roadmap states

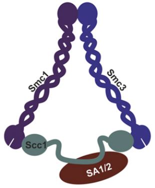
(promoters, enhancers, insulators)



Ernst 2012

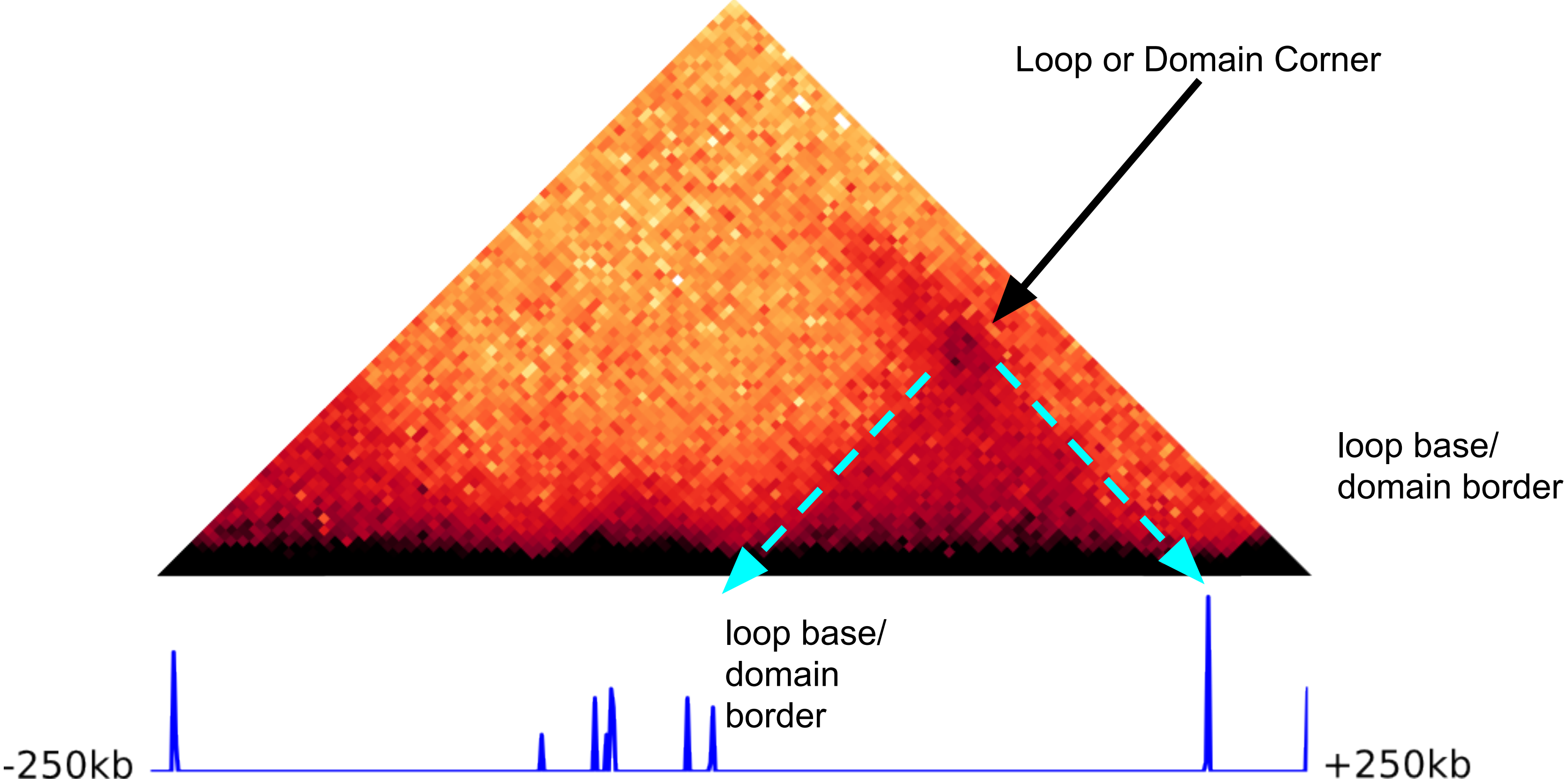
-- architectural proteins

(CTCF, SMC3, Rad21)



Maksimenko 2014
Rudan 2015

Method: aggregate analysis of genomic features around loops or domains

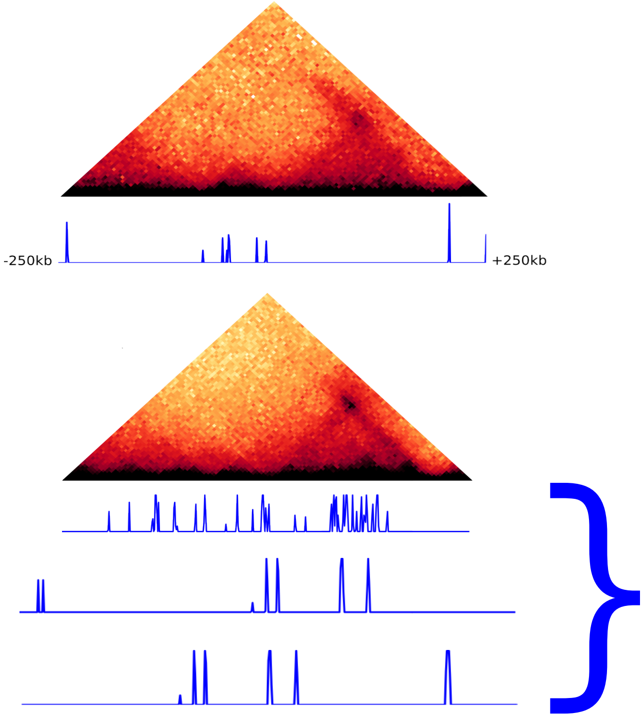


Loop or Domain Corner

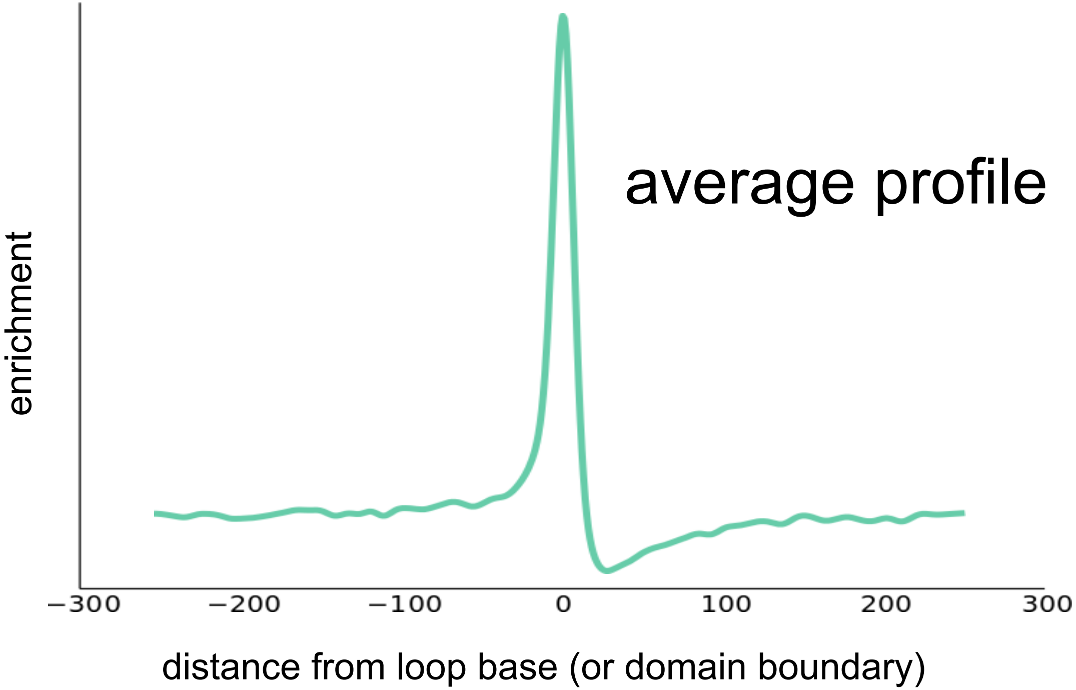
loop base/
domain border

loop base/
domain
border

Method: aggregate analysis of genomic features around loops or domains



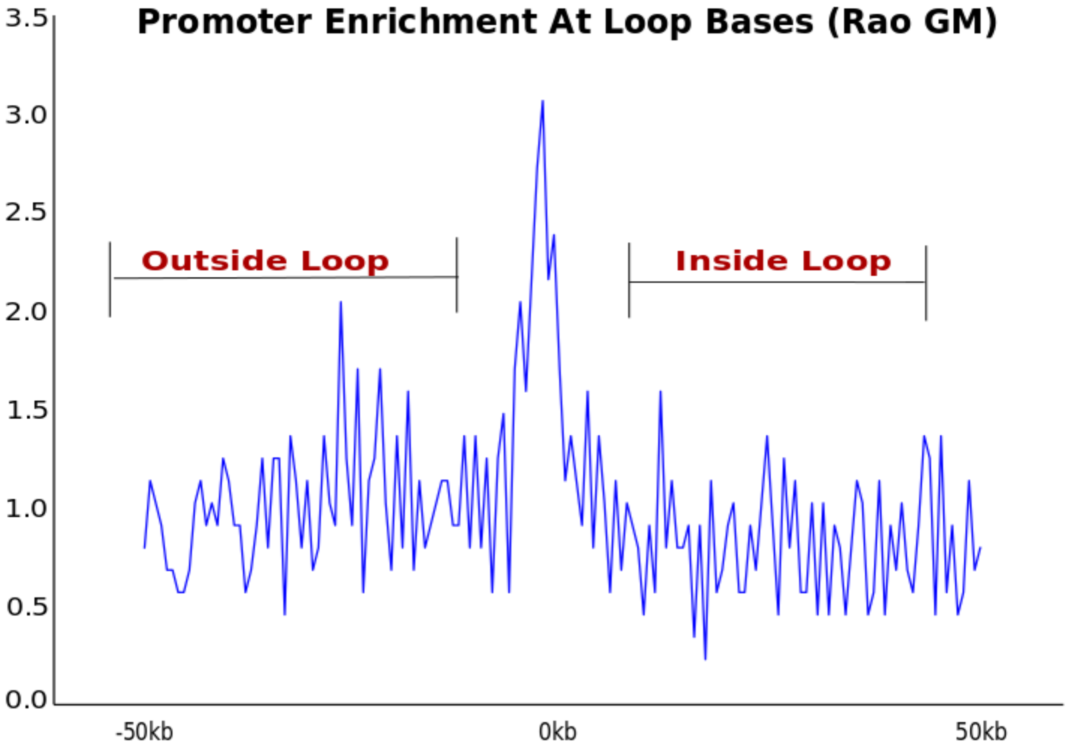
individual profiles



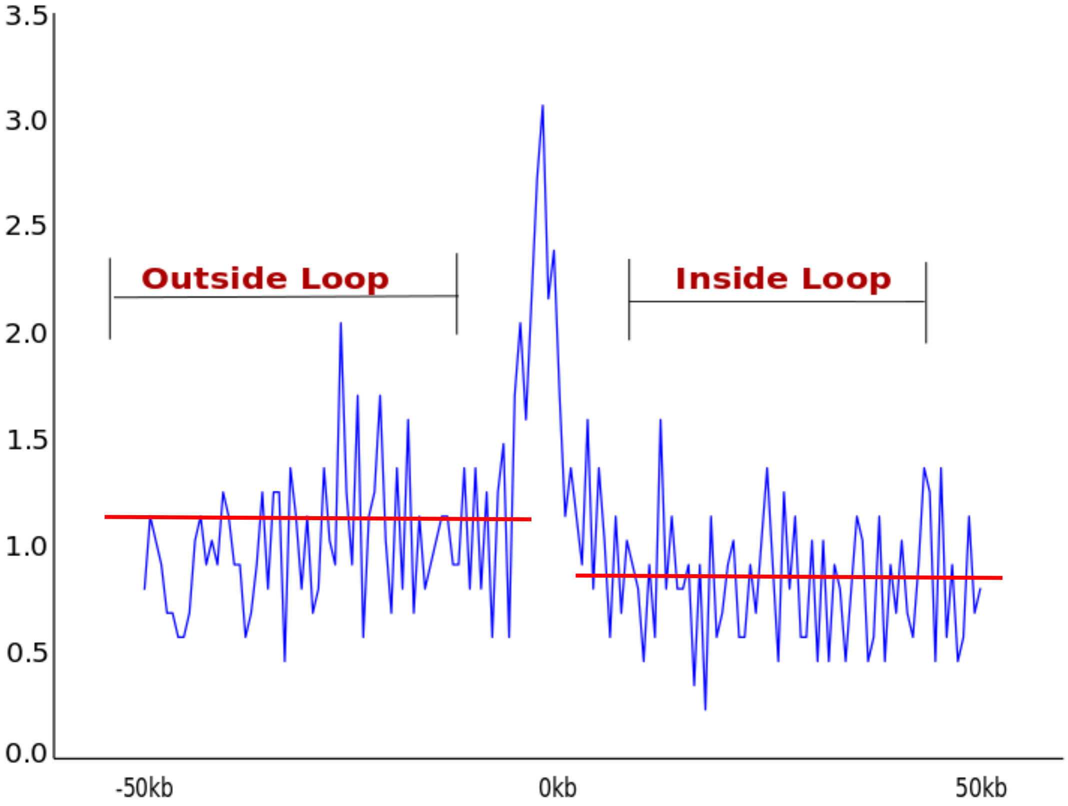
average profile

TSS Enrichment

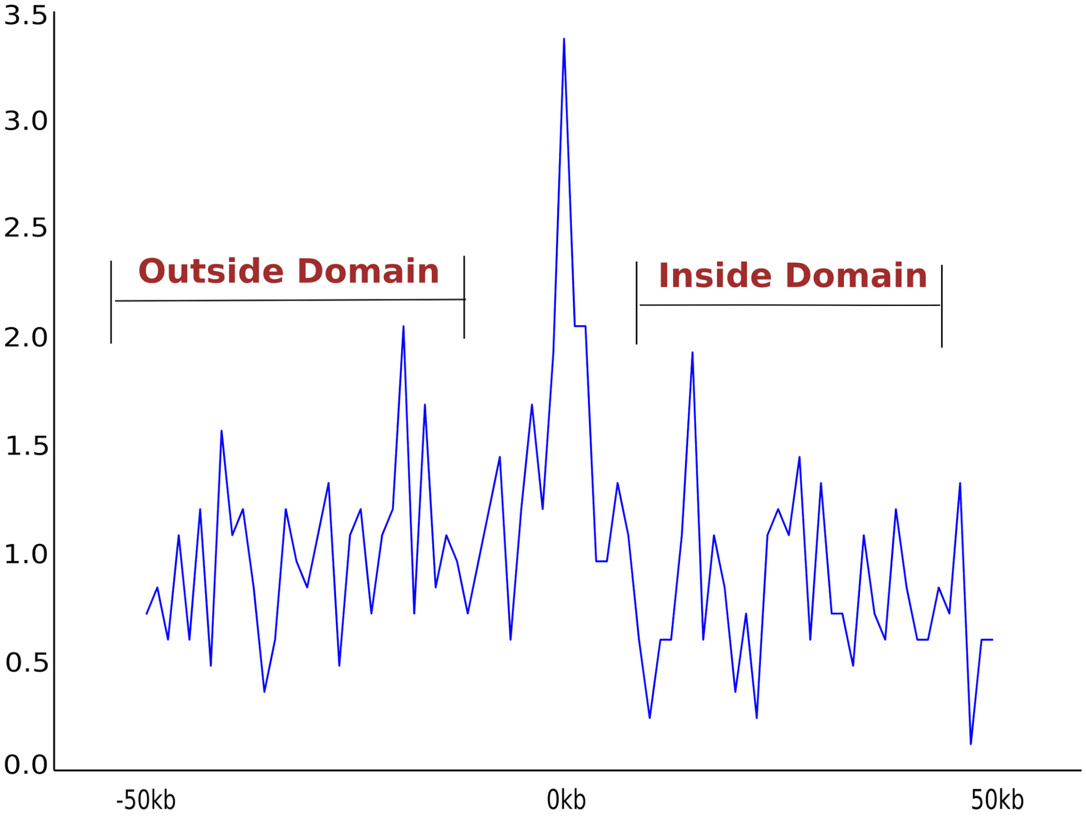
TSS Enriched At Loops



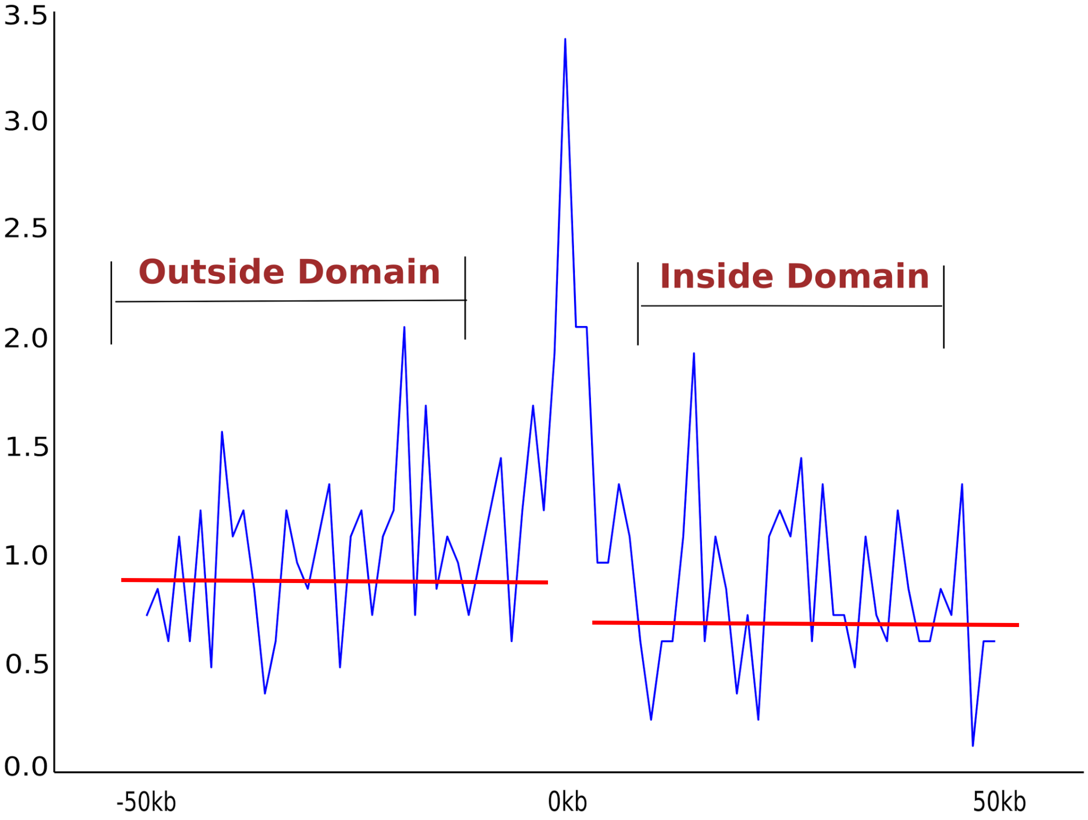
TSS Depleted Inside Loops



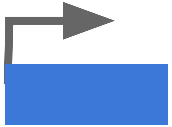
TSS Enriched At Domains

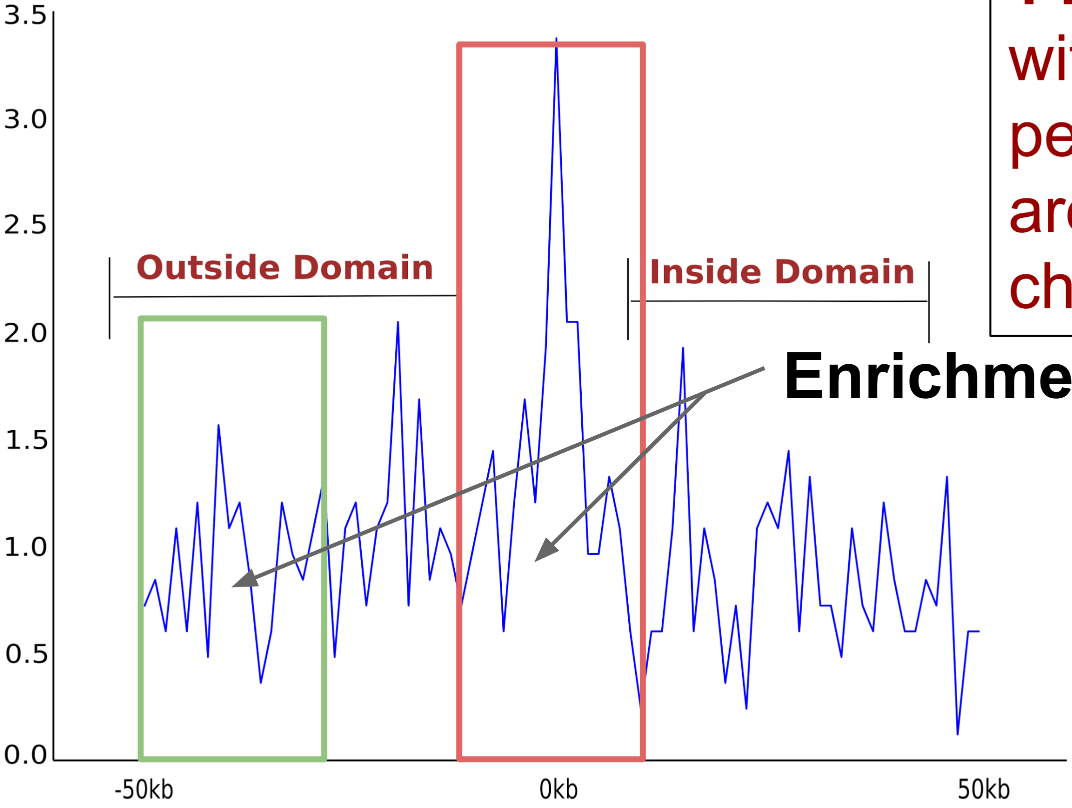


TSS Depleted Inside Domains



2 Fold Enrichment Near Domain

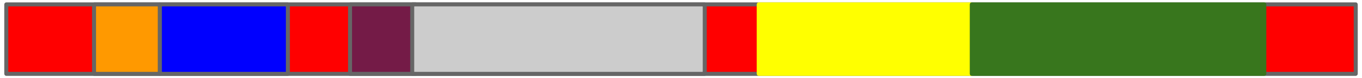




Enrichment

14% TSSs are within 10kb of peaks, while **8%** are expected by chance

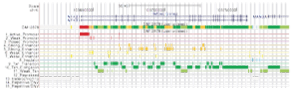
Chromatin States

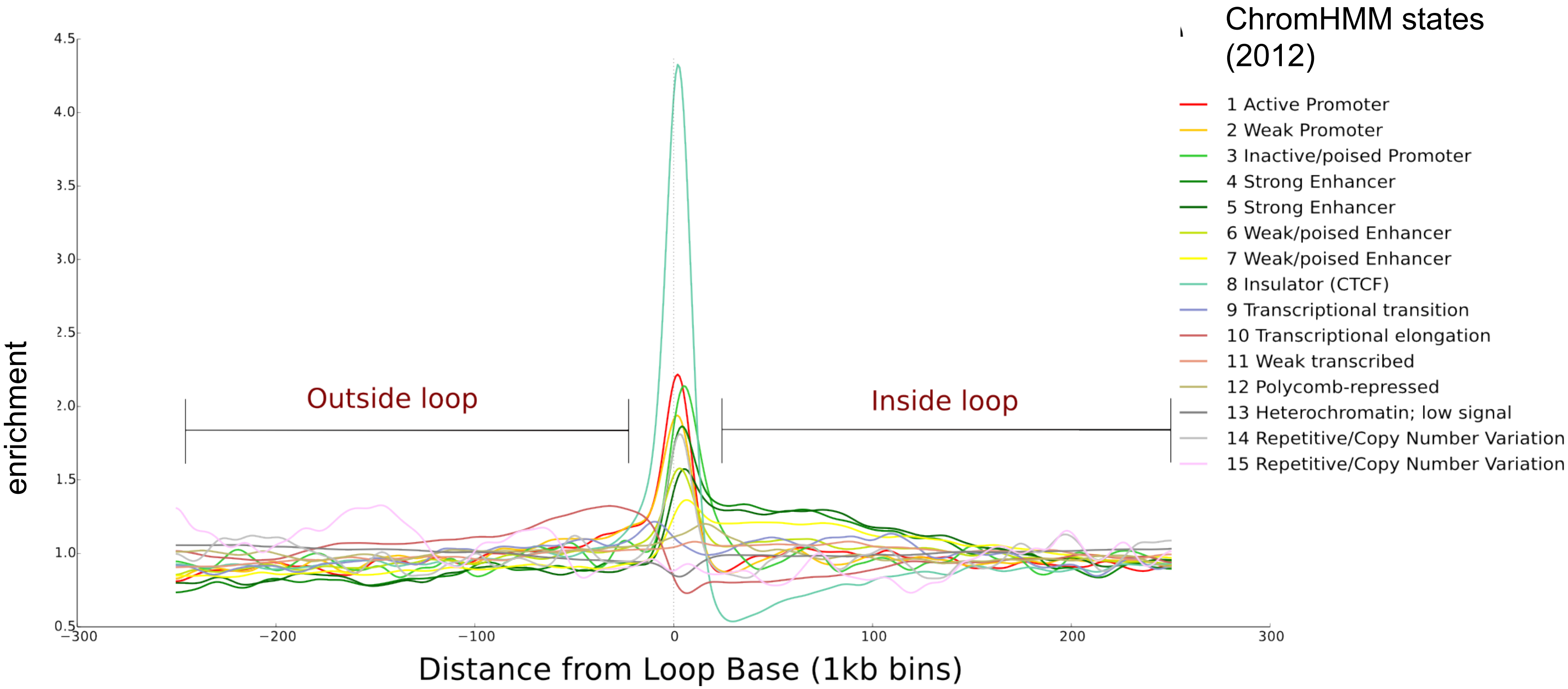


— 1 Active Promoter
— 2 Weak Promoter
— 3 Inactive/poised Promoter
— 4 Strong Enhancer
— 5 Strong Enhancer
— 6 Weak/poised Enhancer
— 7 Weak/poised Enhancer

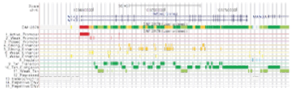
— 8 Insulator (CTCF)
— 9 Transcriptional transition
— 10 Transcriptional elongation
— 11 Weak transcribed
— 12 Polycomb-repressed
— 13 Heterochromatin; low signal
— 14 Repetitive/Copy Number Variation
— 15 Repetitive/Copy Number Variation

Chromatin states have diverse behaviors around loop bases

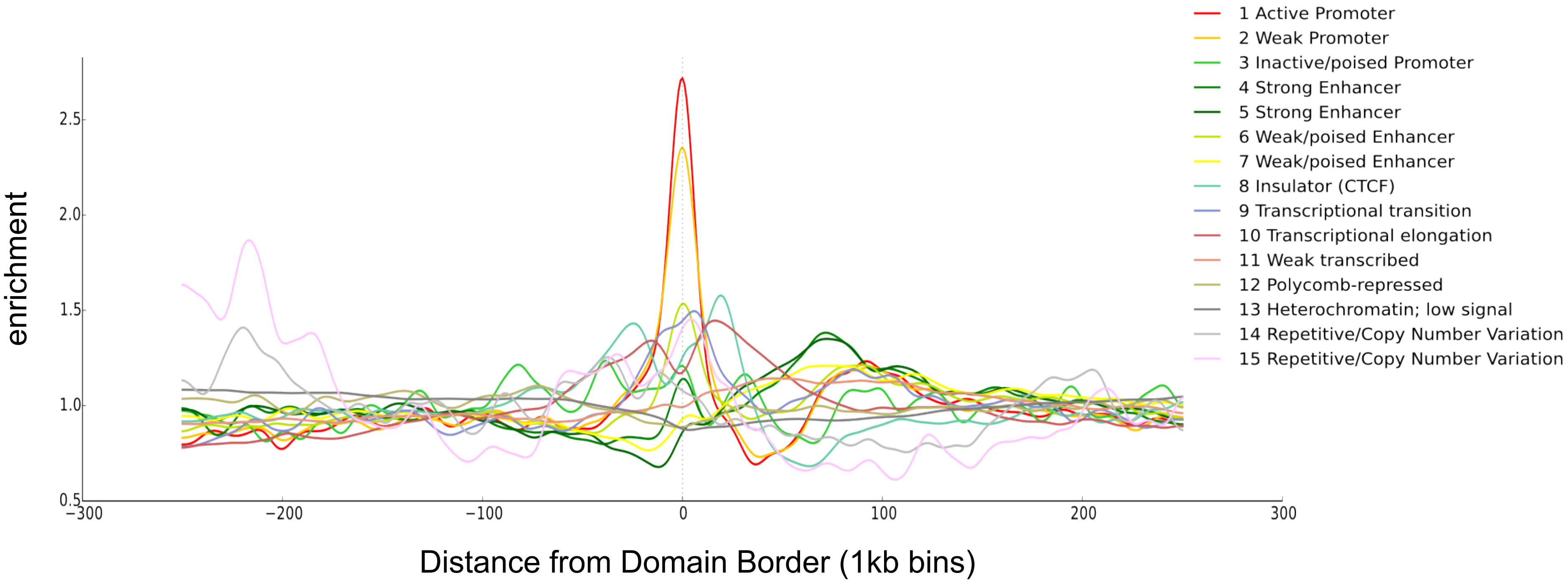




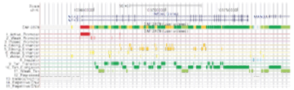
Chromatin states have diverse and different behaviors around domain boundaries



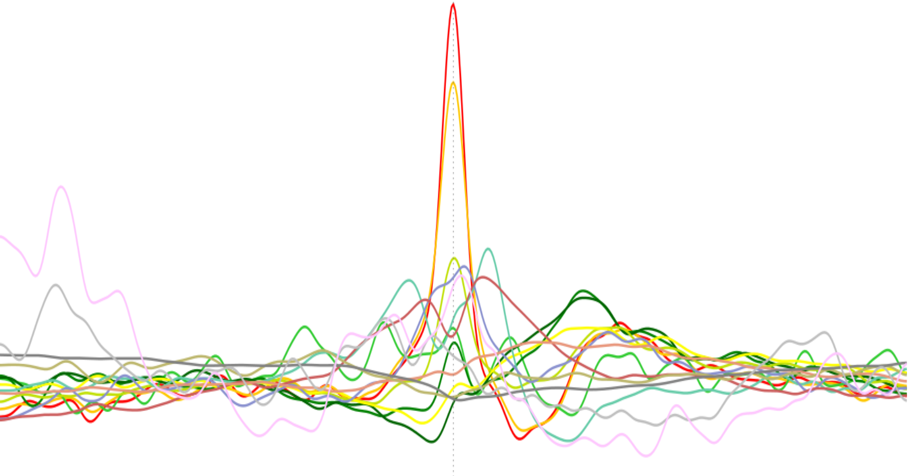
ChromHMM states (2012)



Similar by-state enrichments observed for publicly-available & in house annotations of domain boundaries

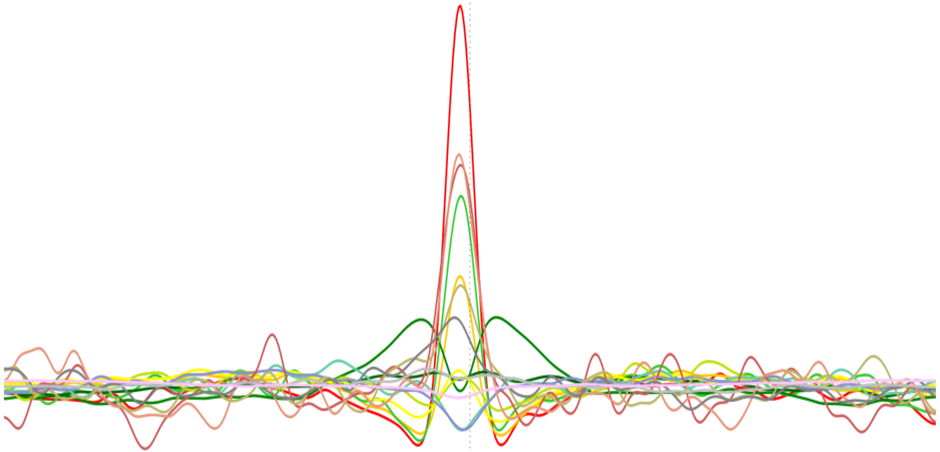


Rao GM Domains vs ENCODE HMM states



Rao et. al (2014)

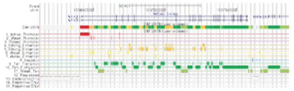
in house GM Domains vs ENCODE HMM states

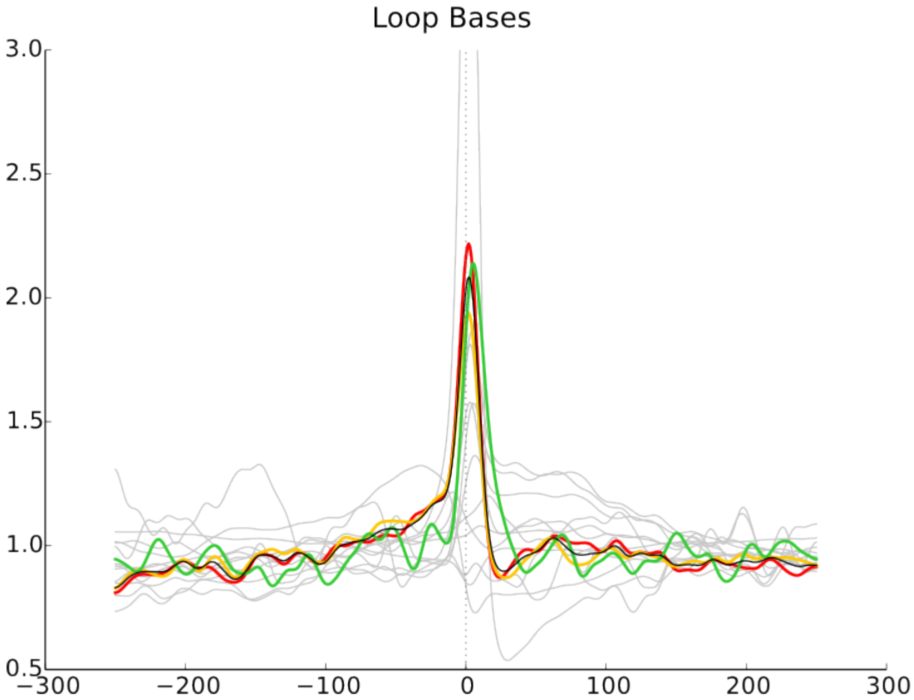


Abdennur et. al (in preparation)

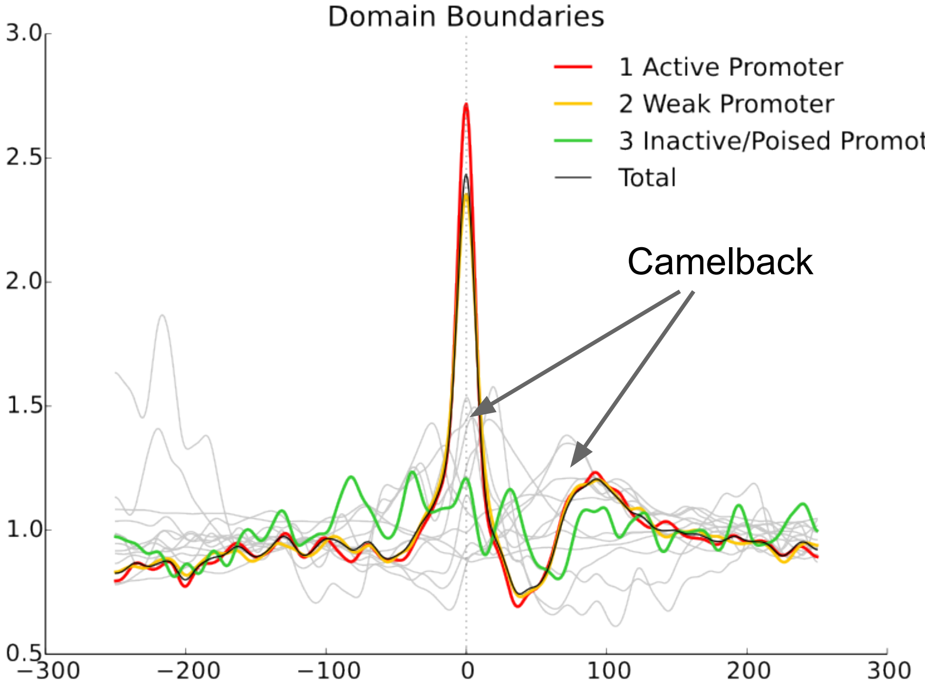
Promoter states

ENCODE promoter states are enriched at loop bases and domain boundaries





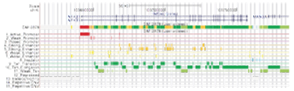
Distance from Loop Base (1kb bins)

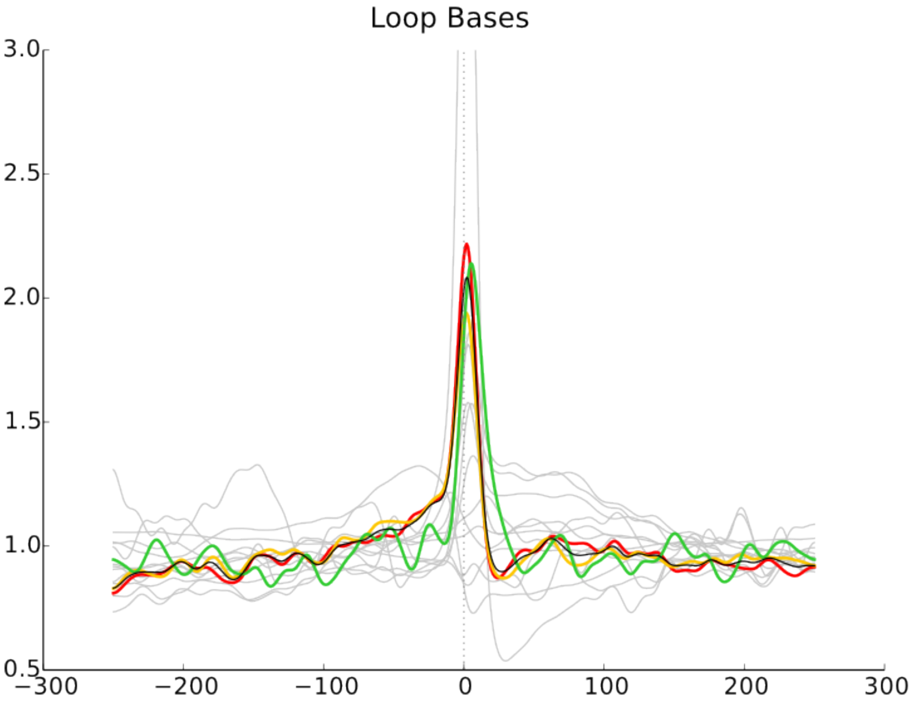


Camelback

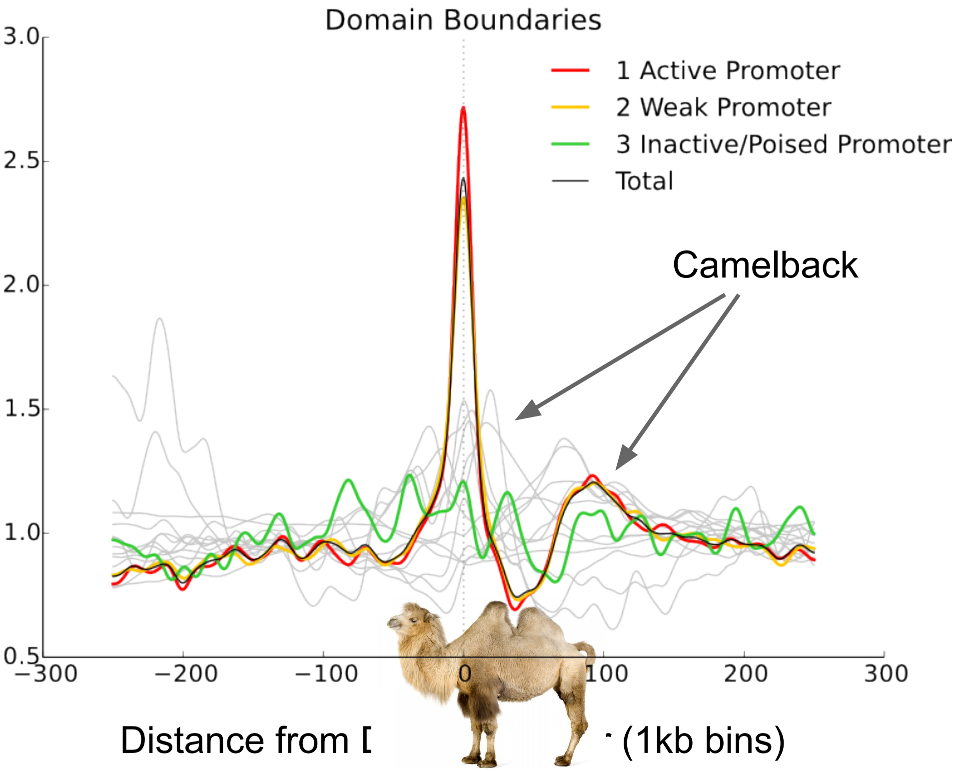
Distance from Domain Border (1kb bins)

ENCODE promoter states are enriched at loop bases and domain boundaries





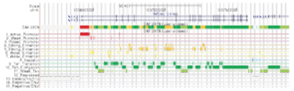
Distance from Loop Base (1kb bins)



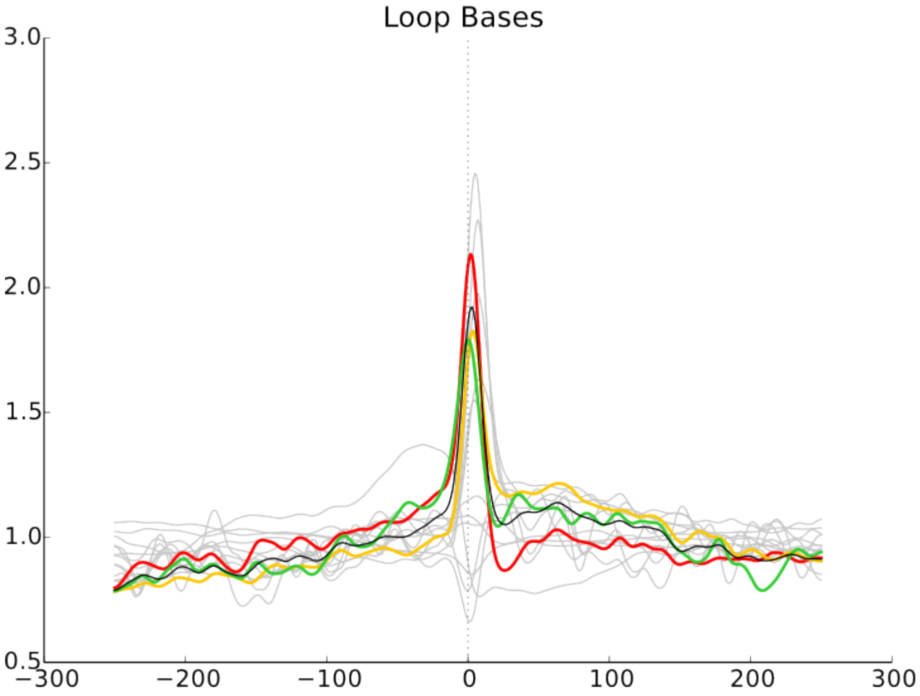
Camelback

Distance from

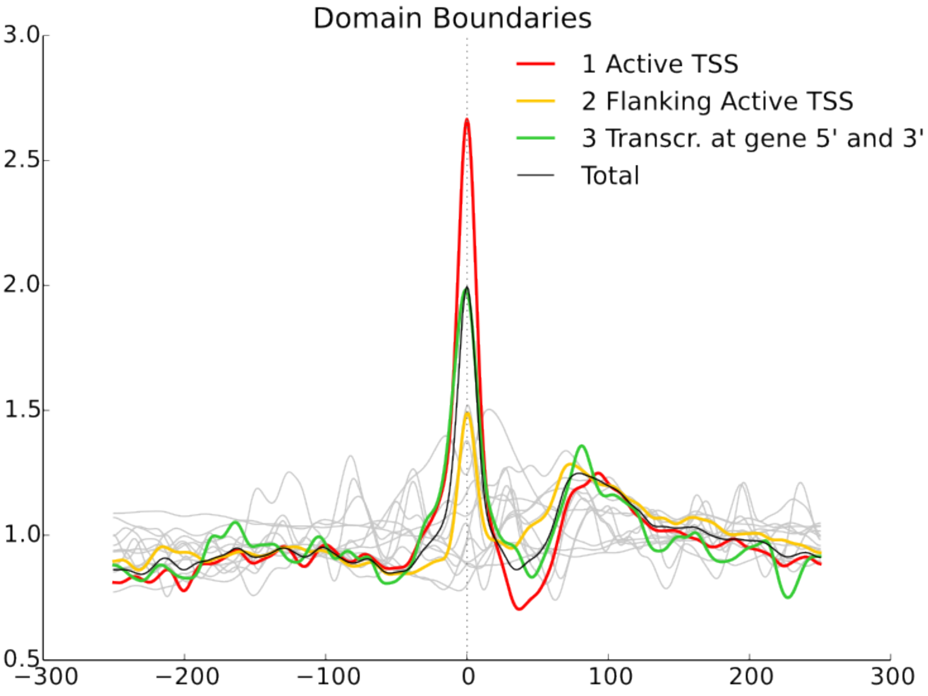
(1kb bins)



Similar behavior for more recent ENCODE Roadmap (2015) TSS states

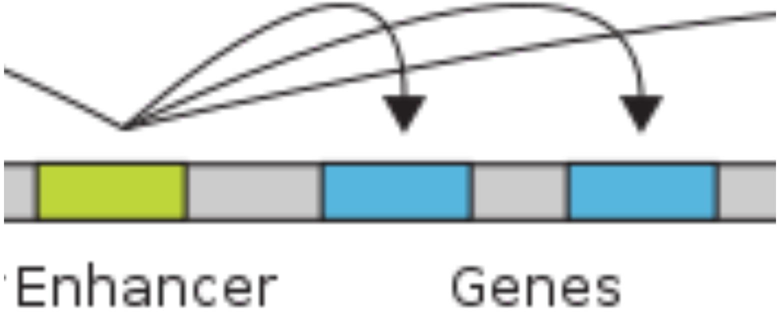


Distance from Loop Base (1kb bins)

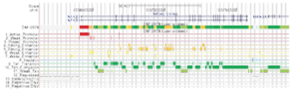


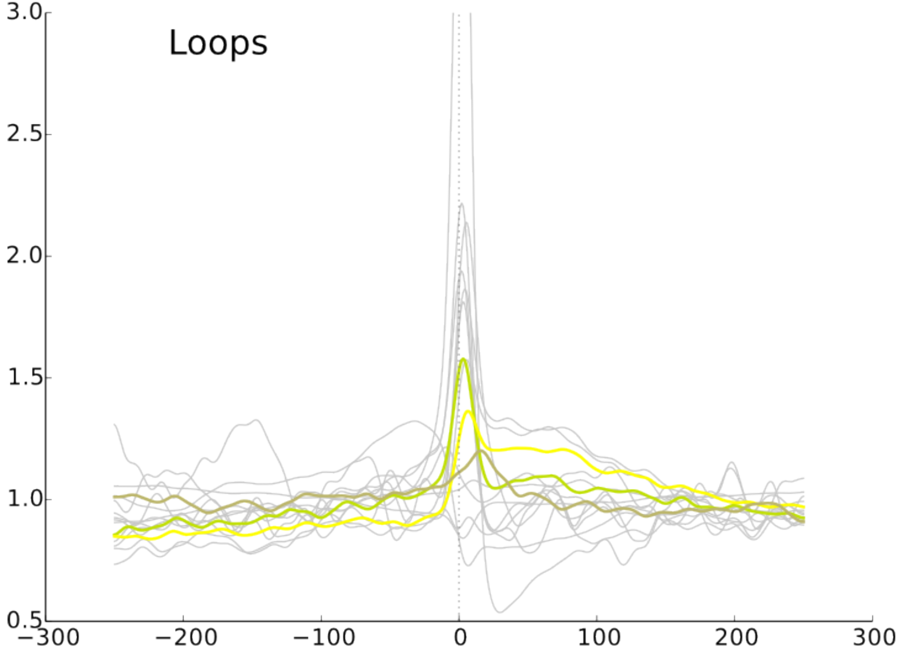
Distance from Domain Border (1kb bins)

Enhancer states

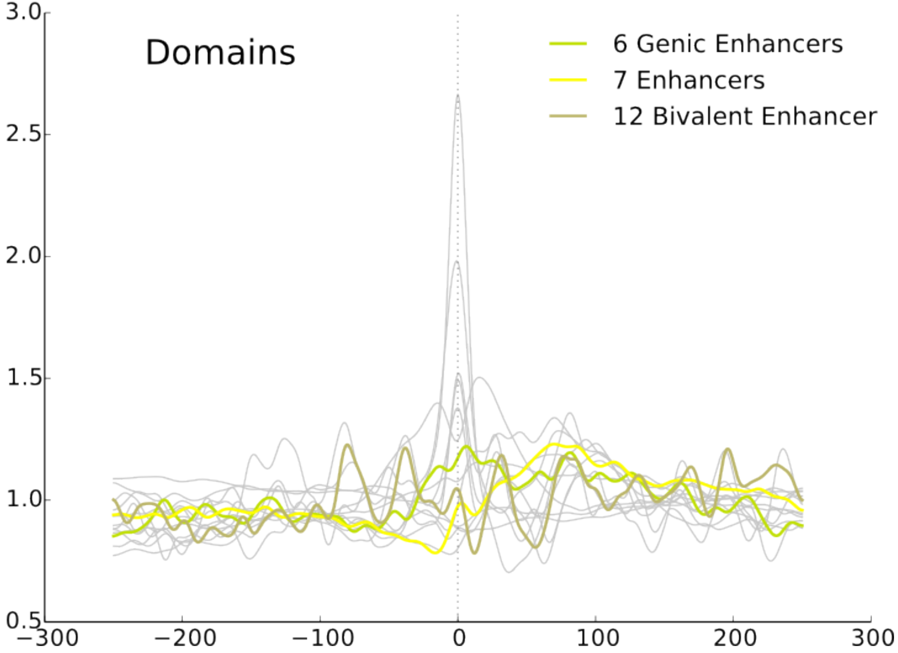


ENCODE enhancer states are enriched at loop bases but not at domain boundaries

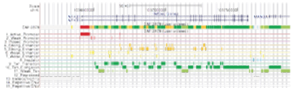




Distance from Loop Base (1kb bins)

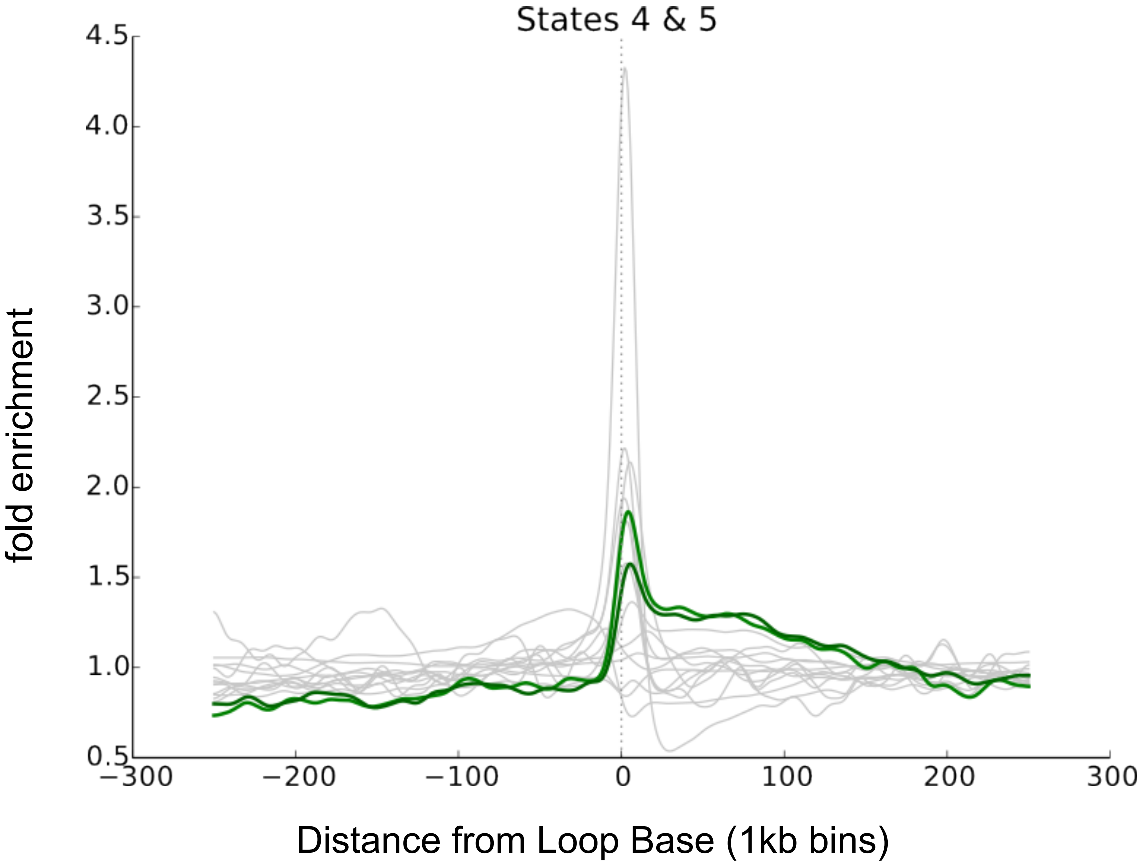


Distance from Domain Border (1kb bins)

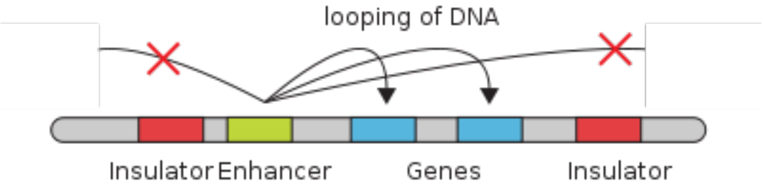


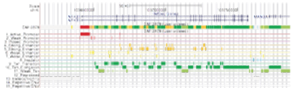
Strong Enhancers

- Moderate peak at domain boundaries
- Stay relatively enriched for a short distance within the loop bases

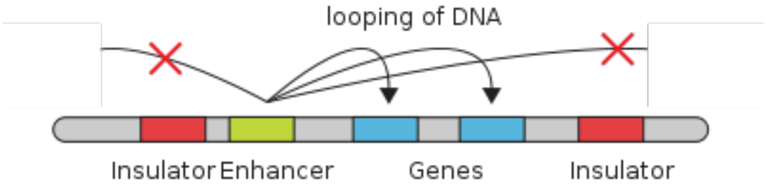


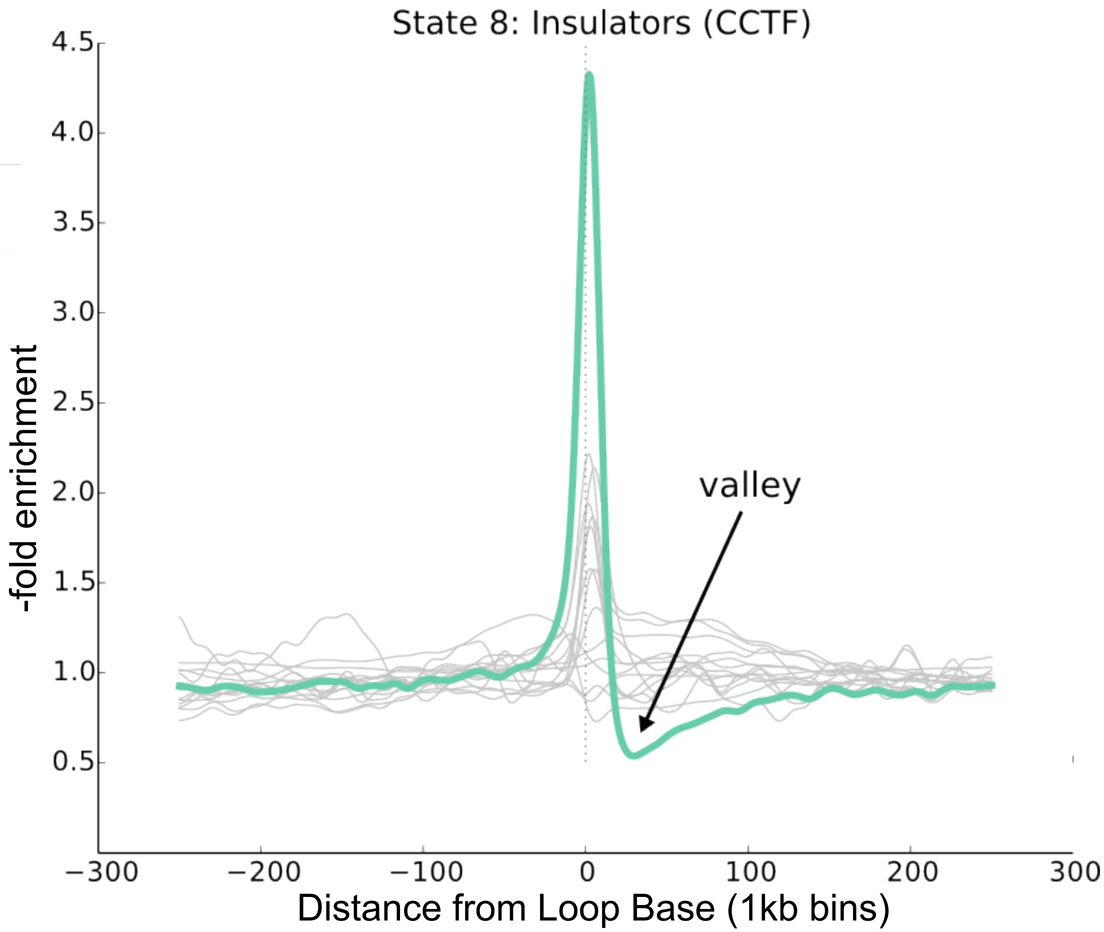
Insulator states

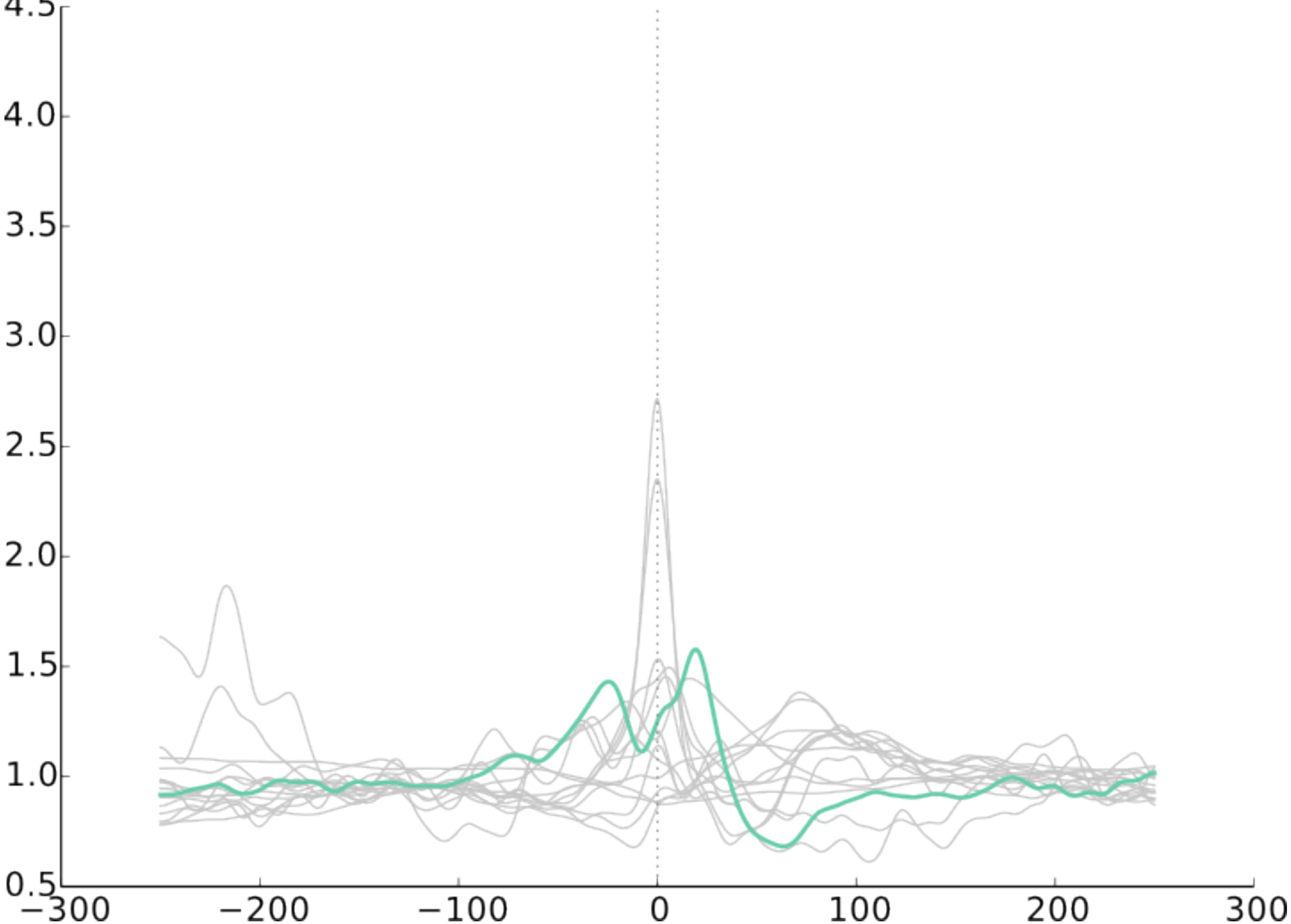




Insulators block interactions between genomic elements

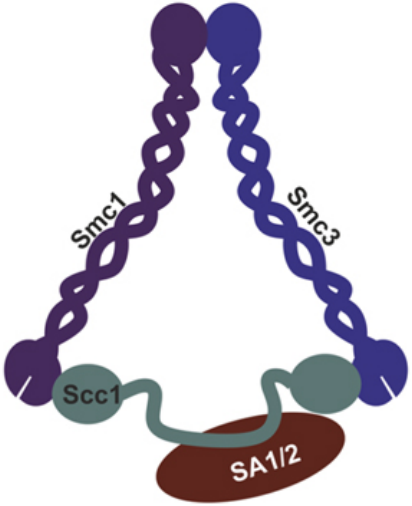



State 8: Insulators (CCTF)

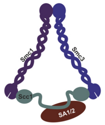


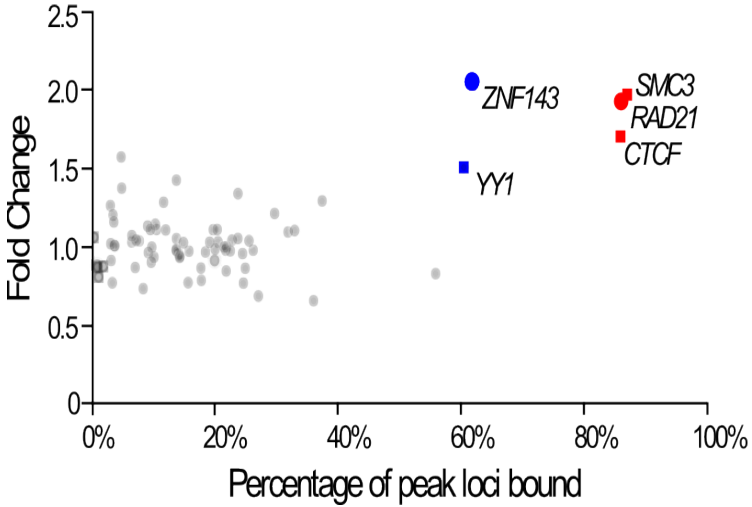
ENCODE Insulators at Domain Boundaries

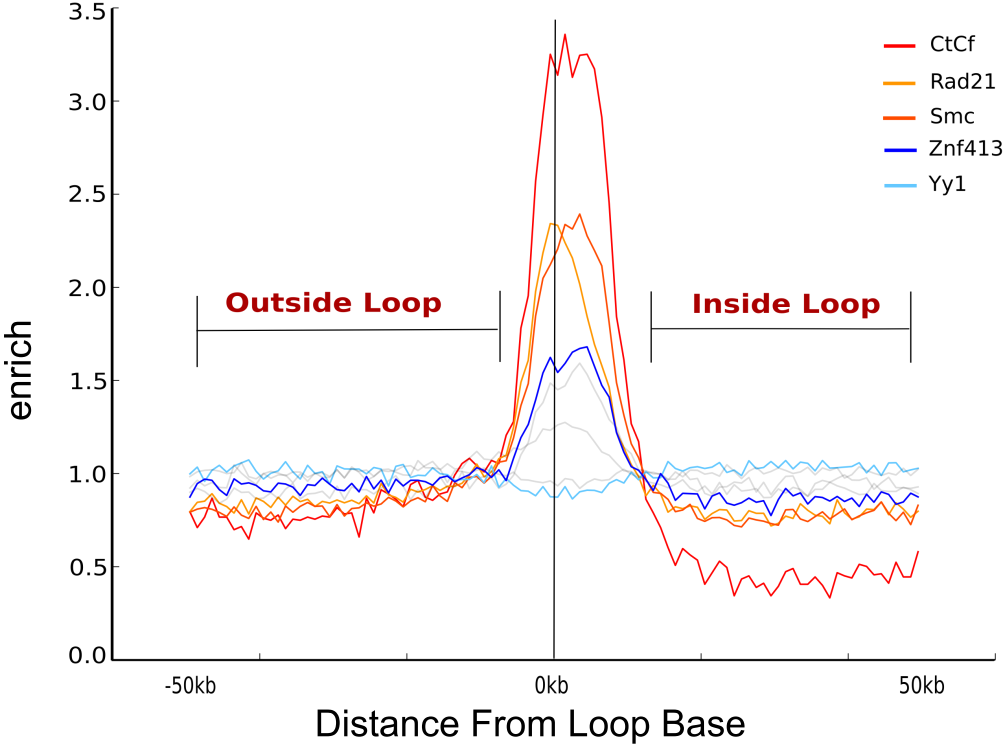
Architectural Proteins



Protein Enrichment around Loop Bases

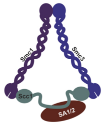


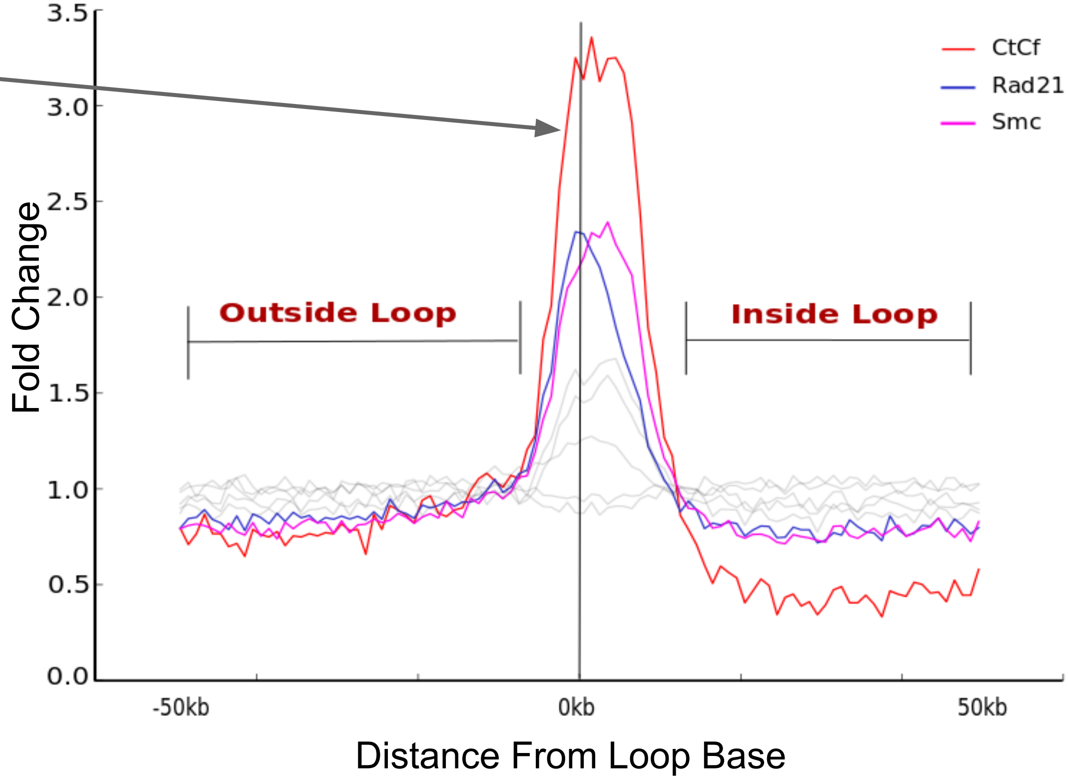




Rao *et al.* 2014

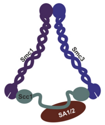
Protein Binding Frequencies

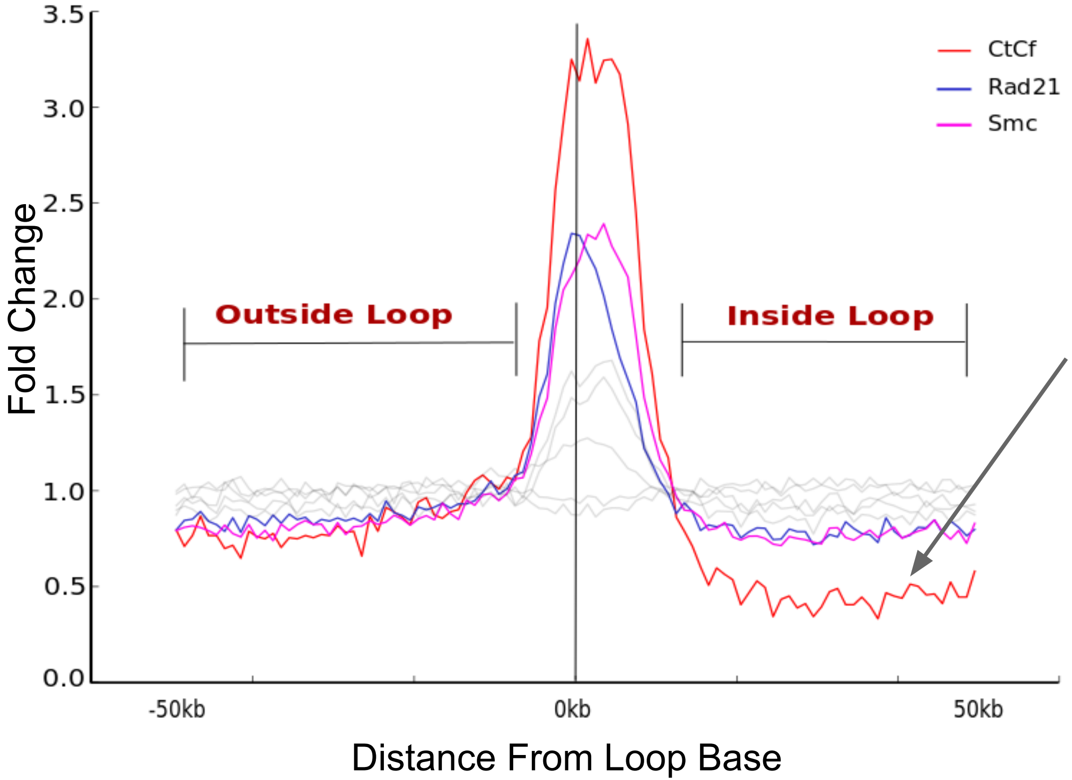




Strong CtCf Peak

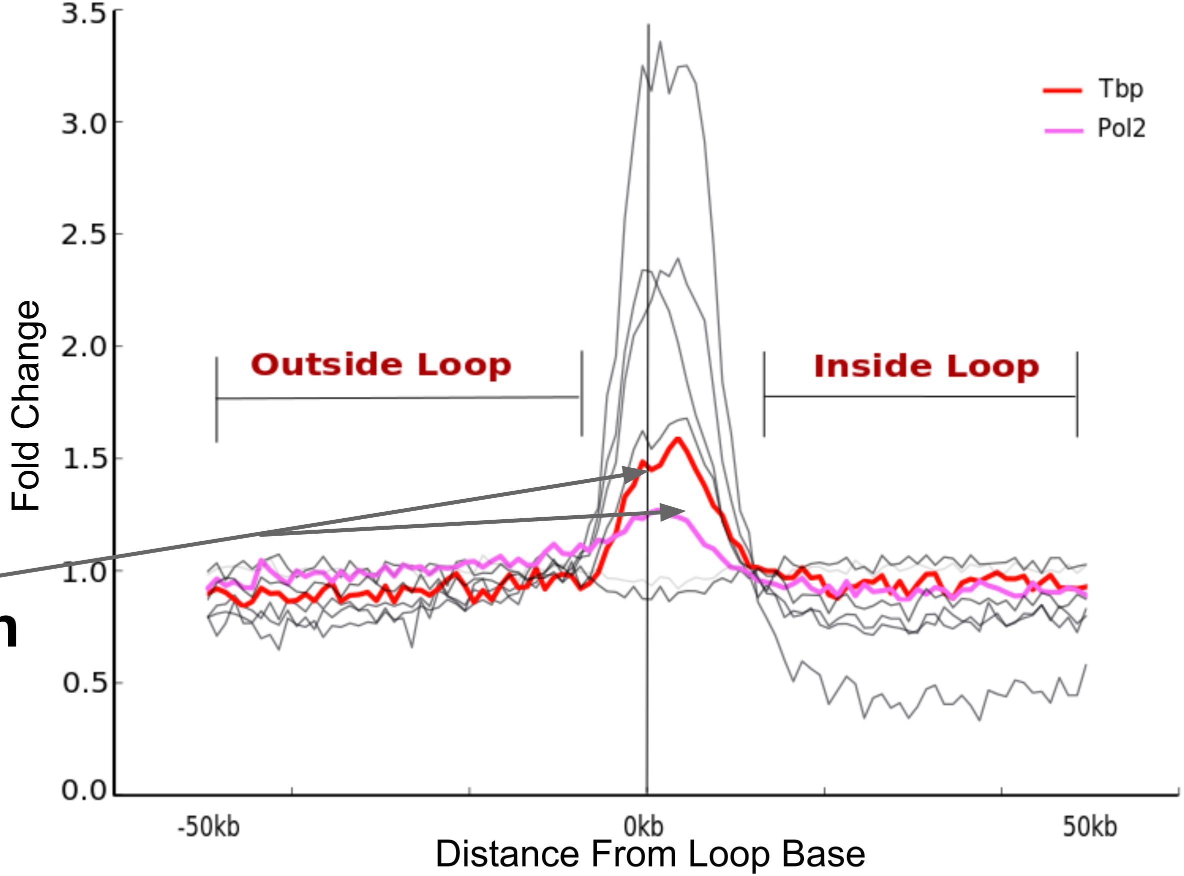
Protein Binding Frequencies





Depleted CtCf
Signal Inside Loop

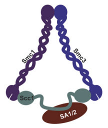
Tbp and Pol2 Enrichment at Loops

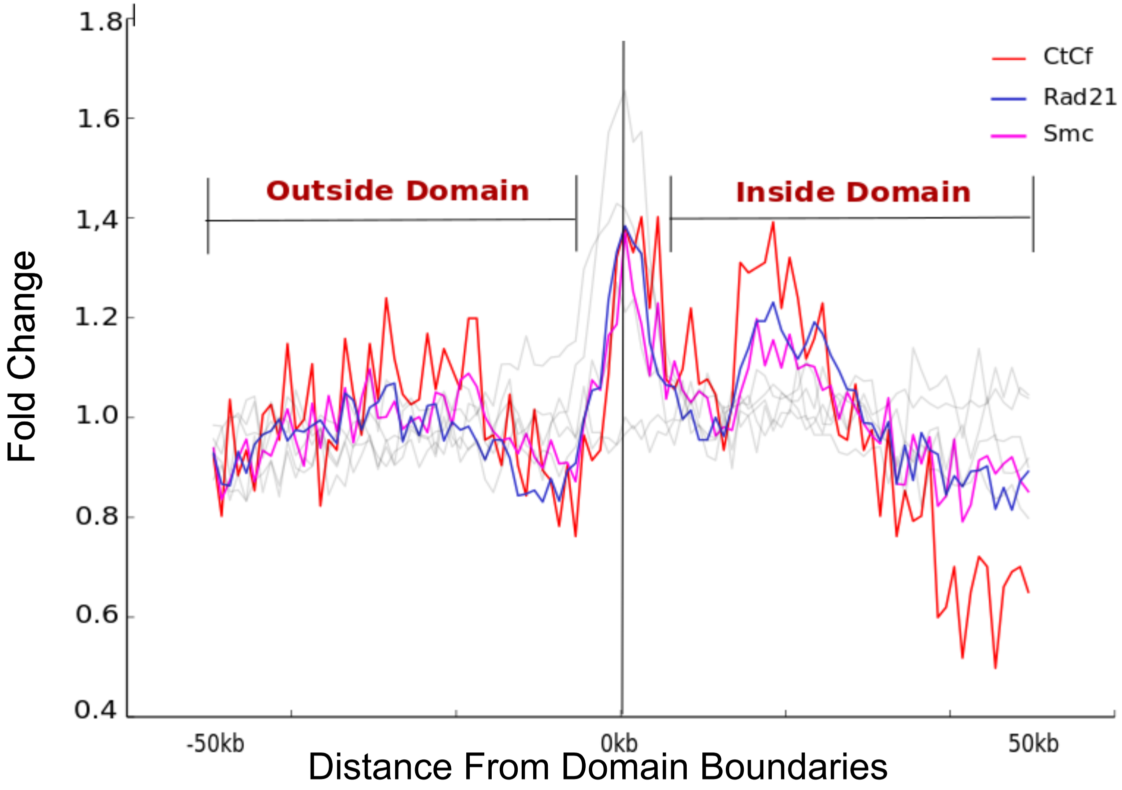


Transcription Related Proteins

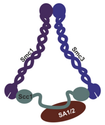
Architectural Proteins:

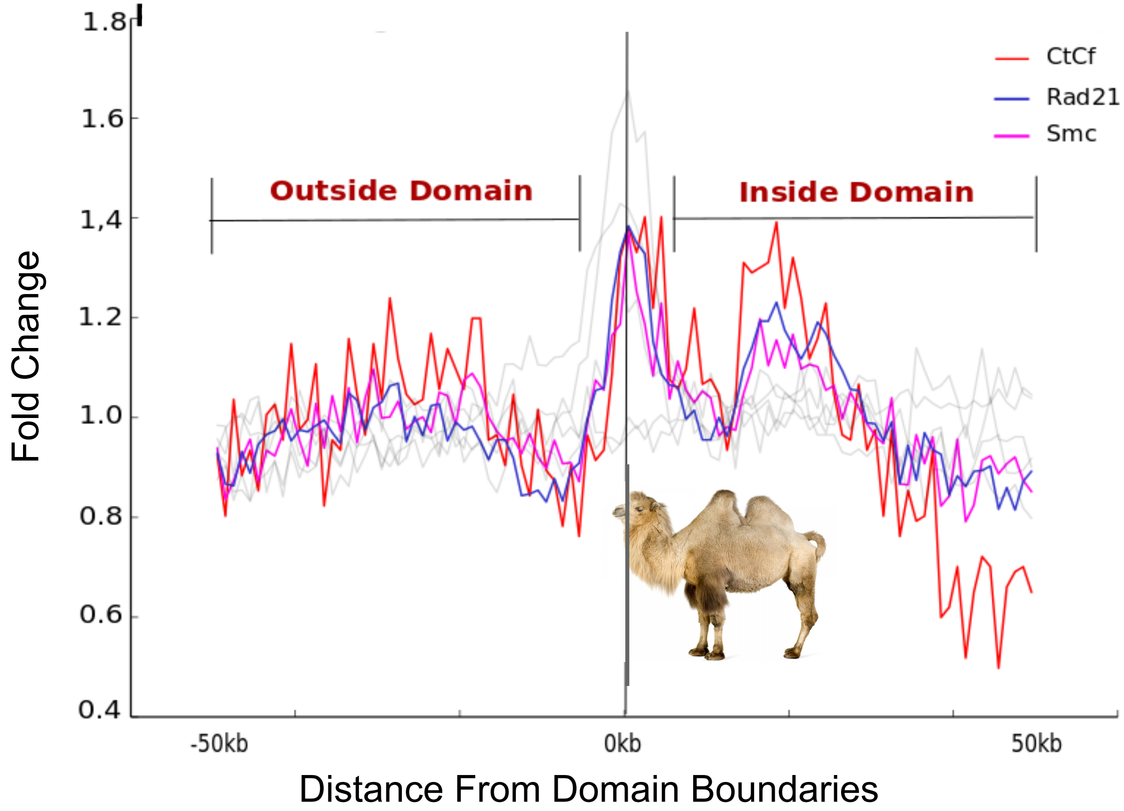
Not as enriched at domain boundaries



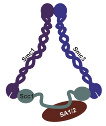


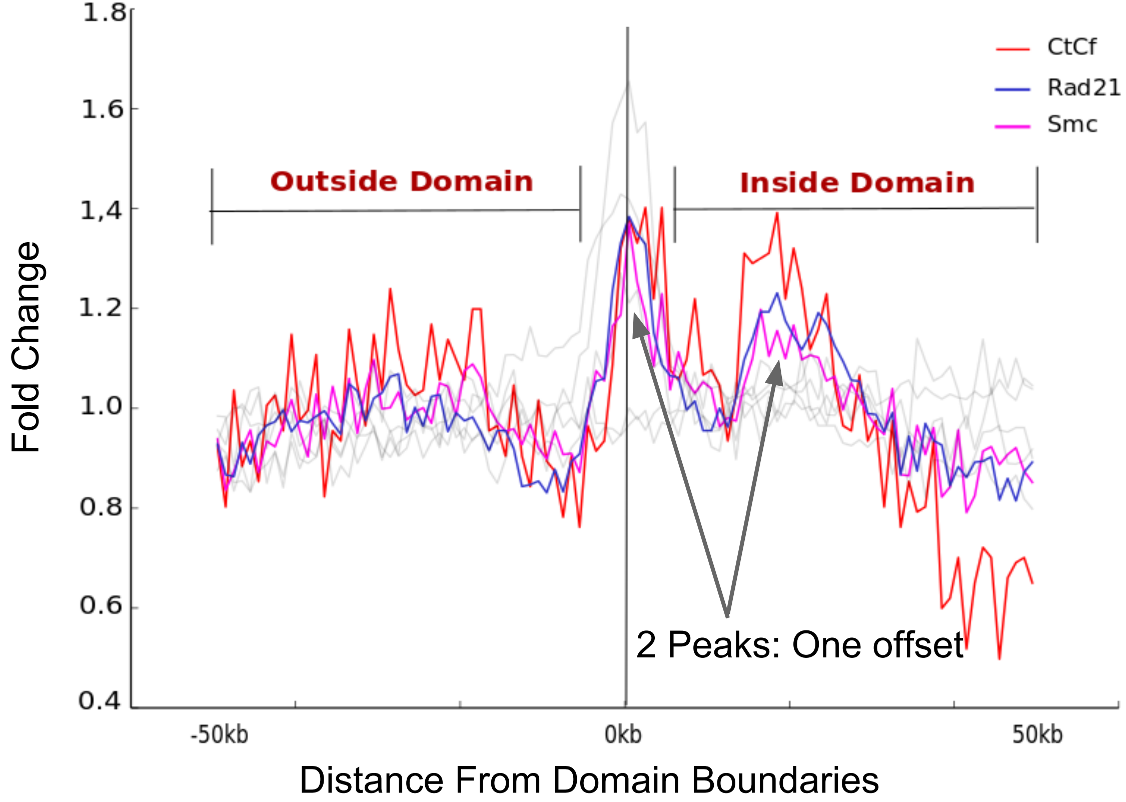
Architectural Proteins
have a “Camel Hump” Enrichment inside of domains





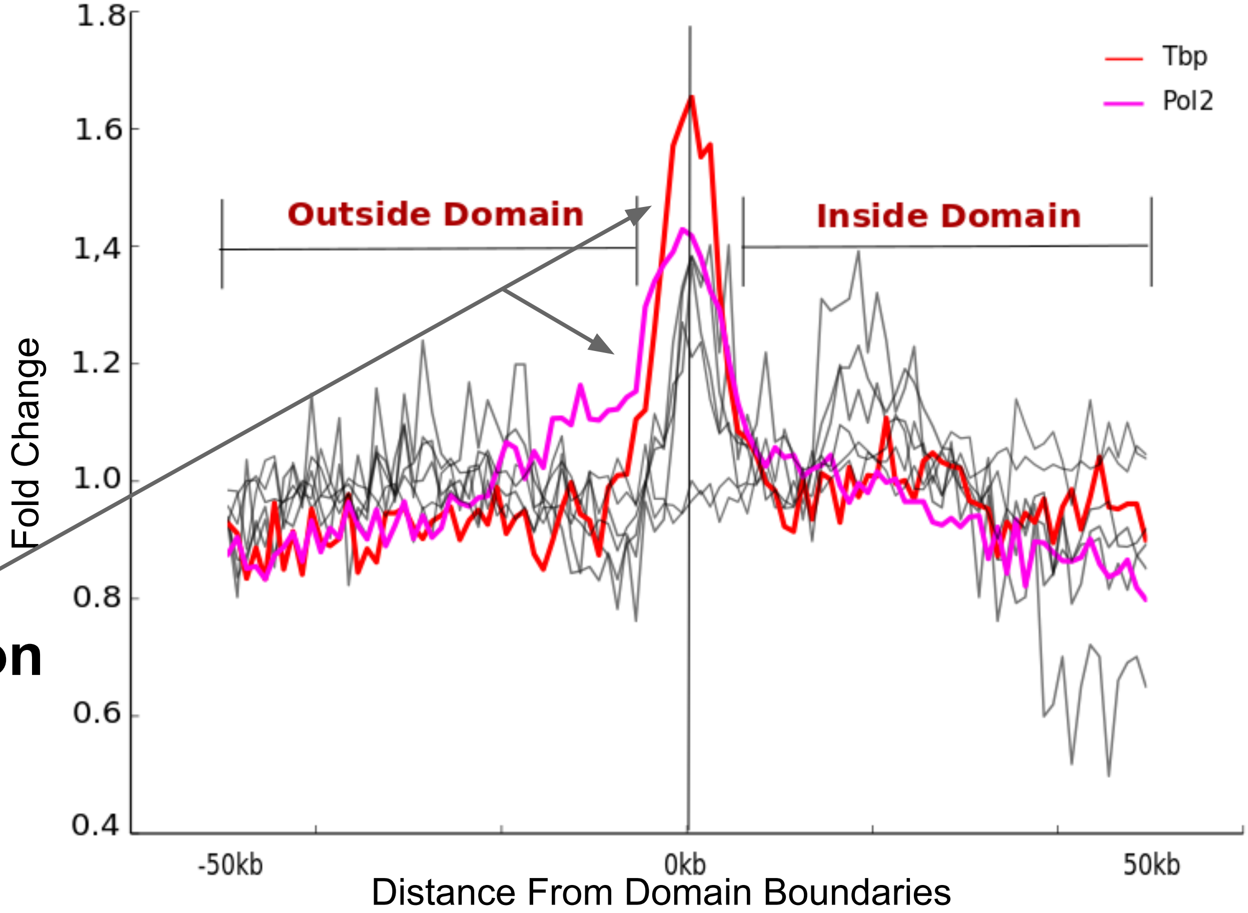
Architectural Proteins have a “Camel Hump” Enrichment inside of domains





2 Peaks: One offset

Domain Enrichment



Transcription Related Proteins

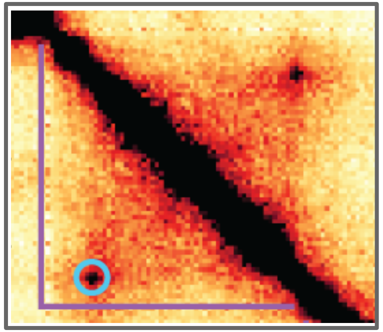
Summary

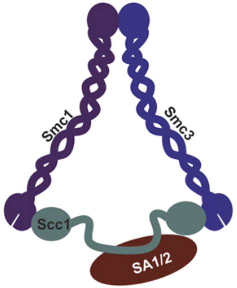
- Promoter chromatin state and TSS are enriched at both domains and loops, but slightly more enriched at domains
- Transcription Proteins: strongest enrichment at domain boundaries
- Insulator chromatin state highly enriched at loops, not domains
- Enhancer state is moderately enriched at loops, not domains
- CTCF, Smc3, Rad21 as Architectural Proteins for loops, not domains

Conclusion:

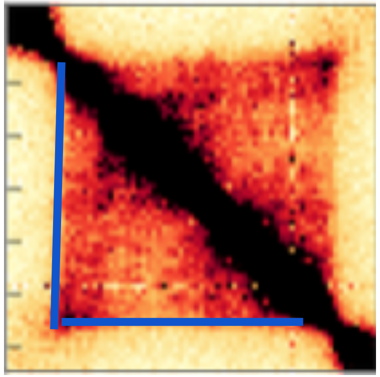
Different mechanisms underlie domain and loop formation

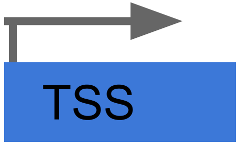
- Loops organized by architectural proteins?





- Domains organized by transcription?



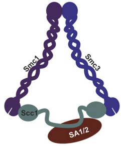


Future Research

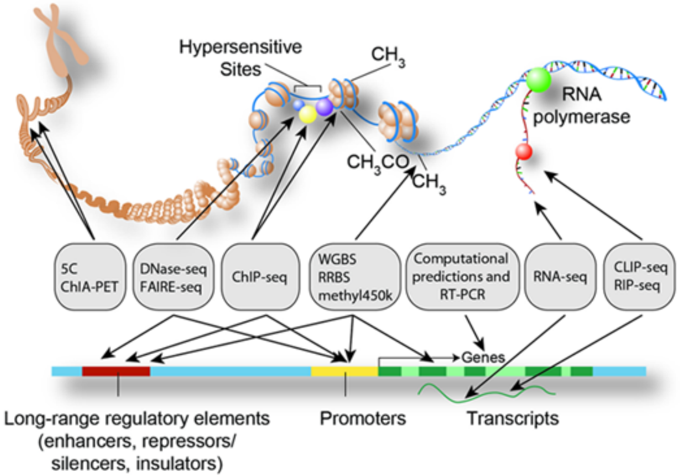
- Explore profiles of other specific factors
- Investigate “valley” and “camelback” profiles
- Look into CTCF motif orientation at loop bases

Acknowledgements / Thank You

Mirny lab:
prof. Leonid Mirny
dr. Geoff Fudenberg
Maxim Imakaev
Boryana Doyle
Nezar Abdennur
Anton Goloborodko



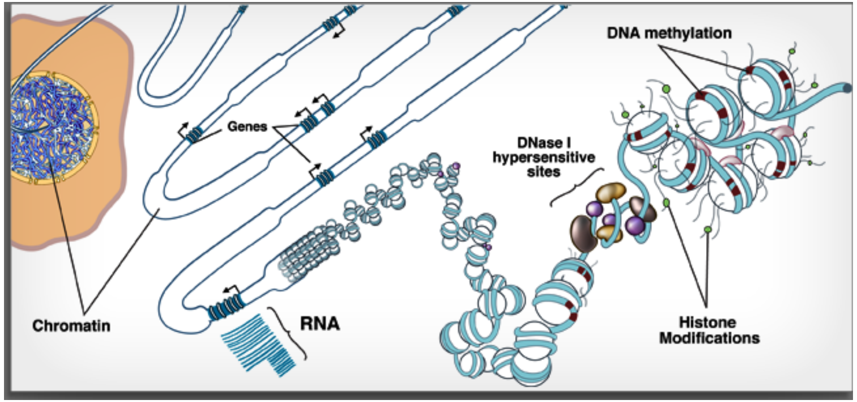
ENCODE (2012)



https://www.encodeproject.org/

Goal: Develop a parts list of **functional elements** in the human genome

ROADMAP (2015)

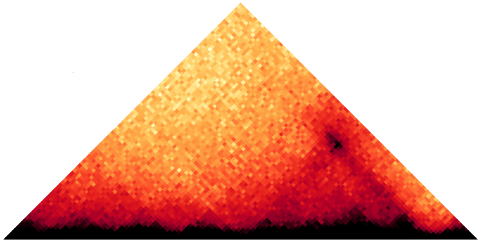


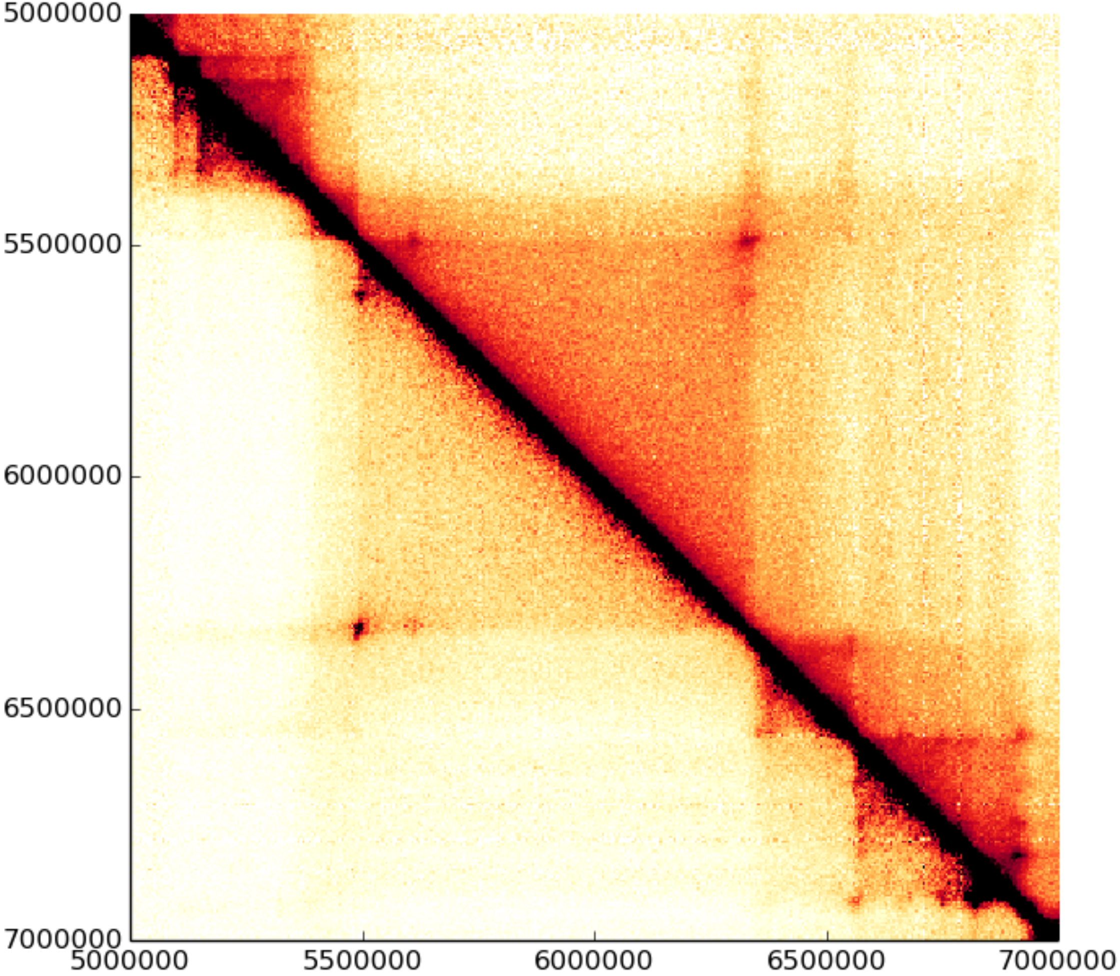
http://www.roadmapepigenomics.org/

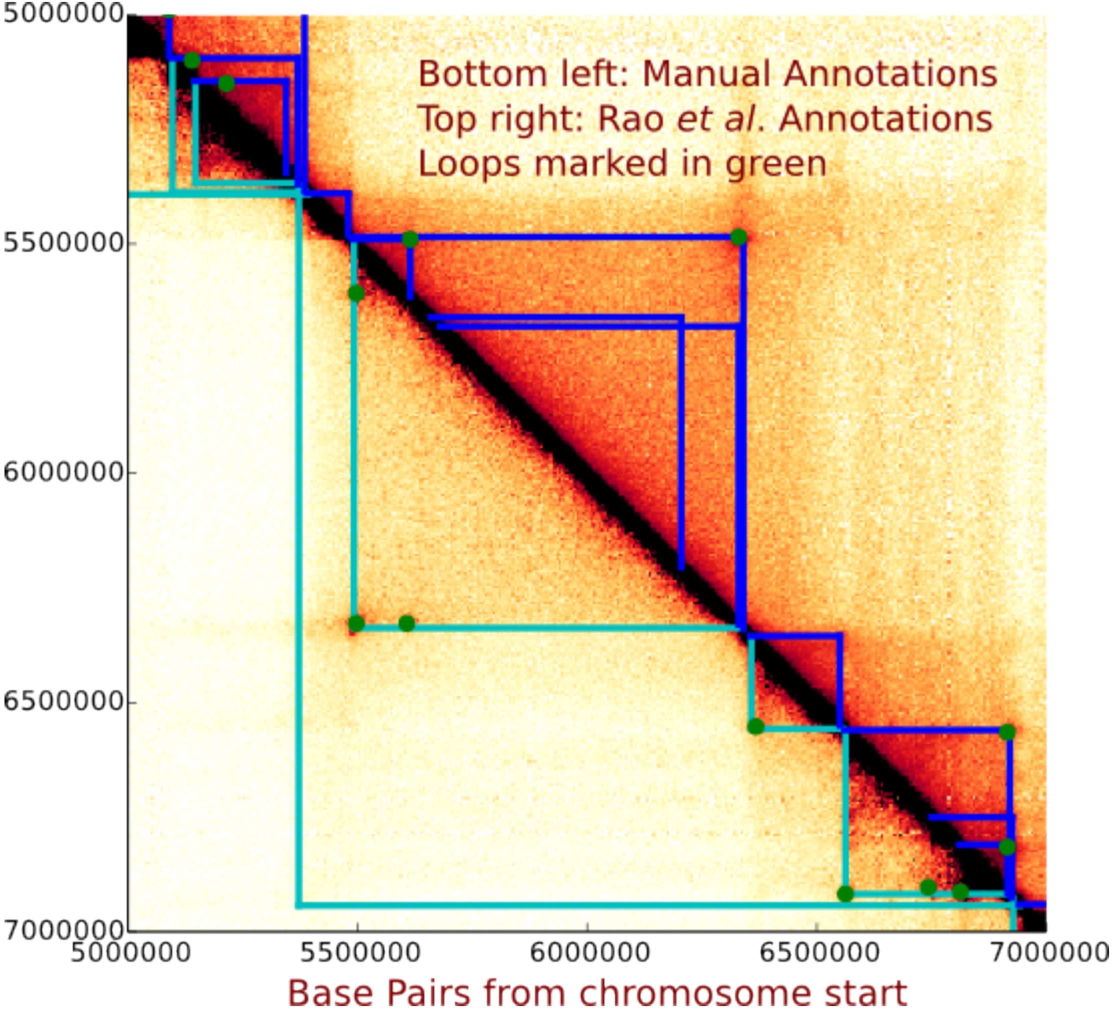
Goal: Systematically characterize functional elements in many **primary human tissues and cells**

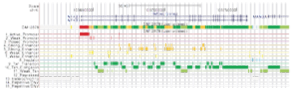
Builds on the work of ENCODE, but covers a large set of healthy cell types to serve as “reference epigenomes”.

Manual Annotations

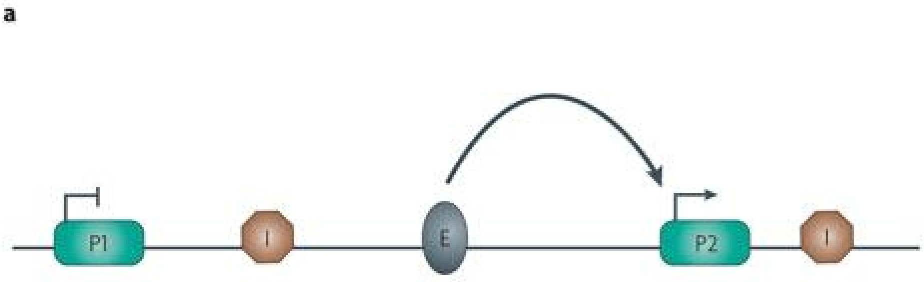


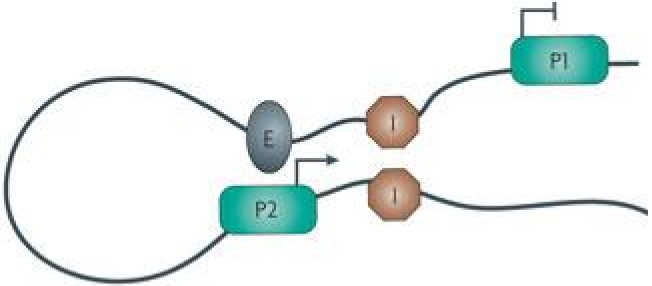

5000000
5500000
6000000
6500000
7000000
5000000
5500000
6000000
6500000
7000000

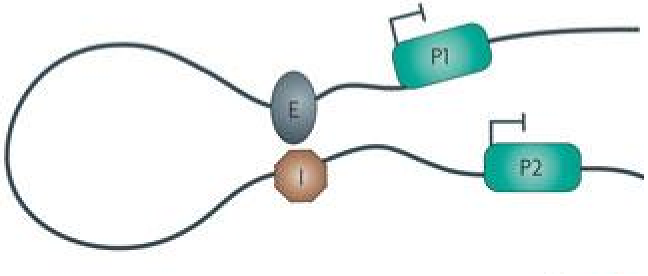

Bottom left: Manual Annotations
Top right: Rao *et al*. Annotations
Loops marked in green
5000000
5500000
6000000
6500000
7000000
5000000
5500000
6000000
6500000
7000000
Base Pairs from chromosome start



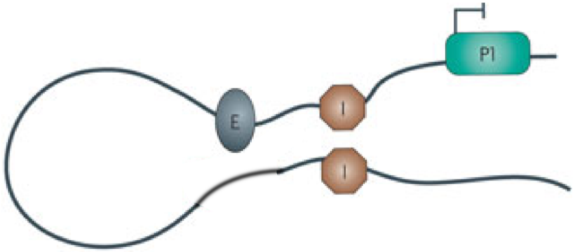
- Enhancers/promoters generally enriched inside domain borders/loop
- Insulators depleted just inside of boundary





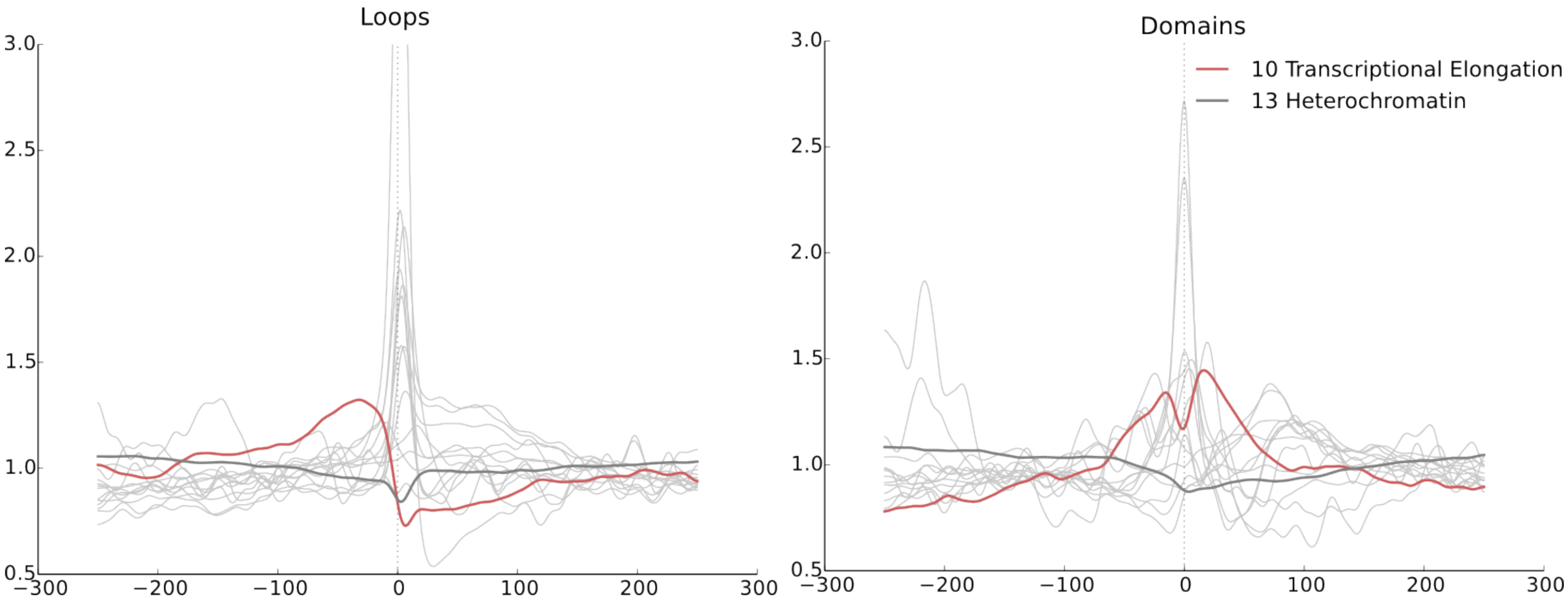


Insulators could possibly bring together elements, or isolate them



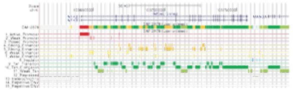
Raab and Kamakaka, 2010

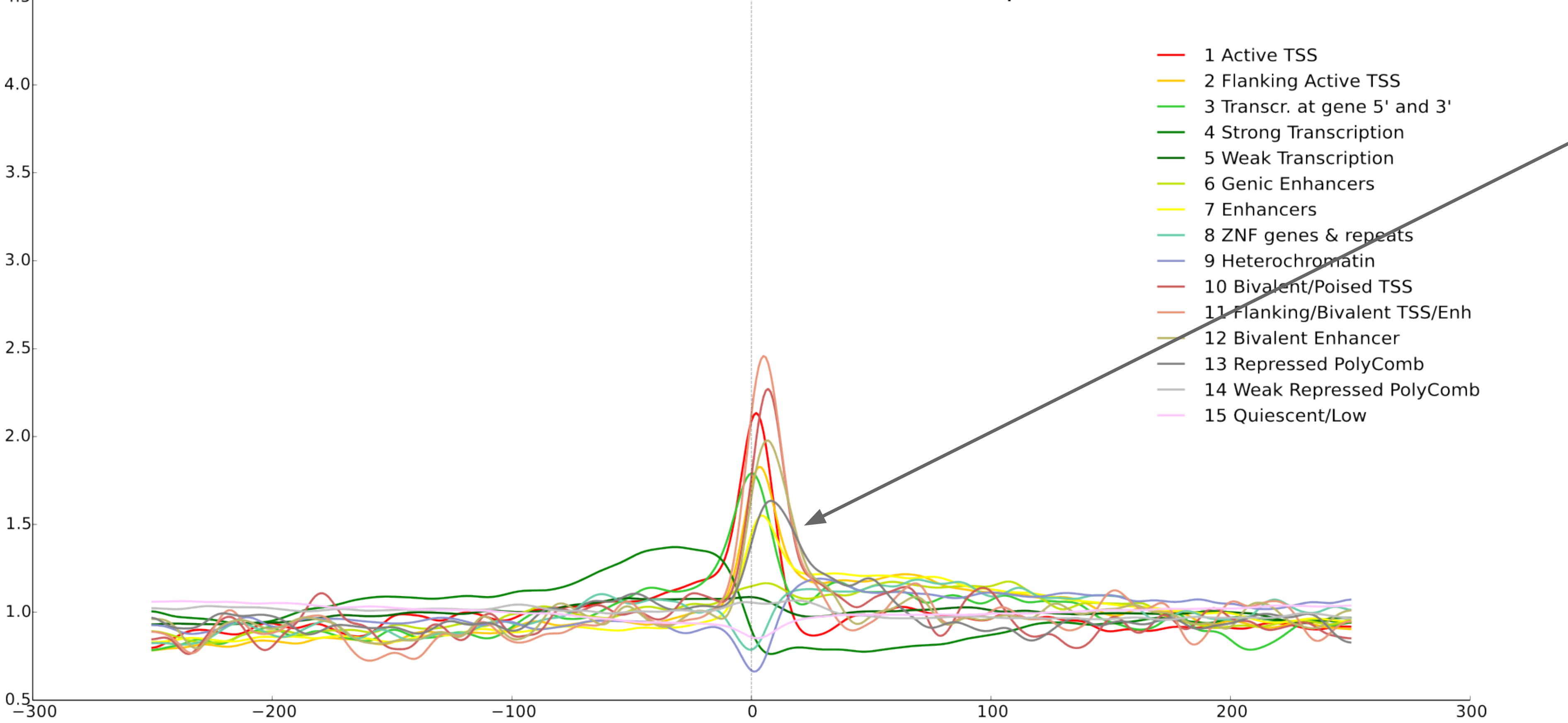
Some states are depleted at domain/loop boundaries

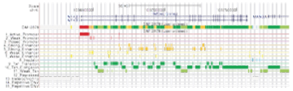


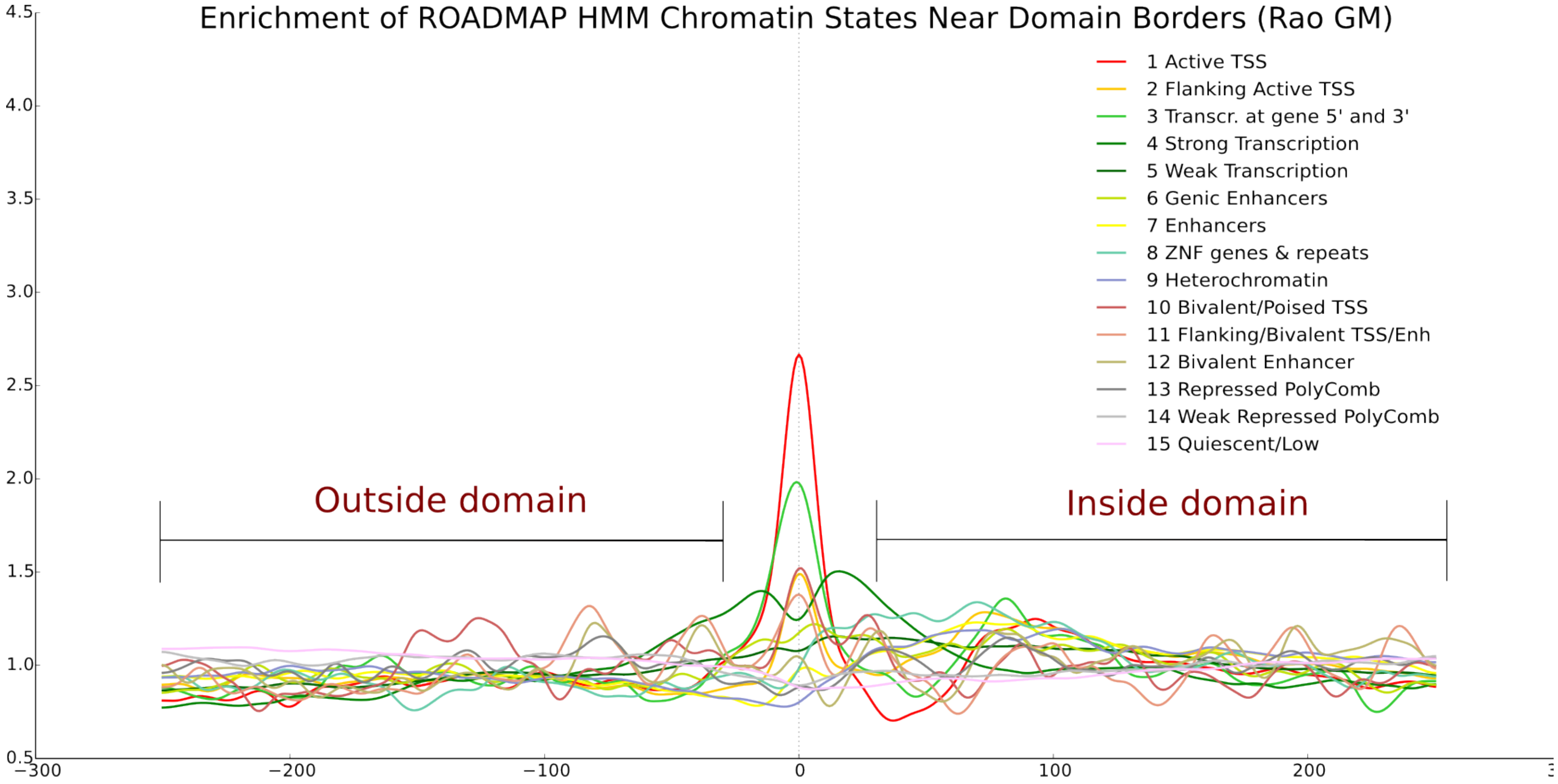
ROADMAP

Enrichment of ROADMAP HMM Chromatin States Near Loop Bases (Rao GM)








Enrichment of ROADMAP HMM Chromatin States Near Domain Borders (Rao GM)
1 Active TSS
2 Flanking Active TSS
3 Transcr. at gene 5' and 3'
4 Strong Transcription
5 Weak Transcription
6 Genic Enhancers
7 Enhancers
8 ZNF genes & repeats
9 Heterochromatin
10 Bivalent/Poised TSS
11 Flanking/Bivalent TSS/Enh
12 Bivalent Enhancer
13 Repressed PolyComb
14 Weak Repressed PolyComb
15 Quiescent/Low
Outside domain
Inside domain
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
−300
−200
−100
0
100
200