Organization and Regulation of Human Genome

June 8-10, 2016
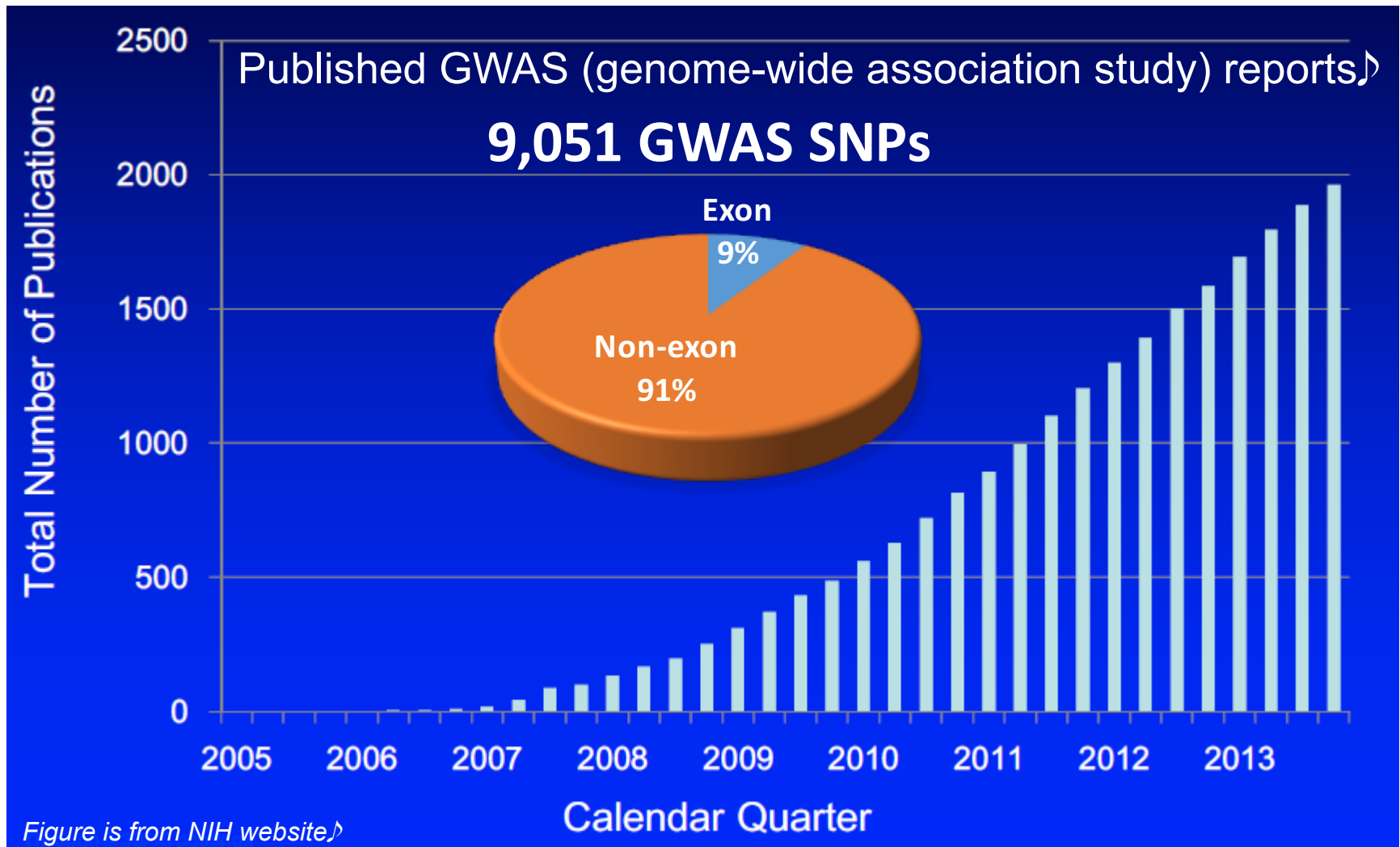
ENCODE 2016 Meeting
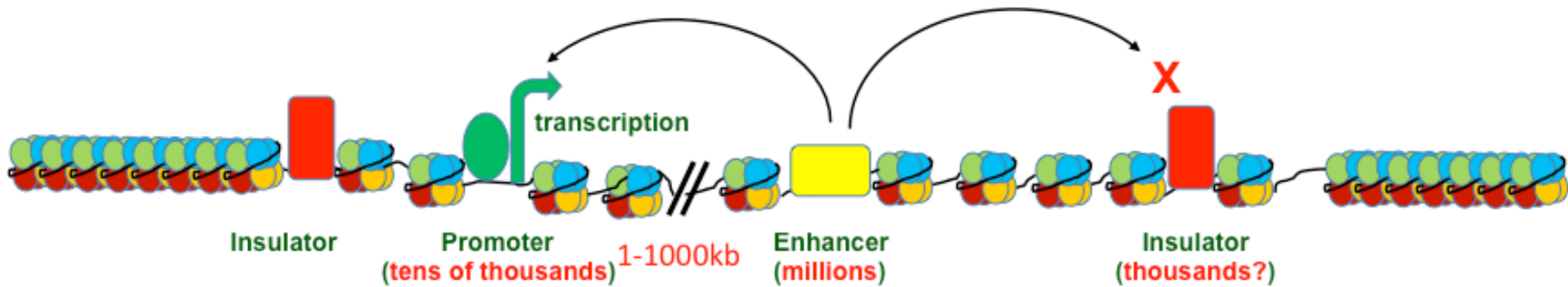
**LUDWIG CANCER RESEARCH**

Bing Ren, Ph.D.

UCSD

The Genetic Variants Associated with Human Diseases Are Mostly Non-coding

- A total of ~13% of the human genome likely devoted to enhancer function, while ~1% to promoter function, in at least one of the 111 cell/tissue types investigated. Approximately 5% of each cell type's genome is marked by signatures associated with cis regulatory elements signatures.

- Enhancers with coordinated activity patterns across tissues are enriched for common gene functions and human phenotypes, suggesting that they represent coordinately regulated modules.

- Regulatory motifs are enriched in tissue-specific enhancers, enhancer modules and DNA accessibility footprints, providing an important resource for gene-regulatory studies.

- Genetic variants associated with diverse traits show enrichments of biochemical signatures in trait-relevant tissues, providing an important resource for understanding the molecular basis of human disease.
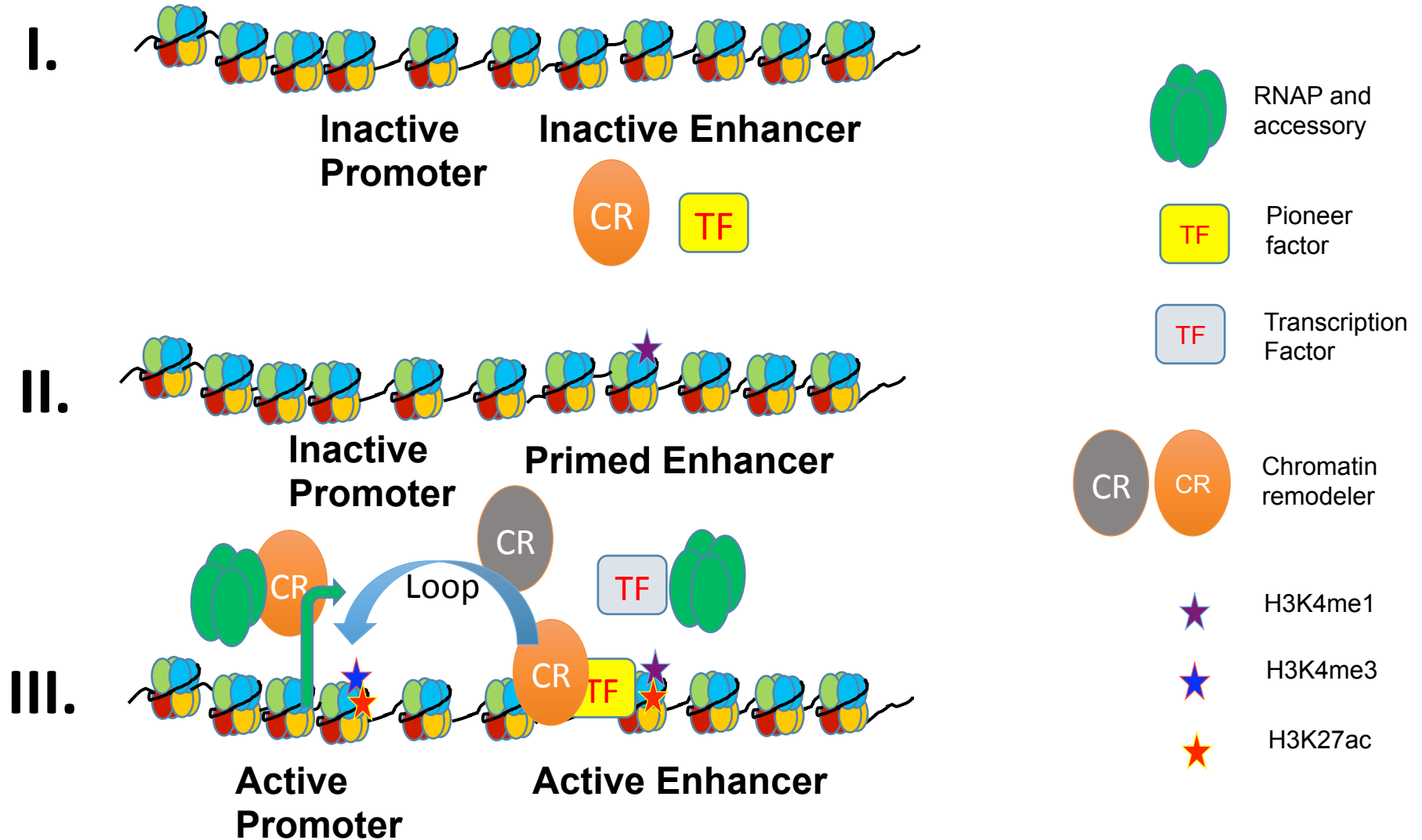
Questions left unanswered
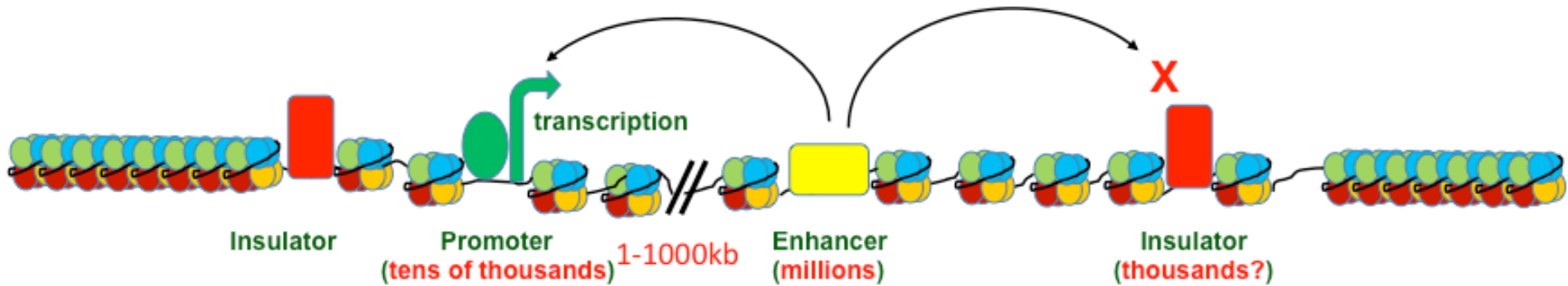
- What mechanisms allow enhancers to target specific genes from a distance?

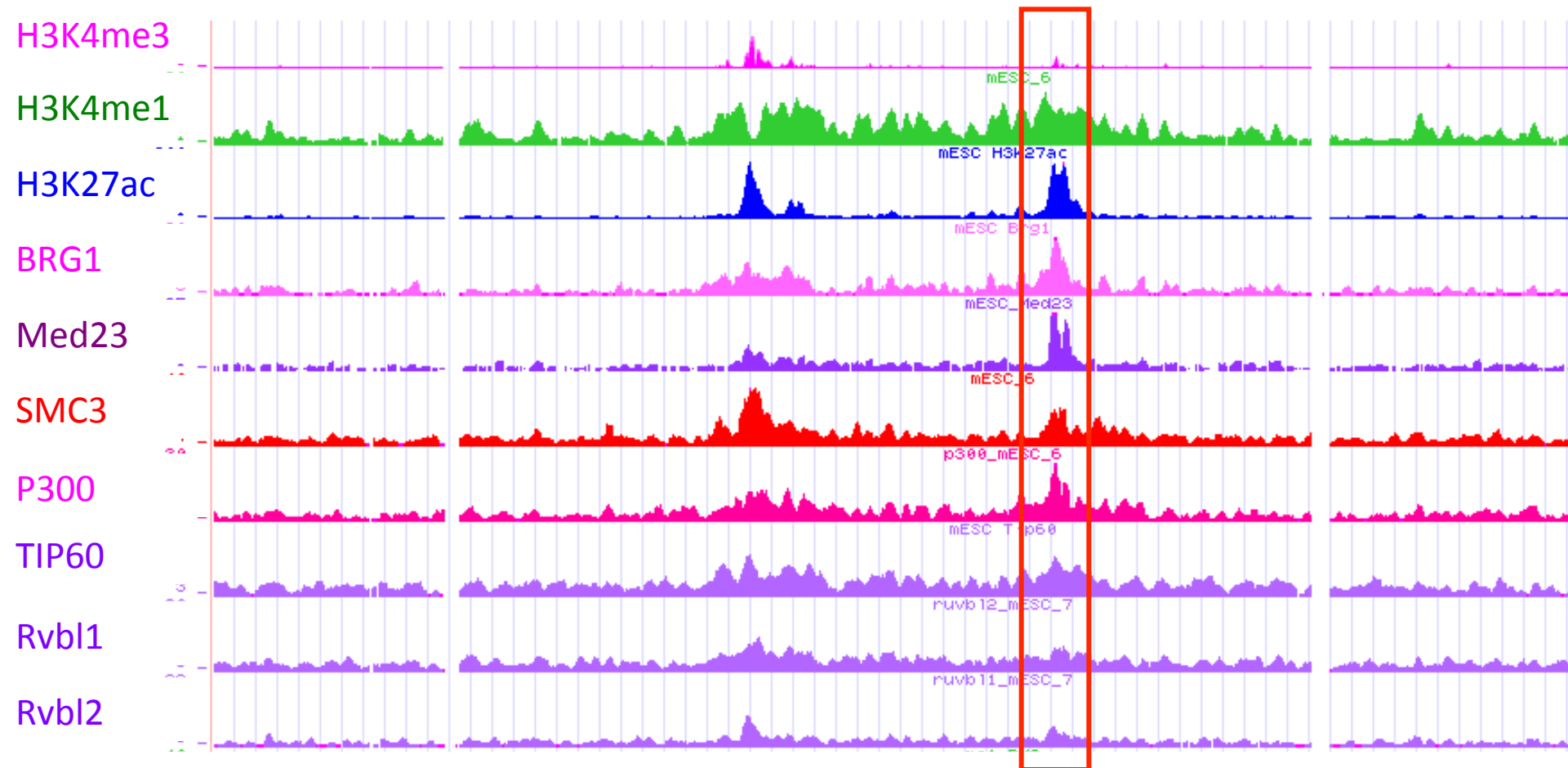- What target genes do the enhancers have in different cell types?
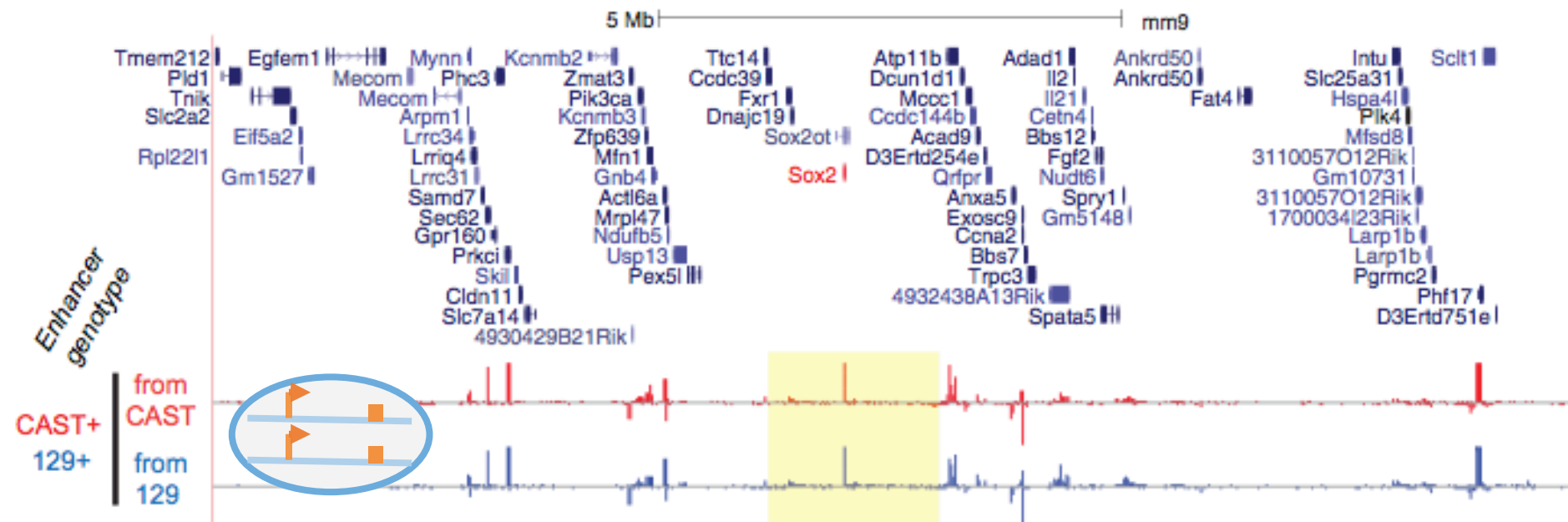
A model for step-wise activation of enhancers

Reviewed by Zaret & Carroll, Genes & Dev 2011;
Calo & Wysocka, Mol Cell 2013; Ren and Yue, 2015

Example: chromatin contact between Sox2 and its enhancer
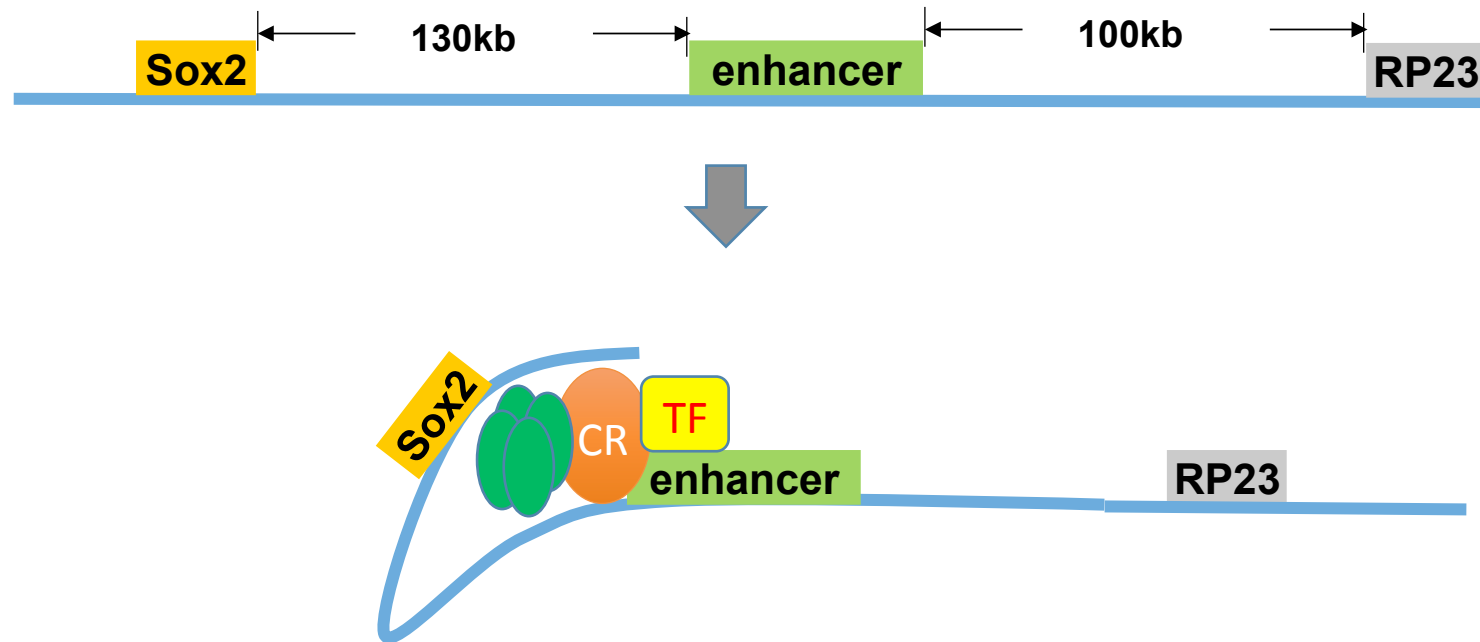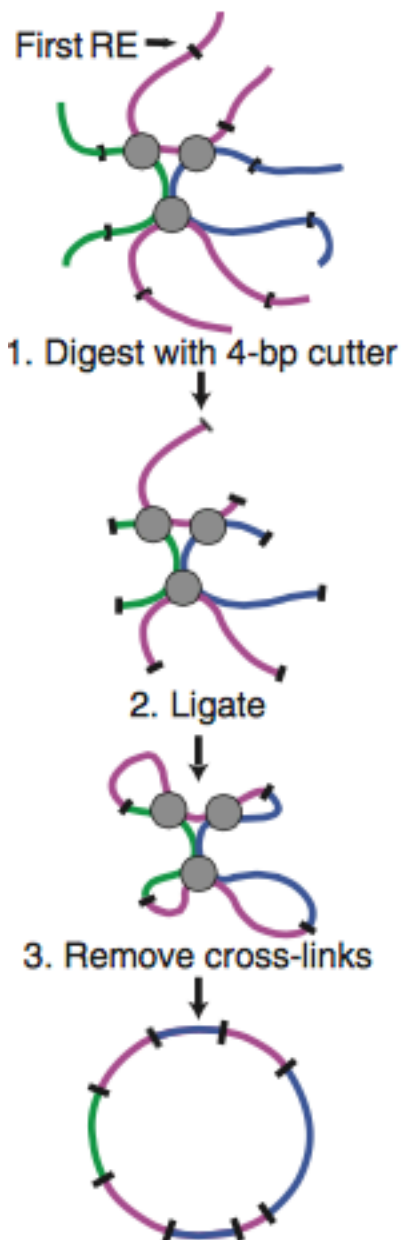
Allelic RNA-seq analysis identified the Sox2 enhancer

Chromatin interactions play a role in regulation of Sox2 gene expression by a distal enhancer
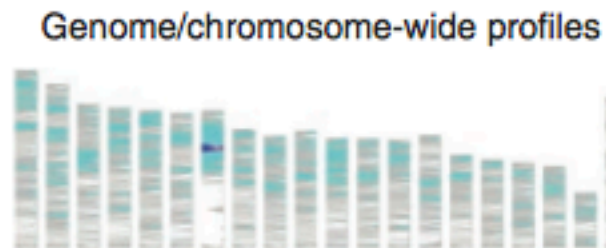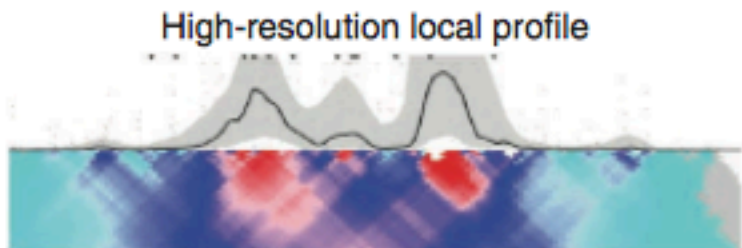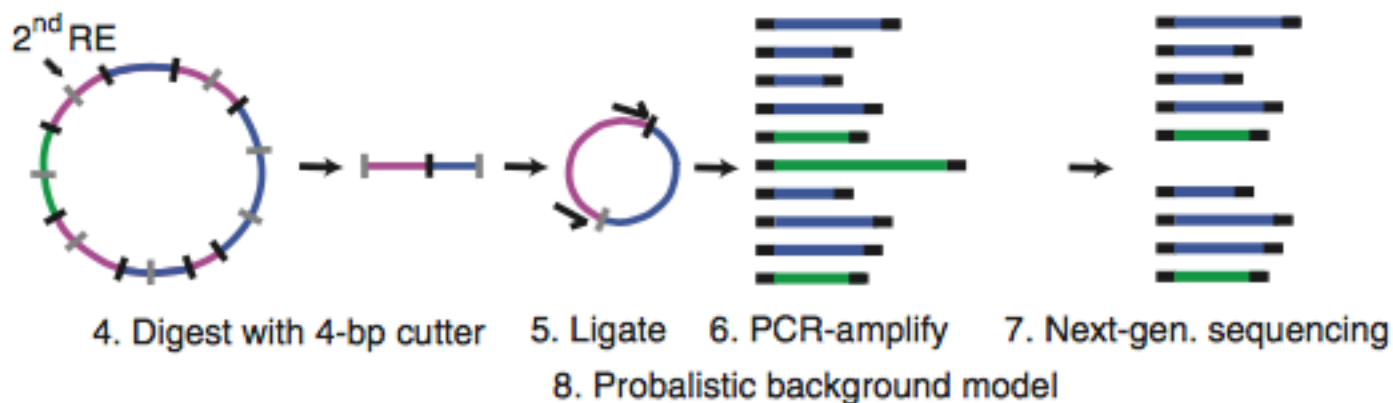
Investigating long-range chromatin interactions using 4C-seq

Simonis, M et al. et al Nature Genetics 2006
van de Werken et al. Nature Method 2012

First RE →

1. Digest with 4-bp cutter

2. Ligate

3. Remove cross-links

High-resolution 4C-seq: PCR amplification after two rounds of digestion-ligation

2nd RE

4. Digest with 4-bp cutter   5. Ligate   6. PCR-amplify   7. Next-gen. sequencing

8. Probalistic background model

High-resolution local profile   Genome/chromosome-wide profiles

Long-range chromatin interaction of Sox2 enhancer/ promoter interactions validated by 4C-seq
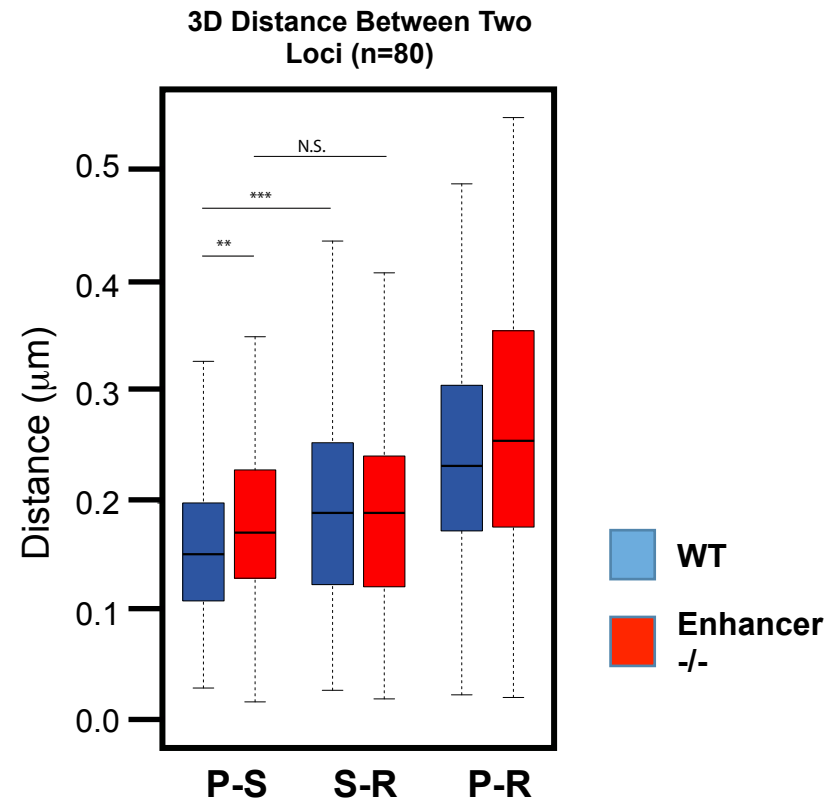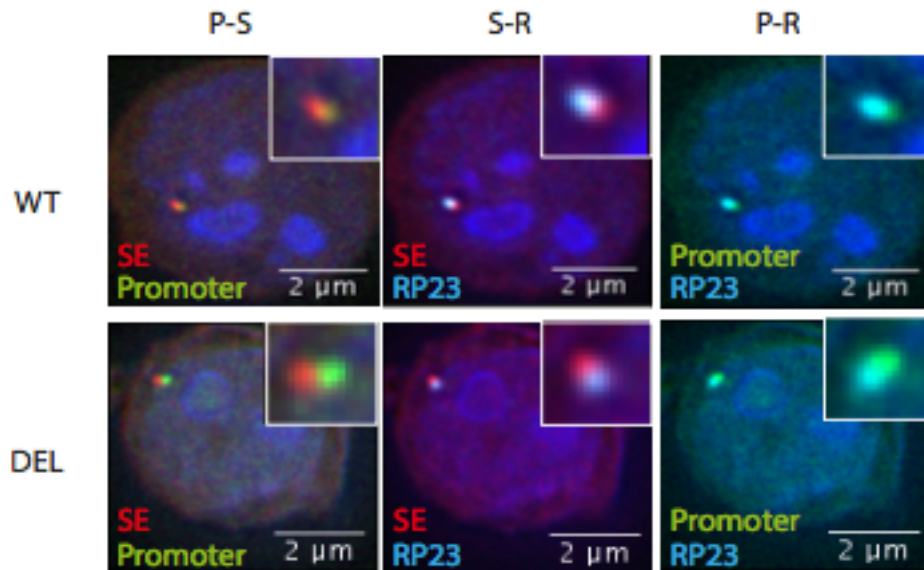

Enh -/-

Sox2 OFF

WT

Sox2 ON
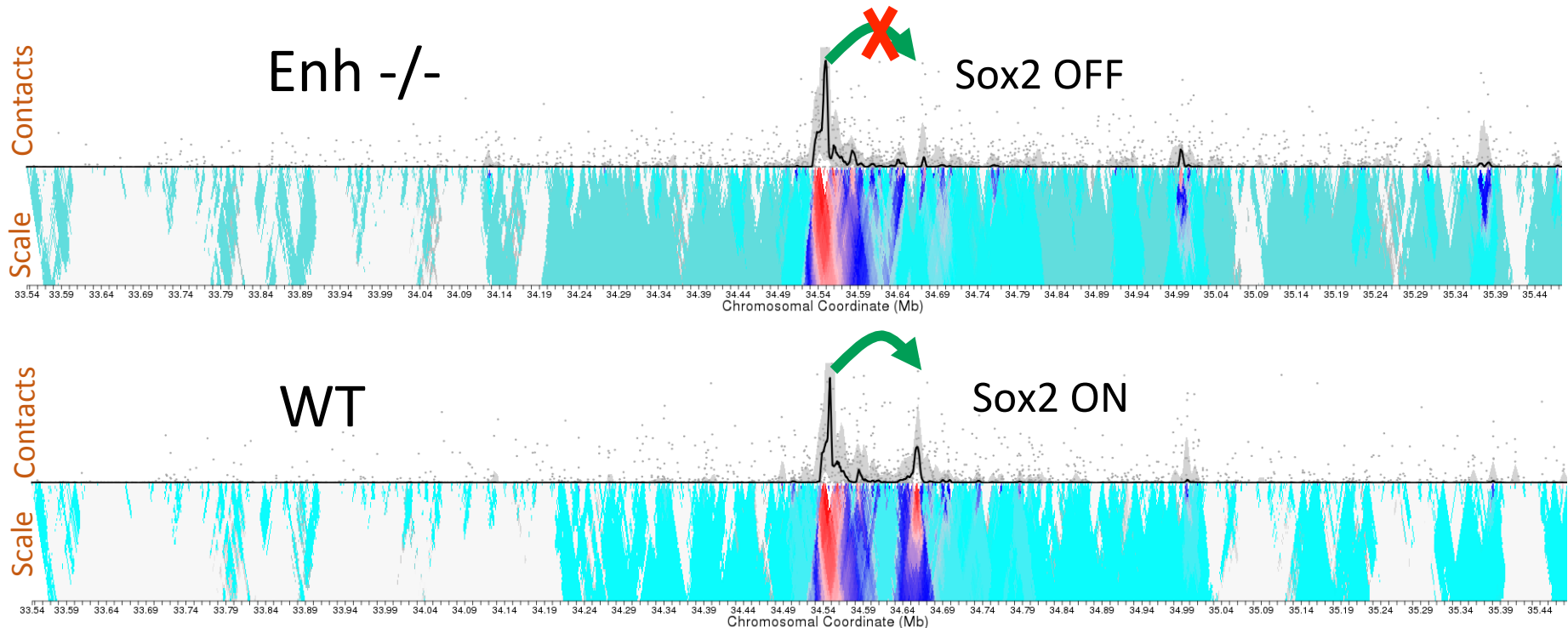
4C-seq analysis with Sox2 as anchor

promoter   enhancer

Sox2

Sox2 enhancer/promoter interactions validated by FISH

Tristin Liu, unpublished
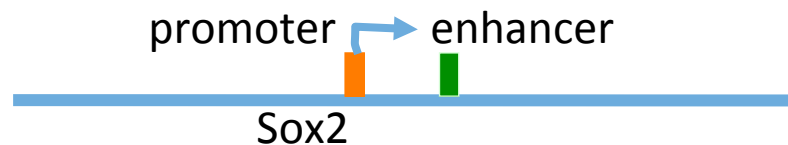
Genome-wide analysis of chromatin interactions

Fix Cells   Digest   Biotin Label, Re-Ligate   Shear, Streptavidin Pull-Down   Sequence, Map

Biotin

GENE A
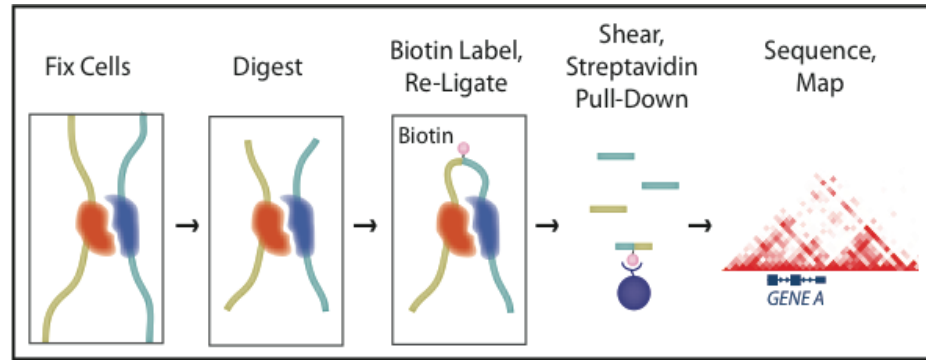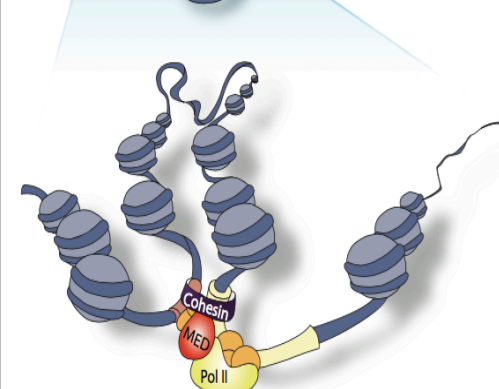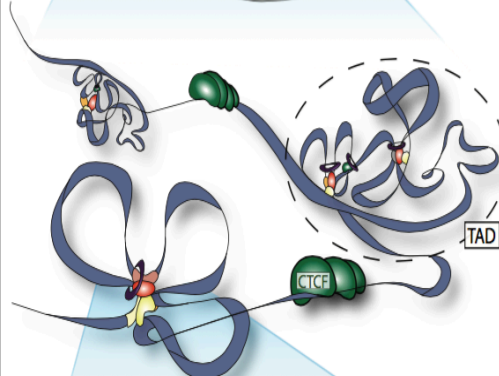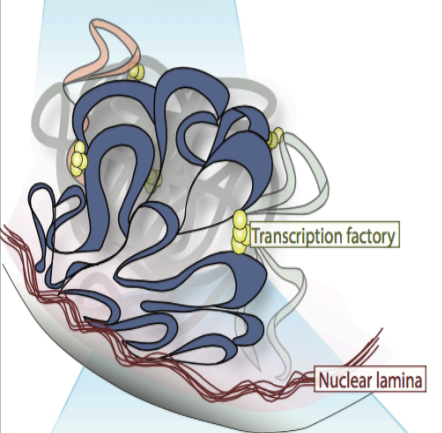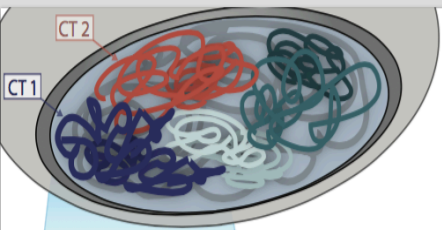
Lieberman … Dekker 2009

Hi-C

Chromosome 3

50Mb   100Mb   150Mb

70
Interaction Counts
0

Scale
chr6: 49500000| 50000000| 50500000| 51000000| 51500000| 52000000| 52500000| 53000000| 53500000| 54000000|

2 Mb

150Mb

B

A
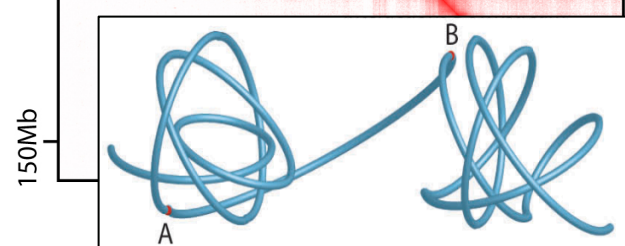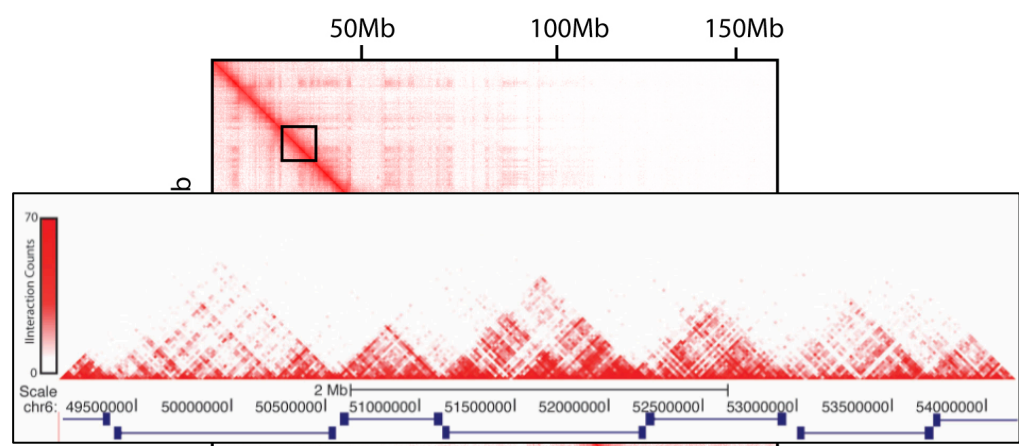
Dixon.. Ren, Nature, 2012

TADs are generally invariant across diverse tissues and cell types
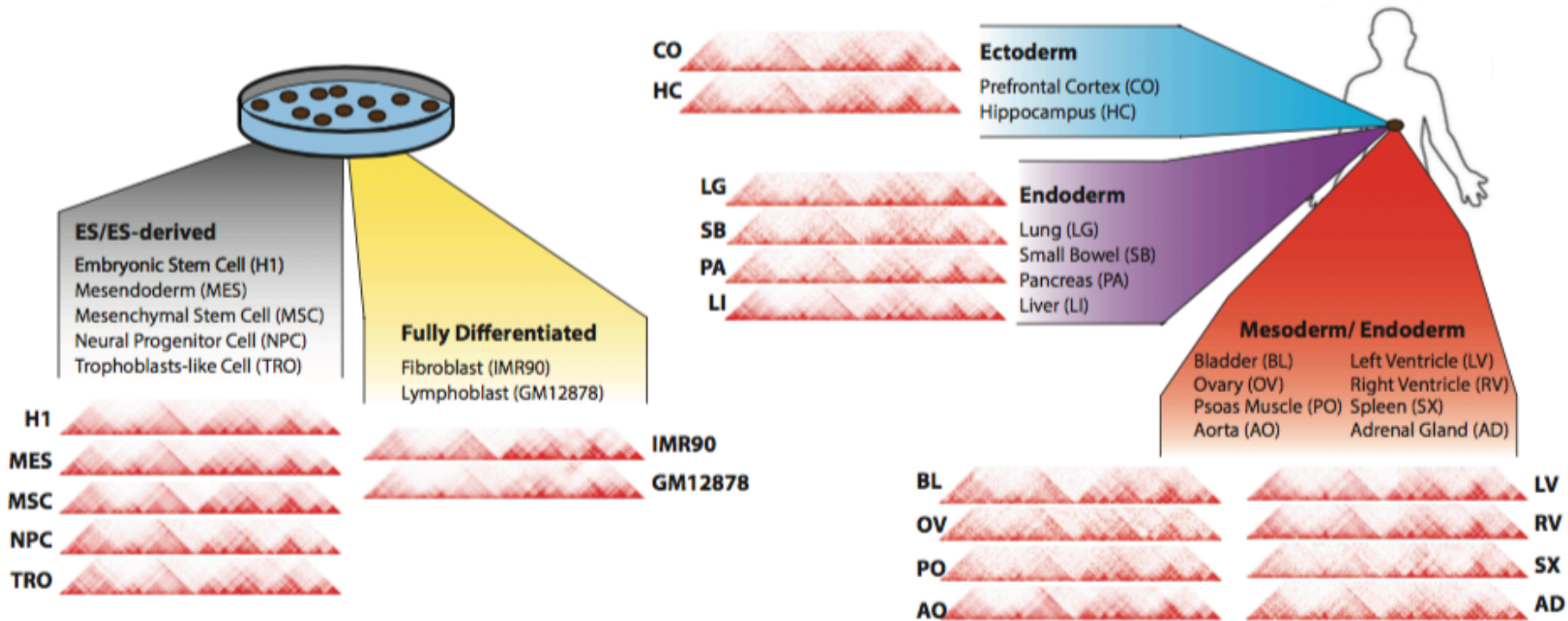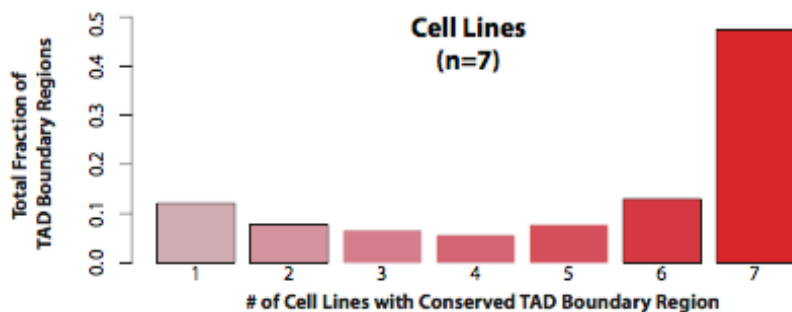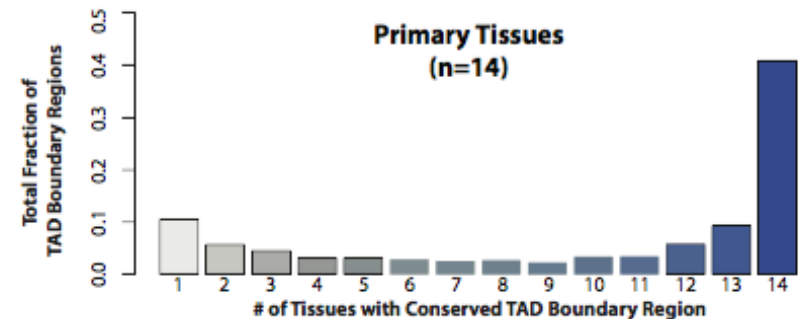
Disruption of TADs Leads to Genetic Disorders in Humans

Lupiáñez, et al. Cell 2015
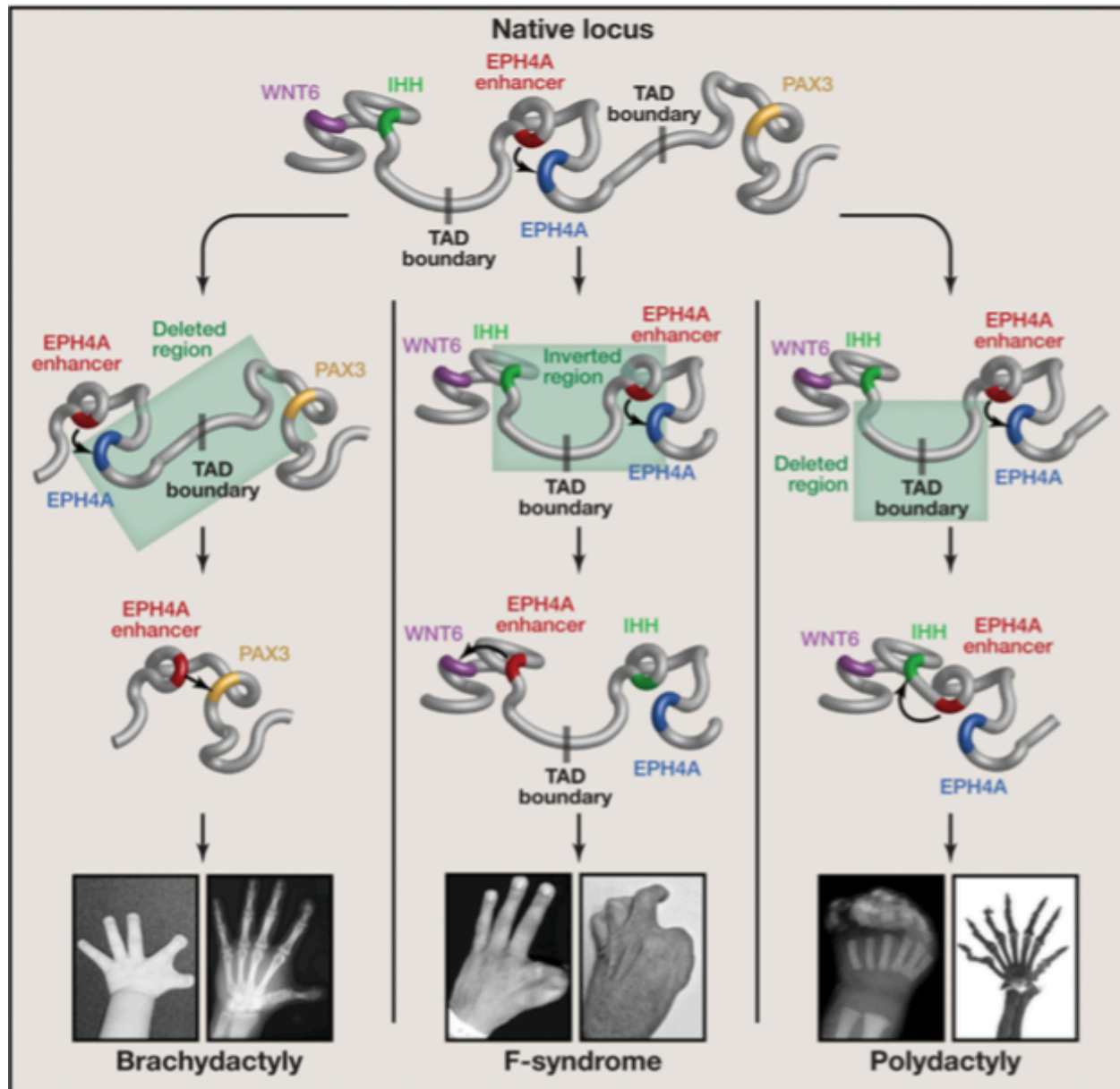
Two Models for TAD Formation

Hand-cuff model

CTCF   CTCF

Cohesin

enhancer

promoter
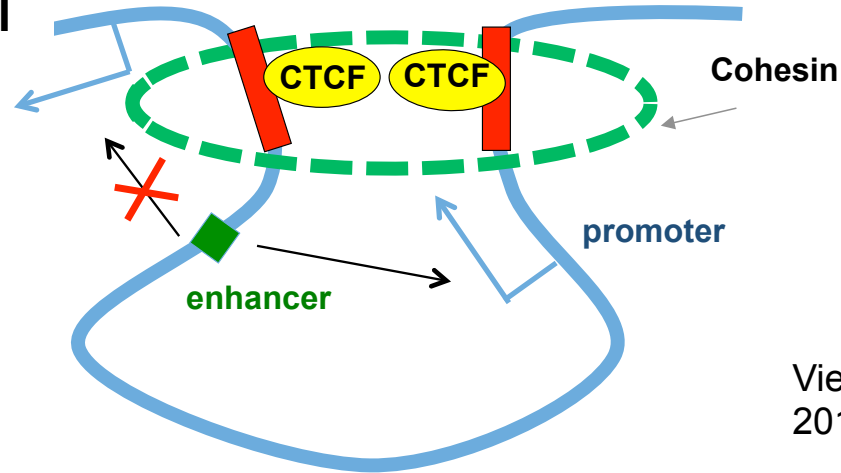
Vietri Rudan and Hadjur,
2015

Extrusion model

CTCF   CTCF

CTCF motif   CTCF motif
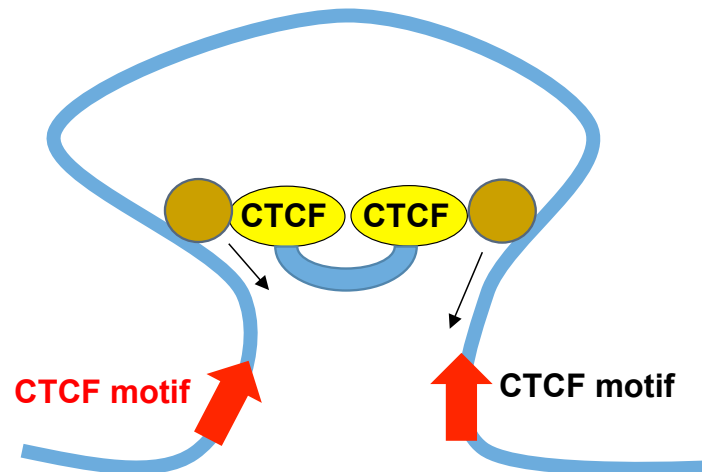
Sanborn … Aiden, 2015
Mirny & Dekker, 2016

ARTICLE

Mediator and cohesin connect gene expression and chromatin architecture

Michael H. Kagey[1]*, Jamie J. Newman[1,2]*, Steve Bilodeau[1]*, Ye Zhan[3], David A. Orlando[1], Nynke L. van Berkum[3], Christopher C. Ebmeier[4], Jesse Goossens[4], Peter B. Rahl[1], Stuart S. Levine[2], Dylan J. Taatjes[4], Job Dekker[3] & Richard A. Young[1,2]
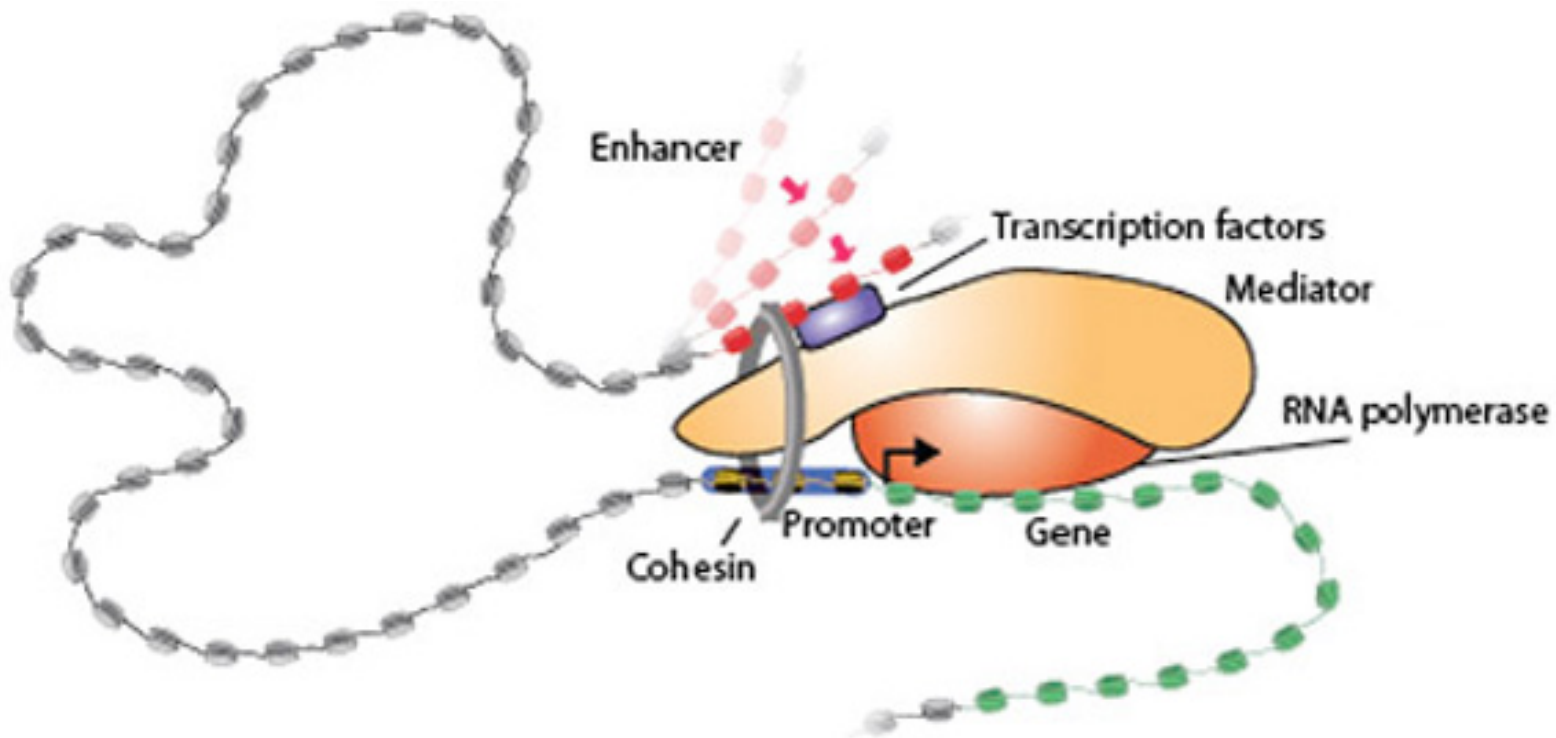
Identify long-range chromatin interactions
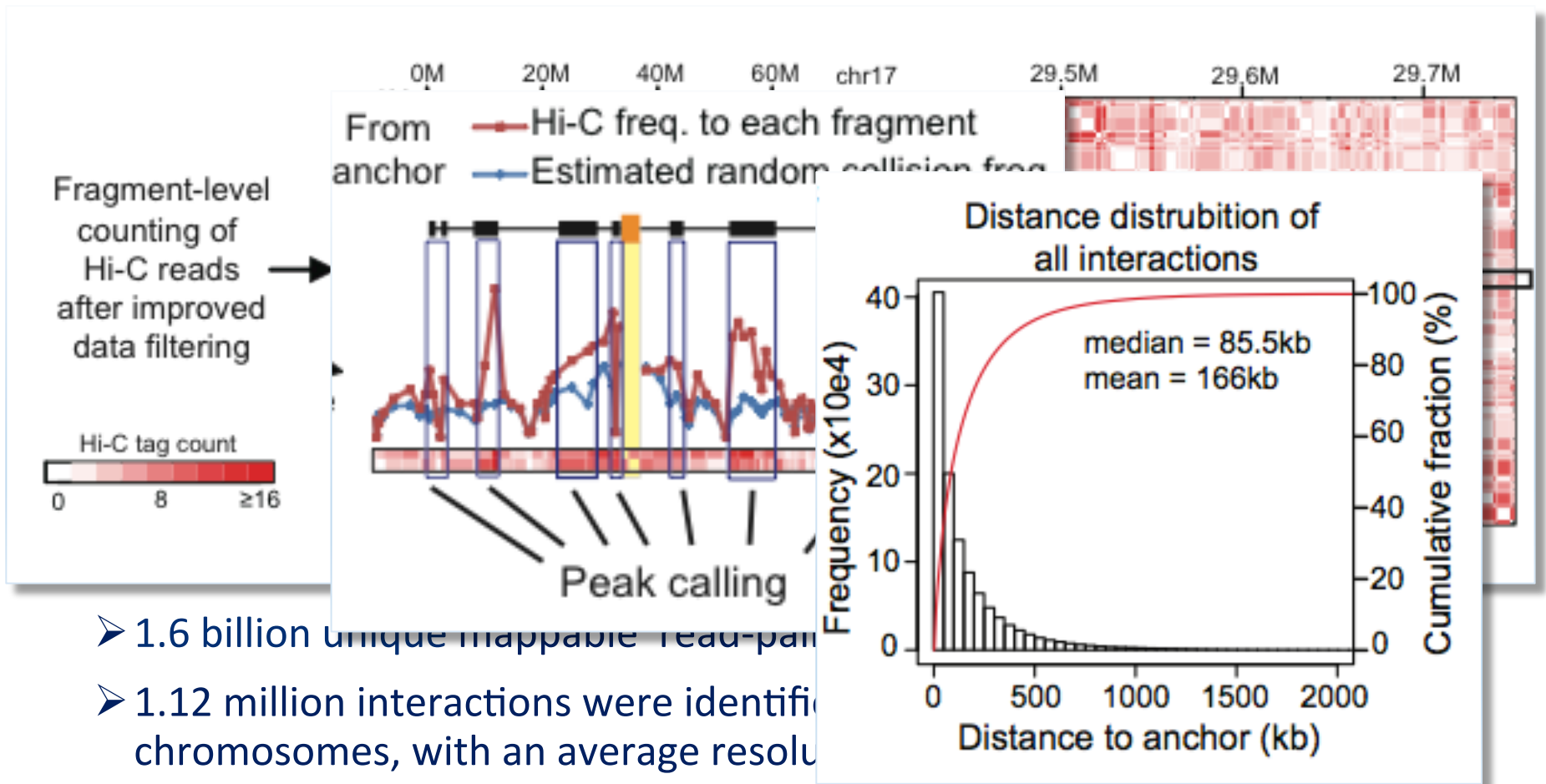
- 1.6 billion unique mappable read-pairs
- 1.12 million interactions were identified on chromosomes, with an average resolution

Promoter capture Hi-C

Design capture probes (RNA)

TSS

Two flanking restriction sites
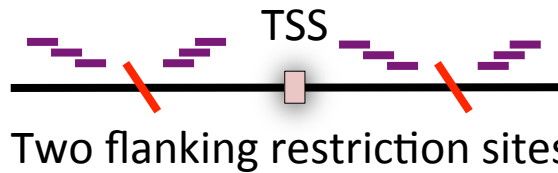
Targeting 96% of protein coding genes

Synthesize biotinylated RNA baits

Hybridize probes to Hi-C library

Construct capture Hi-C library

Sequencing

Similar protocols have been reported by Peter Fraser, Doug Higgs, Rickard Sandberg labs

Jung, Schmitt, unpublished

Promoter capture Hi-C generates high-resolution promoter-centered interaction maps with low cost

IMR90 high resolution HiC (~1.6 billion reads) (Jin, F., et al., Nature 2013)

Promoter capture Hi-C: genome-wide with low sequencing cost

Jung, Schmitt, unpublished

Capture Hi-C in multiple human cell/tissue types identify long-range promoter interactions

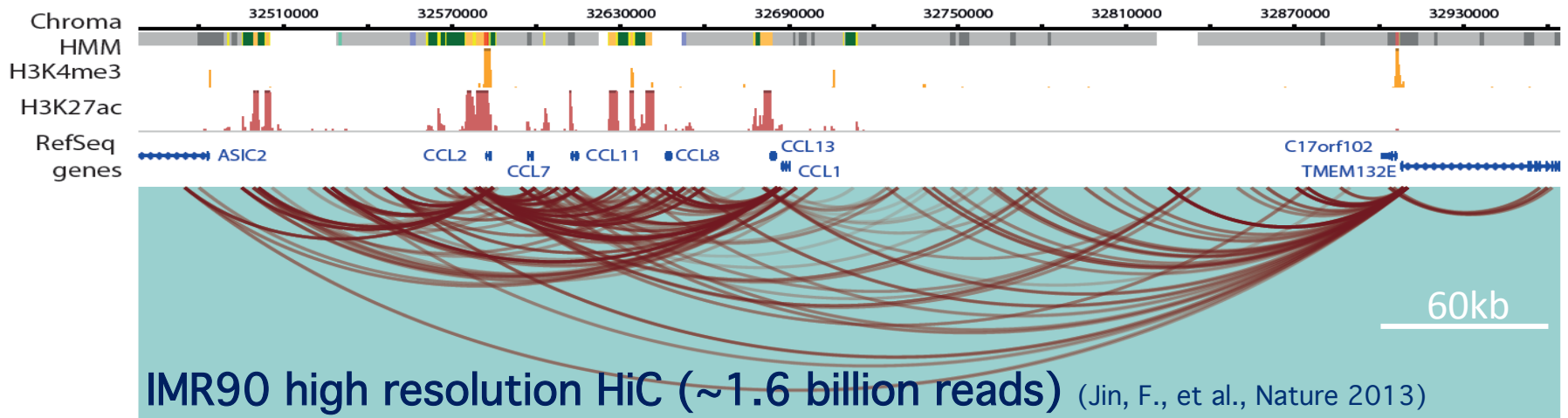| hESC (H1)	Early embryonic cell types (H1-derived cell types)	Ectoderm	Endoderm	Mesoderm
			(Somatic tissue samples isolated from donors)	
	Mesendoderm (ME) Mesenchymal stem cell (MSC) Neural progenitor cells (NPC) Trophoblast-like cells (TB)	Hippocampus (HC) Dorsolateral prefrontal cortex (DLPFC)	Esophagus (EG) Fibroblasts (IMR90) Lung (LG) Liver (LI) Pancreas (PA) Small bowel (SB) Sigmoid colon (SG) Thymus (TH) Bladder (BL)	Adrenal gland (AD) Aorta (AO) Gastric tissue (GA) Left heart ventricle (LV) Right heart ventricle (RV) Right heart atrium (RA) Ovary (OV) Psoas (PO) Spleen (SX) FAT (FT) Lymphoblast (LCL)

27 cell/tissue types
5.5 billion uniquely mapped monoclonal capture HiC (cHiC) reads
6 core hisotne modification marks
RNA-seq samples

In collaboration with Jamie Thomson (U. Wisconsin), Shin Lin (Stanford) and Yiing Lin (Wash U)

Leung, D., Jung, I., Rajagopal, N., et al., Nature (2015)
Schultz, MD., He, Y., et al., Nature (2015)

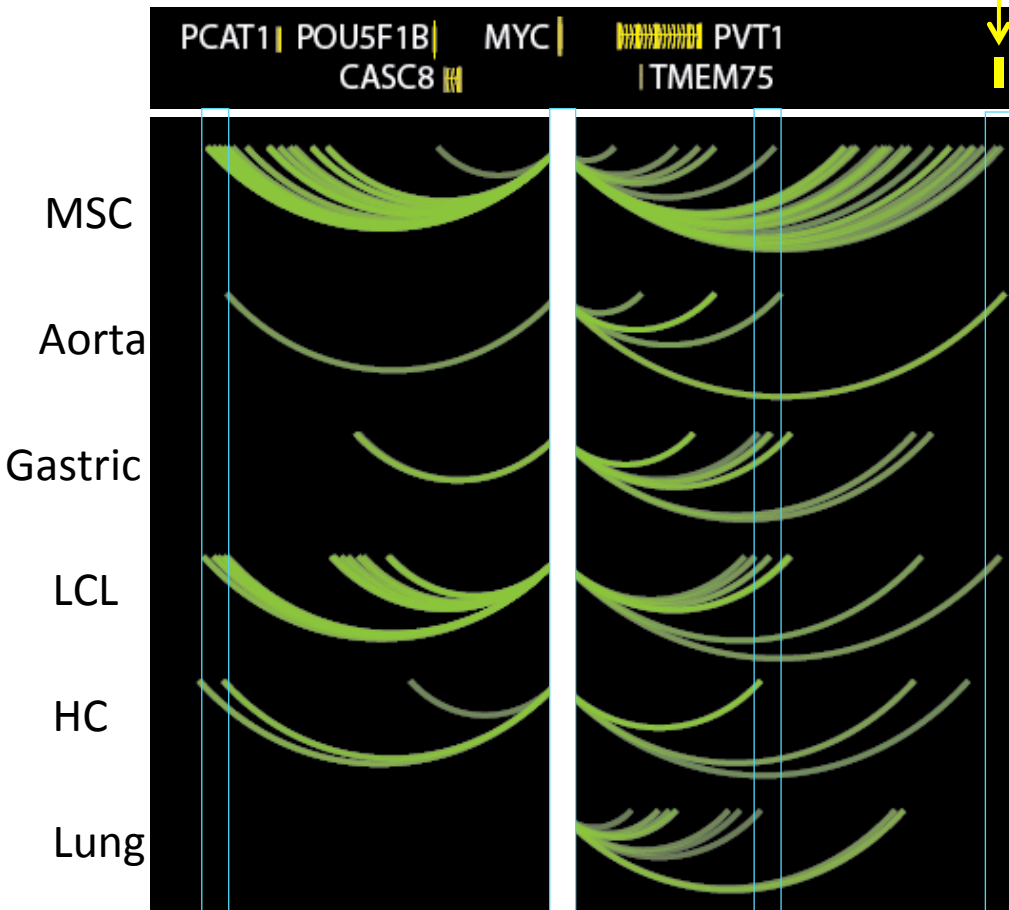# Constructing long-range promoter-centered interactome maps

300kb

MYC enhancer

| PCAT1 | POU5F1B | MYC | | PVT1 |
| | CASC8 | | | TMEM75 |

MSC

Aorta

Gastric

LCL

HC

Lung

➤Number of interactions in 27 tissues: ~700k unique ones
➤Resolution: 5.4kb
➤Average interaction distance: 238kb average distance

P-E (37%)

P-PE (6.5%)

P-P (7.5%)

P-O (49%)
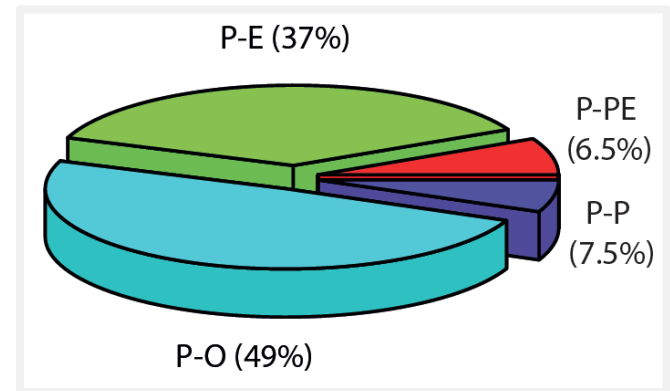
P: Promoter
E: Enhancer
O: Other regions

Herranz, D., et al., Nature Medicine (2014)

Jung, Schmitt, unpublished

Validation of the indentified interactions by eQTL discovered in lymphoblastoid cell lines
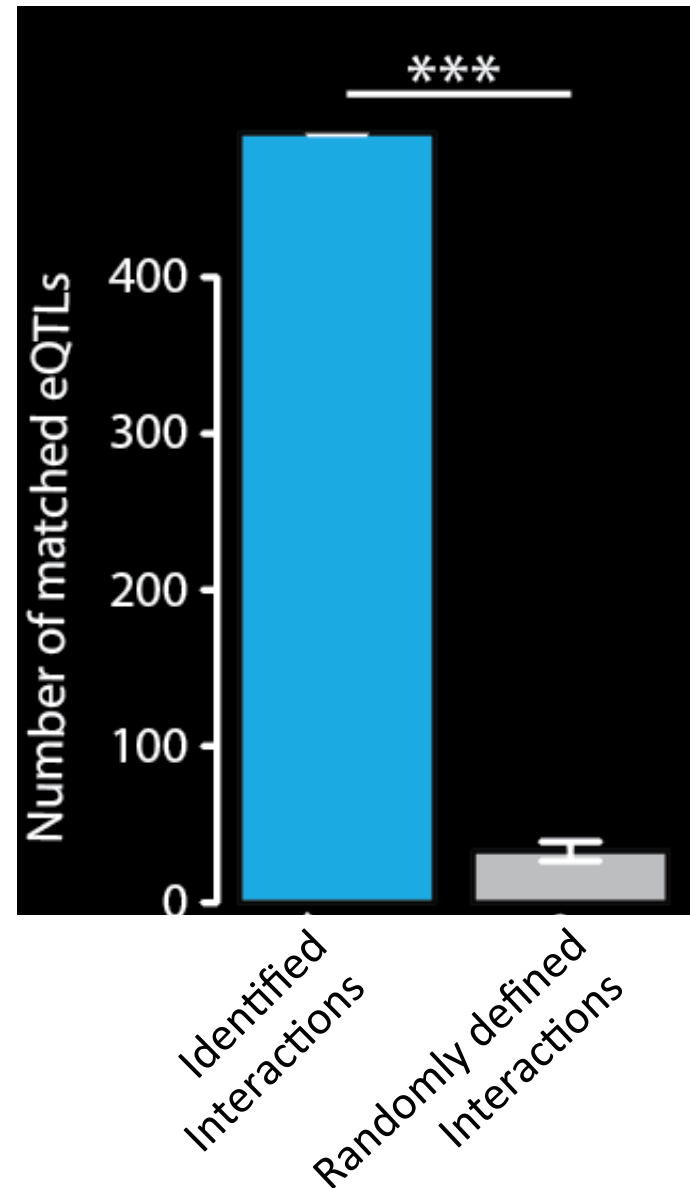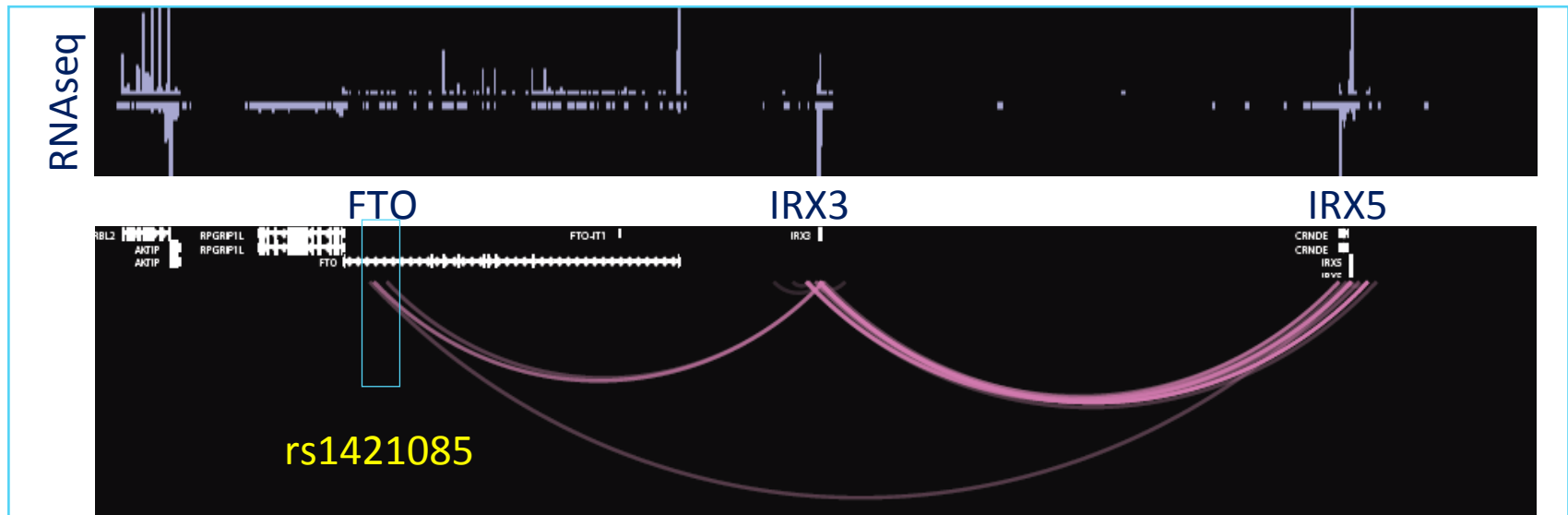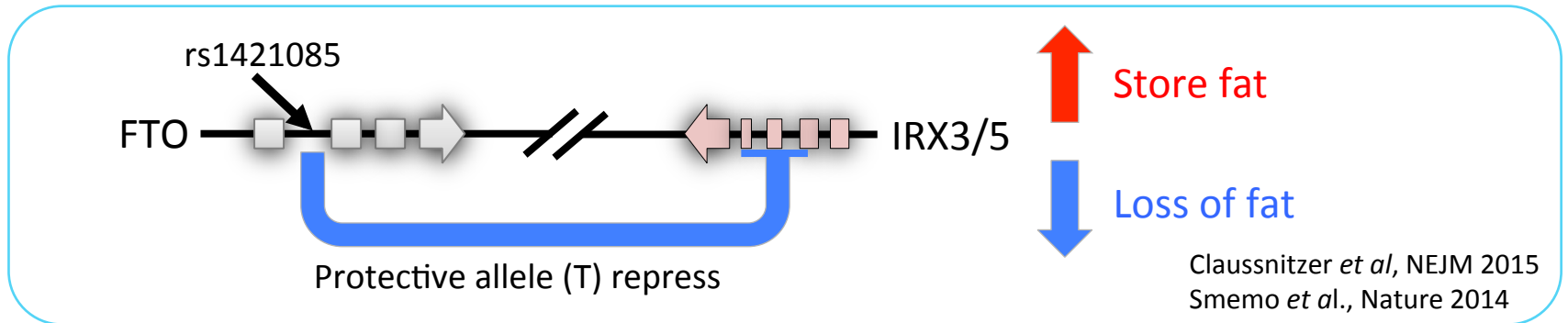
eQTLs (expression quantitative trait loci)
: Identification of genomic loci that
contribute to variation in gene expression

eQTLs from lymphoblastoid cell lines with
462 individuals (Lappalainen, T. *et al* Nature 2013)

Jung, Schmitt, unpublished

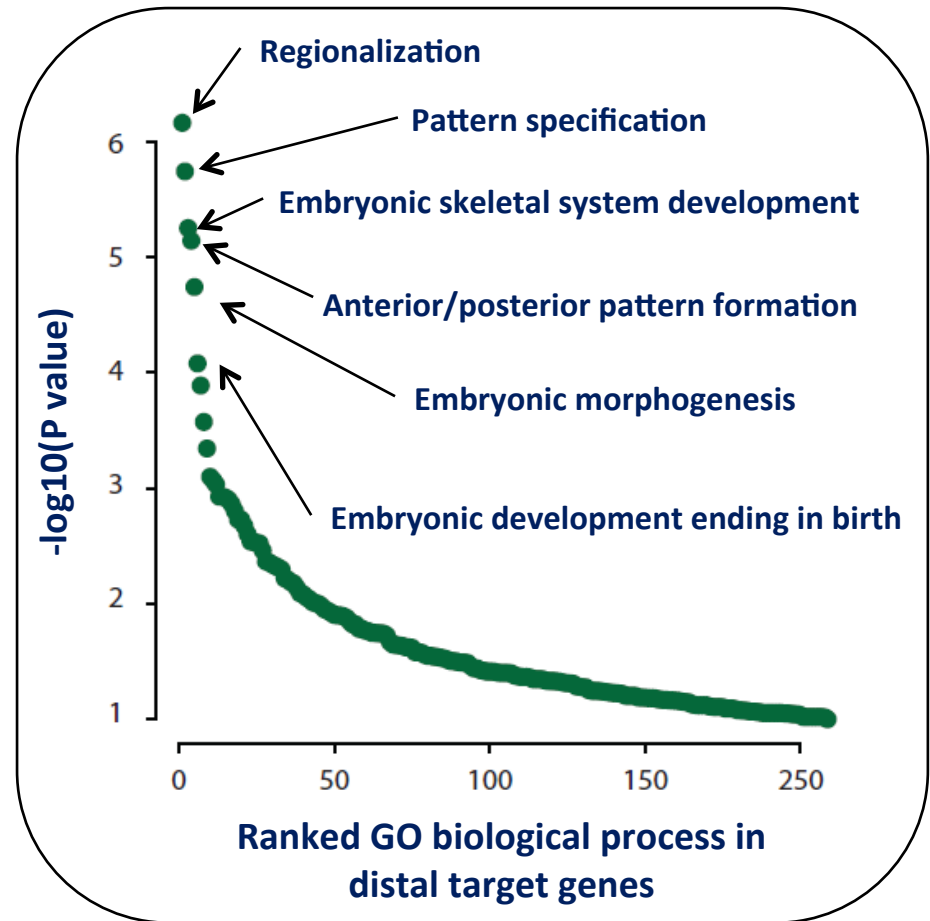# Promoter capture Hi-C identifies a known target genes for a SNP linked to human obesity risk

rs1421085

FTO

IRX3/5

Store fat

Loss of fat

Protective allele (T) repress

Claussnitzer *et al*, NEJM 2015
Smemo *et al*., Nature 2014
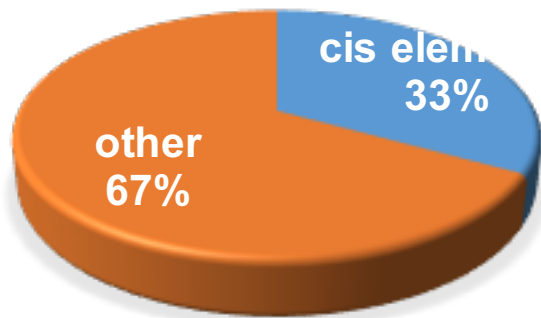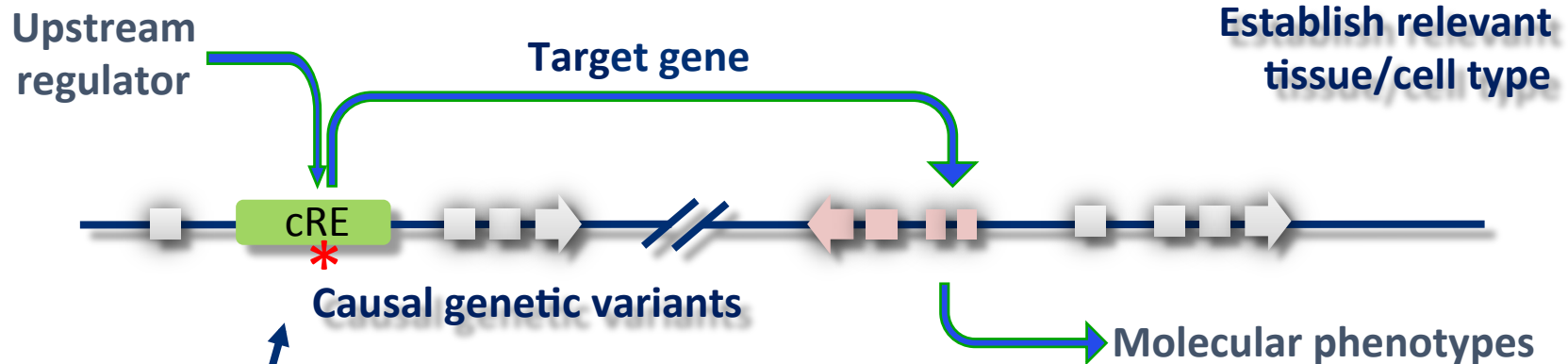
RNAseq

FTO   IRX3   IRX5

rs1421085

"The identified long-range promoter interactions can be useful to reveal distal target genes of disease associated SNPs"

Promoter-centered interactome maps suggest putative target genes for neural disorder SNPs

GWAS SNPs in brain disorders	Number of total target genes	Number of nearest genes	Number of distal genes
908	2,102	326 (16%)	1776 (84%)

From long-range interactome maps

From nearest gene information (788)

115   27   60

Number of matched genes in OMIM database

Regionalization

Pattern specification

Embryonic skeletal system development

Anterior/posterior pattern formation

Embryonic morphogenesis

Embryonic development ending in birth

-log10(P value)

Ranked GO biological process in distal target genes

Jung, Schmitt, unpublished

Toward a better understanding of non-coding GWAS SNPs

- Do the identified cis regulatory elements function to regulate target gene expression?

- Do SNPs affect binding of transcription factors (TF) to the cis regulatory elements? If so which ones?

- Does disruption of the TF binding to a candidate cis elements lead to changes in transcriptional programs?
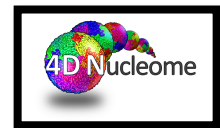
Acknowledgments

TAD discovery

- Jesse Dixon, MD/ PhD (now Fellow at Salk)
- Sid Selvaraj, PhD

Chromatin Interaction Maps

- Anthony Schmitt
- Inkyung Jung, Ph.D.
- Jian Yan, Ph.D.
- Tristin Liu
- Fulai Jin, Ph.D. (now Ass. Prof. @ Case Western)
- Yan Li, Ph.D. now Ass. Prof. @ Case Western)

Acknowledgements

Past

Postdoctoral Fellows

Zirong Li
Tae Hoon Kim (U. Texas, Dallas)
Kun Wang
David Hawkins (U of Washington)
Andrea Local
Fulai Jin (Case Western Reserve U.)
Celso Espionoza
Feng Yue (Penn State U.)
Yin Shen (UCSF)
Tingting Du
Yan Li (Case Western Reserve U)
Wei Xie (Tsinghua University, China)
Danny Leung (HKUST)

Graduate Students

Leah O. Barrera
Nate Heintzman
Gary Hon (UT Southwestern)
Nate Maynard
Nisha Rajagopal
Saurabh Agarwal
Chloe Rivera
Jesse Dixon (Salk Institute)
Siddarth Selvaraj

Lee Edsall
Ulrich Wagner
Keith Ching

Present

Postdoctoral Fellows

Inkyung Jung
David Gorkin
Yarui Diao
Jian Yan
Guoqiang Li
Sebastian Preissl
Miao Yu

Graduate Students

Anthony Schmitt
Anugraha Raman
Yunjiang Qiu
Yuan Zhao
Rongxin Fang

Haruhiko Ishii
Bin Li
Zhen Ye
Samantha Kuan
Ah Young Lee
Sora Chee
Tristin Liu

Collaborators

Wei Wang (UCSD)
Maike Sander (UCSD)
Kelly Frazer (UCSD)
Kang Zhang (UCSD)
Sheng Zhong (UCSD)
Paul Mischel (LICR/UCSD)
Huilin Zhou (LICR/UCSD)
Neil Chi (UCSD)
Sylvia Evans (UCSD)
Kun Zhang (UCSD)
Dong Wang (UCSD)
Kun-liang Guan (UCSD)
Joe Ecker (Salk)
Len Pennacchio (LBNL)
Axel Visel (LBNL)
Michael Zhang (UTD)
Jamie Thomson (UW)
Chuan He (UofC)
Shin Lin (Stanford)
Yiing Lin (Wash. U.)
Ming Hu (NYU)
Jun Liu (Harvard)
Kai Ge (NIDDK)
Cathy Barr (U. Toronto)
Kerstin Wendt (Erasmus MC)
Victor Lobanenkov (NIAID)
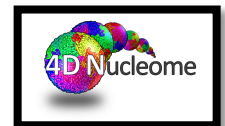
Challenges in deciphering the function of non-coding sequences

- A substantial fraction of the non-coding variants are suspected to act by affecting gene regulation, but advancing this hypothesis requires a deeper understanding of the gene regulatory processes in human cells;

- The *cis* regulatory sequences in the human genome remain to be clearly defined

- There is an insufficient molecular understanding of the interplays among signaling pathways, transcription factors, chromatin regulators, chromatin structure and *cis* regulatory sequences
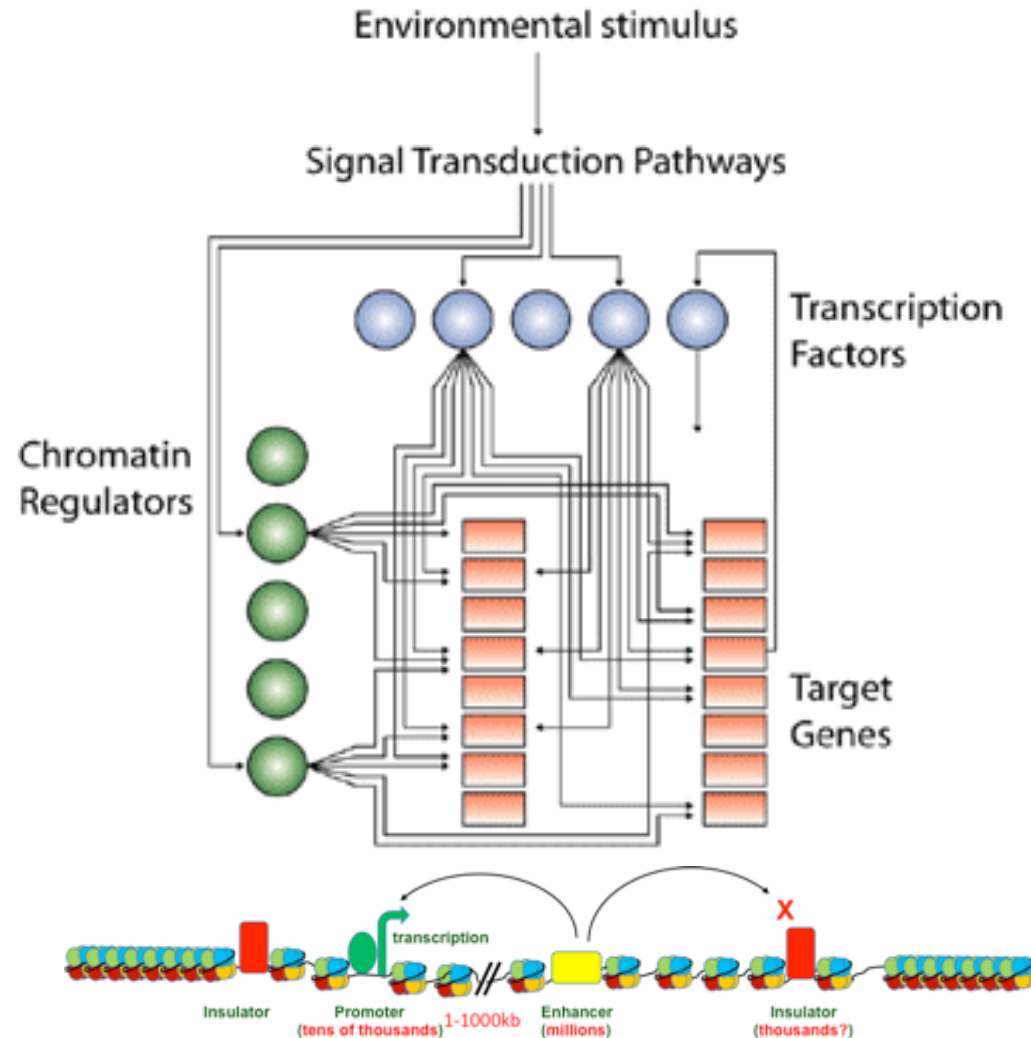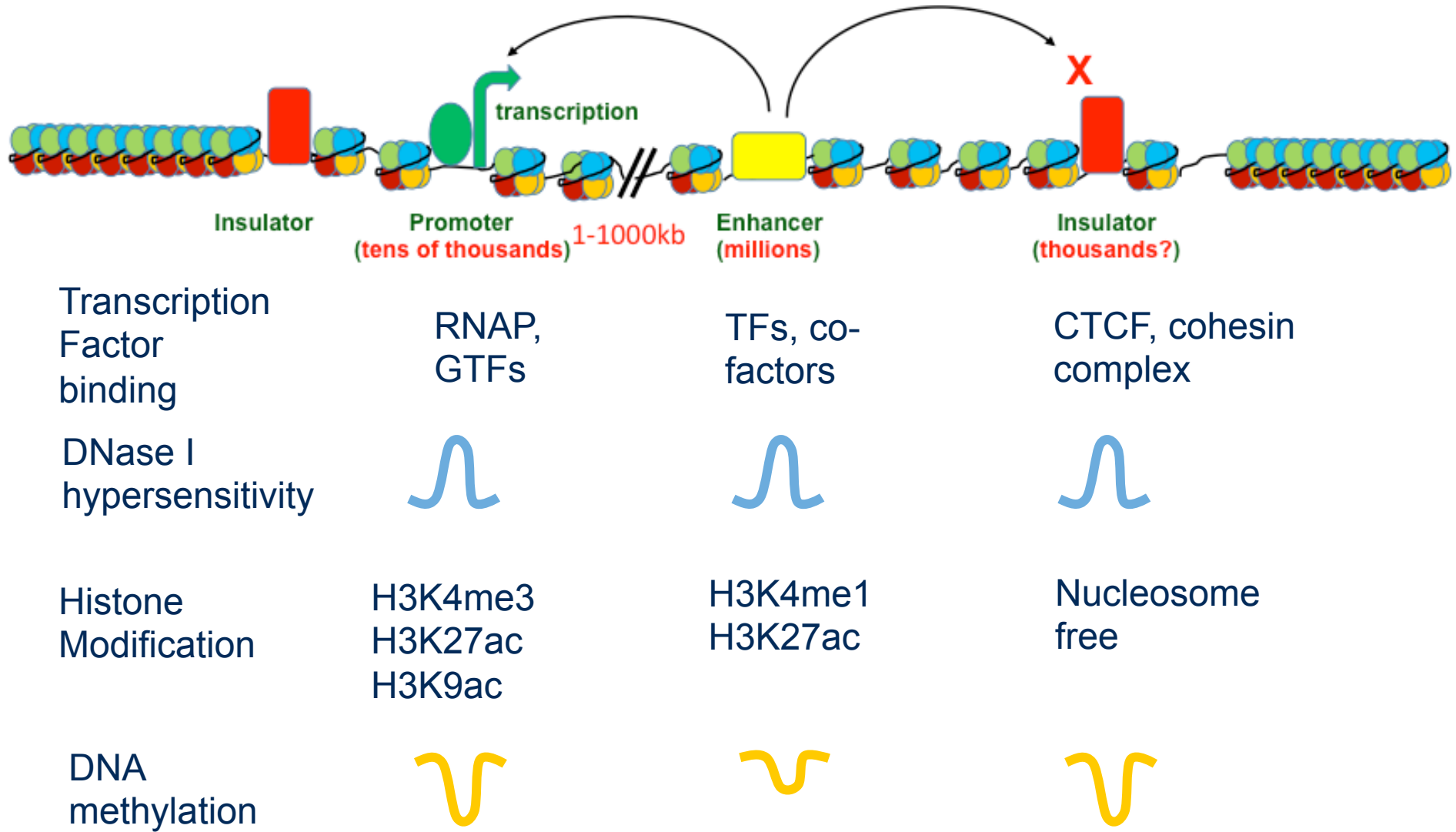
# Multidimensional Genomic Analysis Enabled by Next-gen Sequencing Technologies

Reviewed by Rivera & Ren, Cell 2013
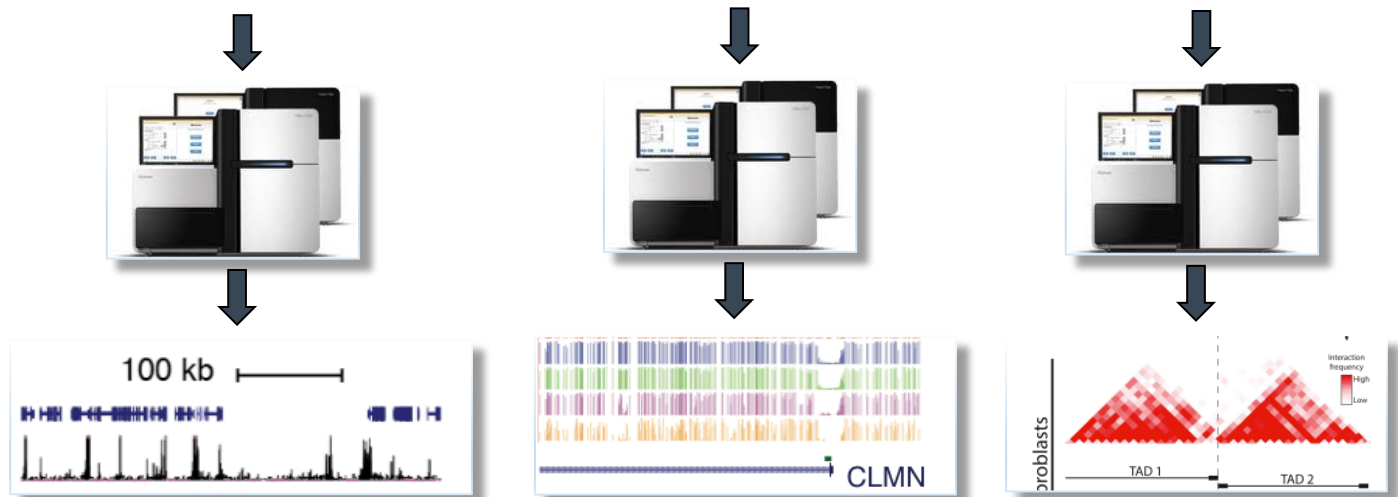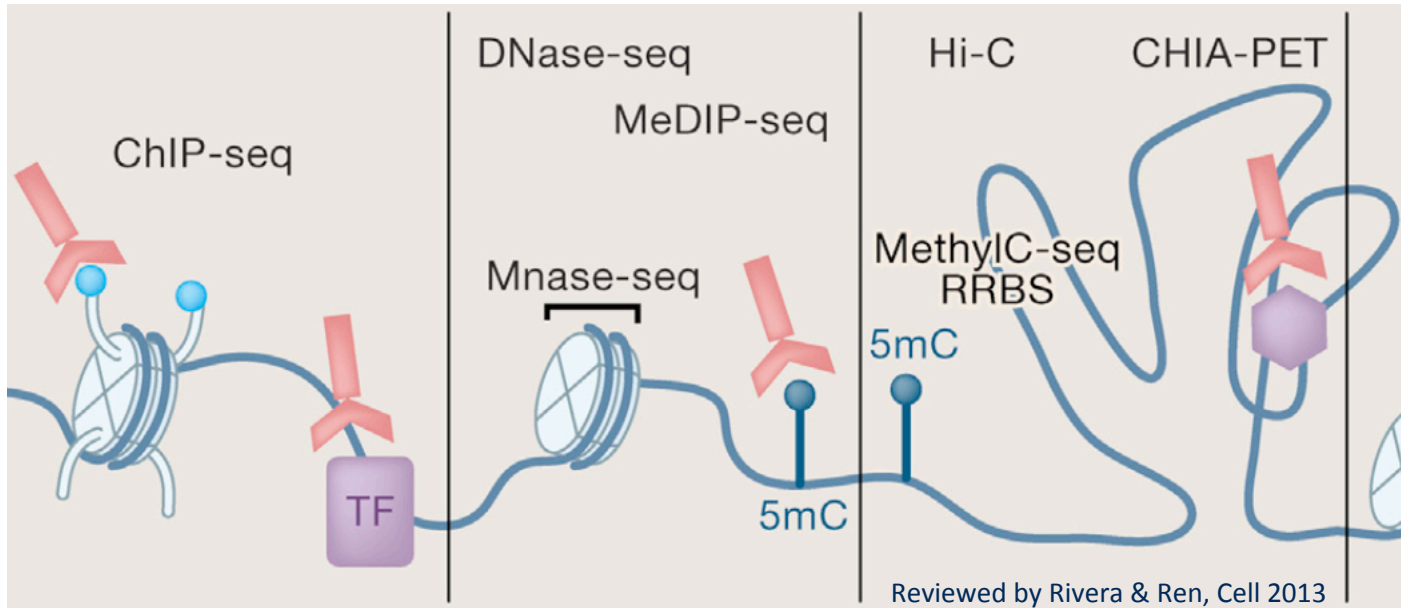
Chromatin signatures of *cis* elements

Rivera & Ren, Cell 2013

Topological Associating Domains (TADs)

- TADs are stable during cellular differentiation and between cell types (Dixon et al., Nature 2012; Dixon et al. Nature 2015)

- TADs are evolutionarily conserved (Dixon et al., Nature 2012)

- TADs have been independently observed in flies (Sexton et al. Cell 2012; Hou et al, Mol Cell 2012) and other species, using a variety of approaches (5C, Nora et al., Nature 2012)

- TADs correspond to DNA replication domains (Hope … Ren, Gilbert, Nature 2014). They disappear in mitosis but reestablishes in G1 (Naumova et al. 2014)

- Partitioning of the genome into TADs provides a structural understanding of how enhancers come into contact with their target promoters (Nora et al. Nature 2012; Lupiáñez, et al. Cell 2015; Nerandra et al, Science 2015)
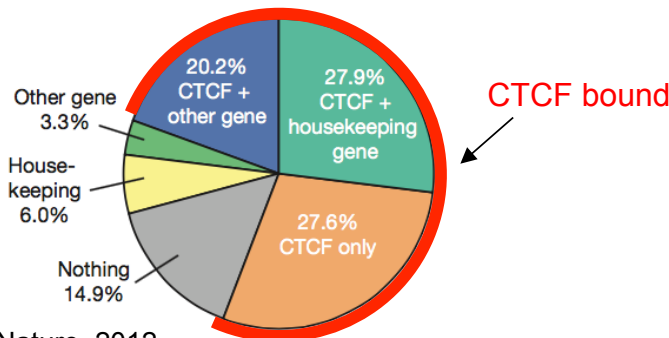
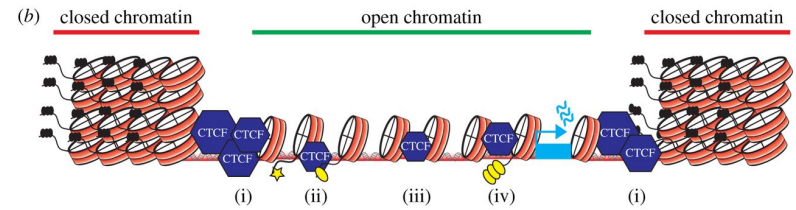# An Insulator-binding Protein, CTCF, is Enriched at TAD Boundaries
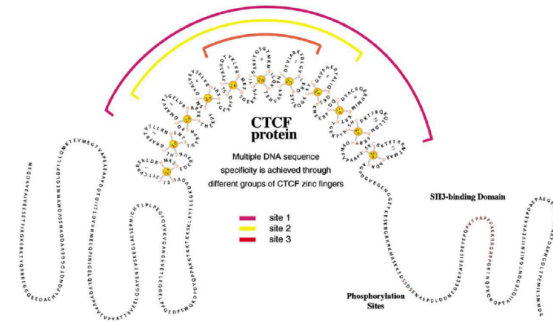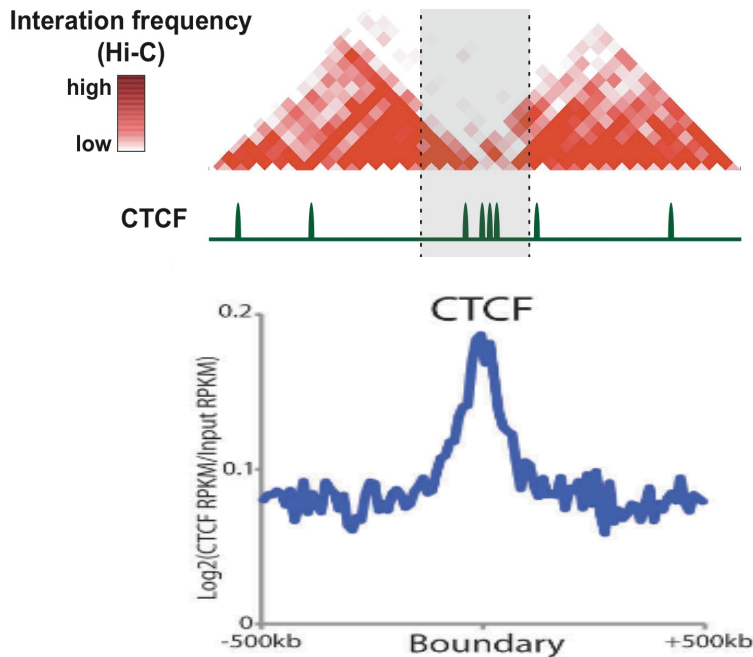
Lobanenkov, 1993; Felsenfeld, 1999
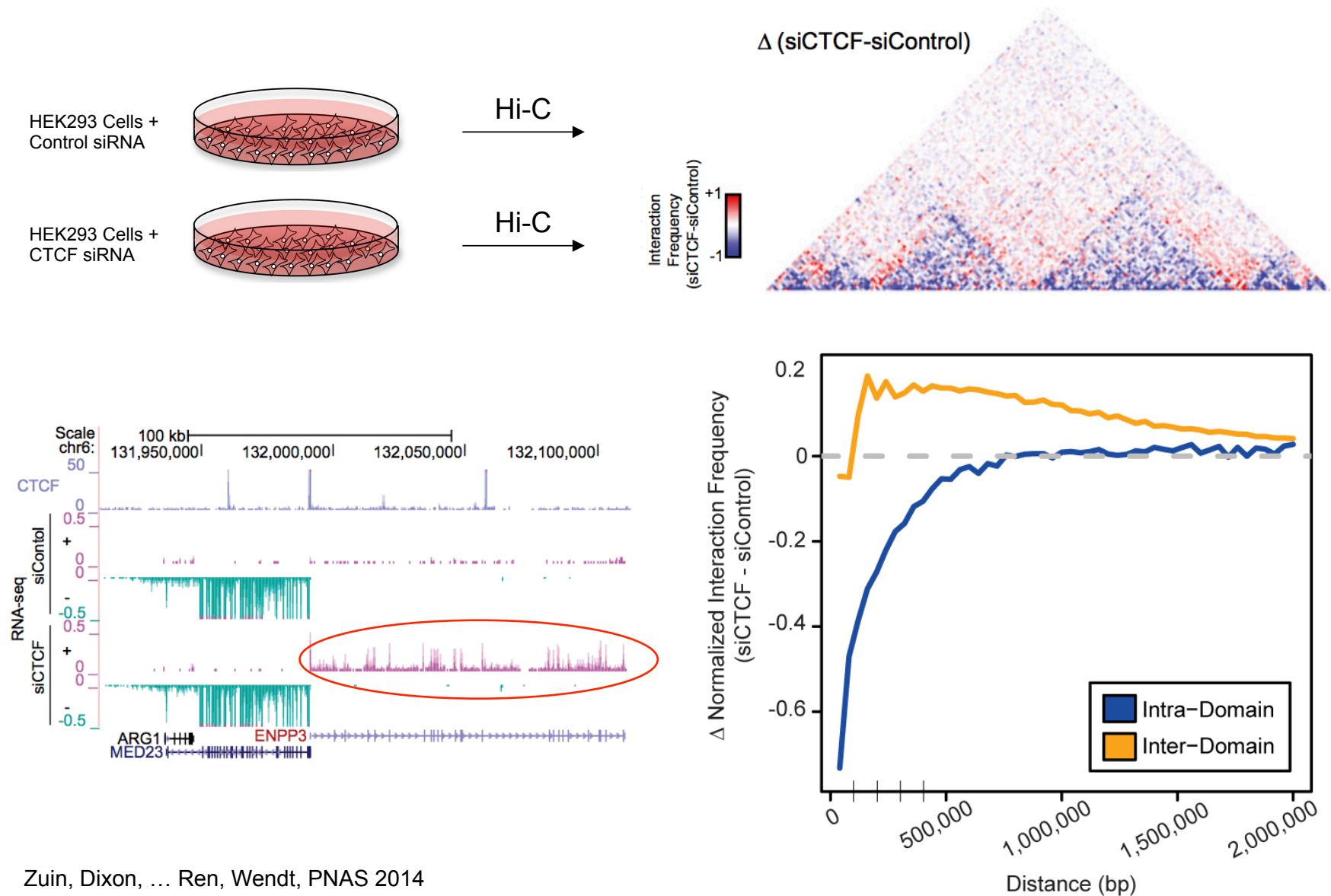
Kim .. Ren, 2007; Cuddapah .. Zhao, 2009

Dixon.. Ren, Nature, 2012

Holwerda & de Laat, 2013

Depletion of CTCF Protein in Cells Leads to Increased Inter-TAD Interactions While Reduced Intra-TAD Interactions

Zuin, Dixon, … Ren, Wendt, PNAS 2014

Orientation of CTCF Binding Sites is Critical For Local Chromatin Organization

HEC-1B cells

Guo et al., Cell 2015