Super-Enhancers and Transcriptional Condensates

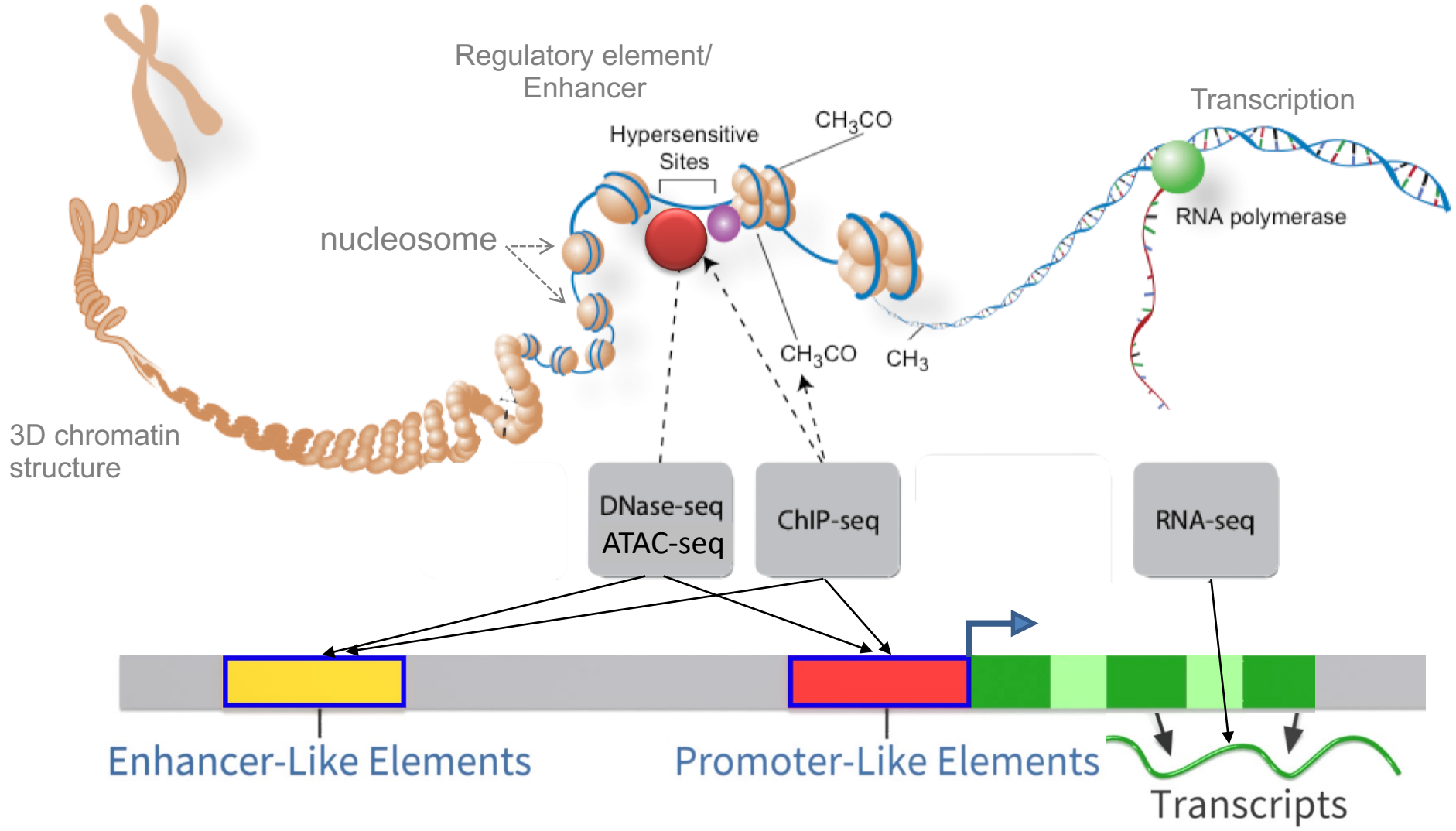
Chongzhi Zang
zang@virginia.edu
zanglab.org

May 3, 2022

Outline

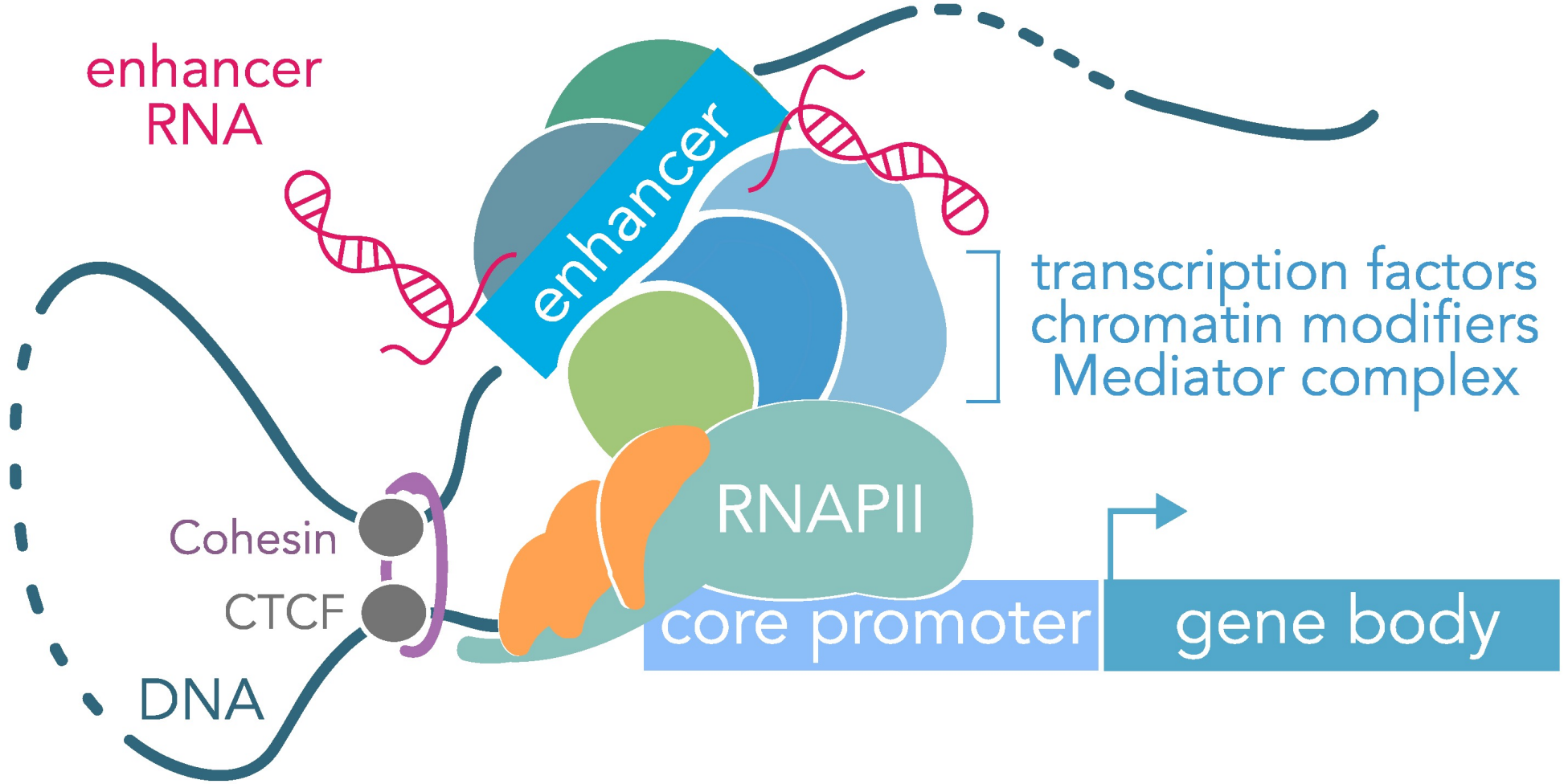
- Enhancers
- Super-enhancers/stretch enhancers
- Super-enhancers in cancer transcription control
- Super-enhancers and transcriptional condensate

Enhancers: Cis-regulatory elements in the genome



Modified from ENCODE, Darryl Leja (NHGRI), Ian Dunham (EBI)

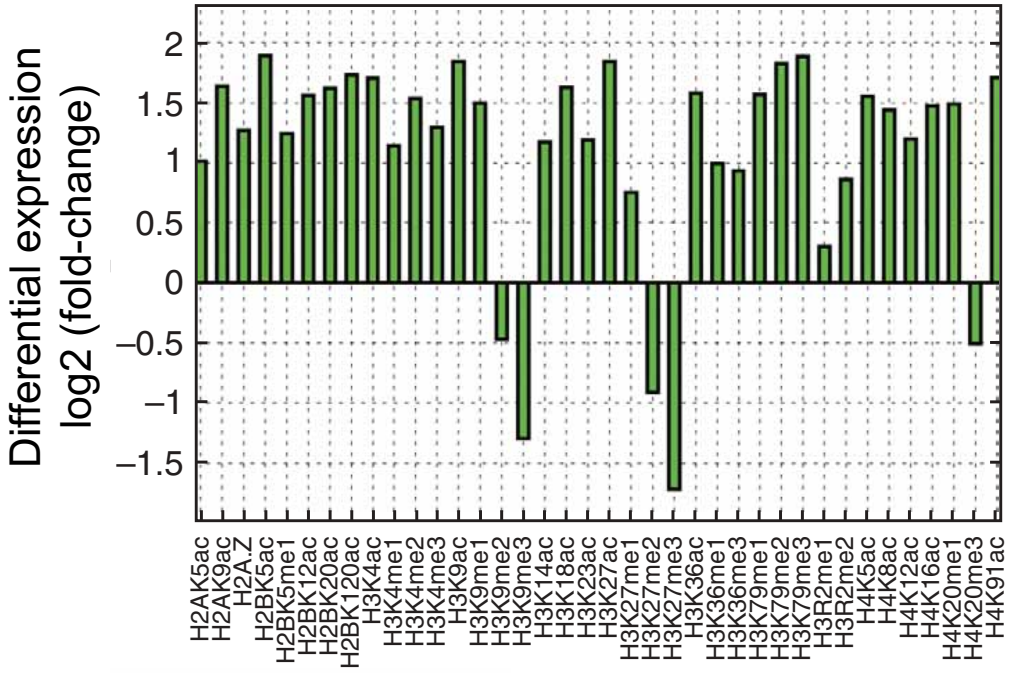
Enhancers



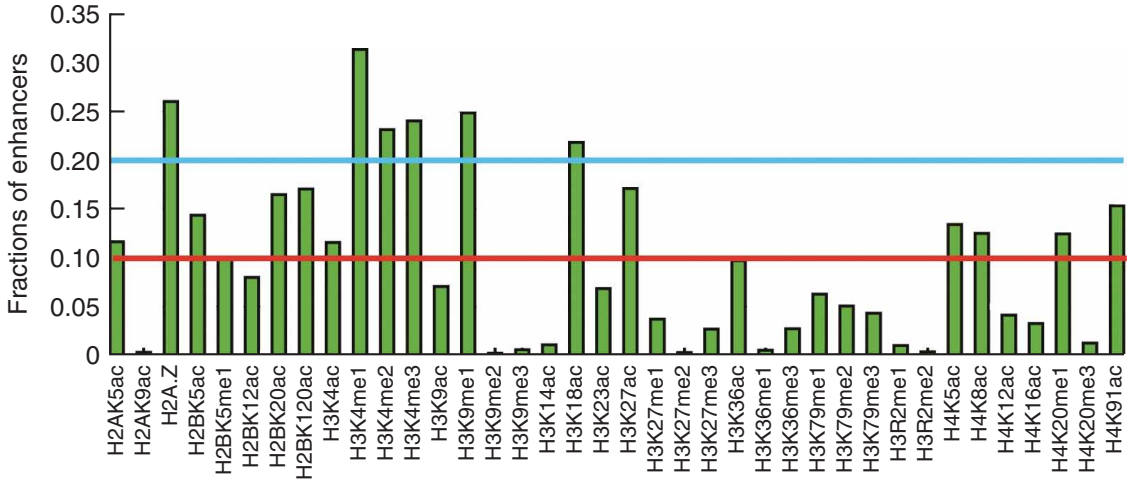
Carullo & Day 2019

Histone modifications associate with enhancers

Promoters



Putative enhancers



Wang, Zang et al. *Nat Genet* 2008

Histone modifications at enhancers

PNAS

SEE COMMENTARY

Histone H3K27ac separates active from poised enhancers and predicts developmental state

Menno P. Creyghton[a,1], Albert W. Cheng[a,b,1], G. Grant Welstead[a], Tristan Kooistra[c,d], Bryce W. Carey[a,e], Eveline J. Steine[a,e], Jacob Hanna[a], Michael A. Lodato[a,e], Garrett M. Frampton[a,e], Phillip A. Sharp[d,e], Laurie A. Boyer[e], Richard A. Young[a,e], and Rudolf Jaenisch[a,e,2]

[a]Whitehead Institute for Biomedical Research, Cambridge, MA 02142; [b]Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, MA 02142; [c]Division of Health Sciences and Technology, Harvard–Massachusetts Institute of Technology, Boston, MA 02115; [d]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02142; and [e]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

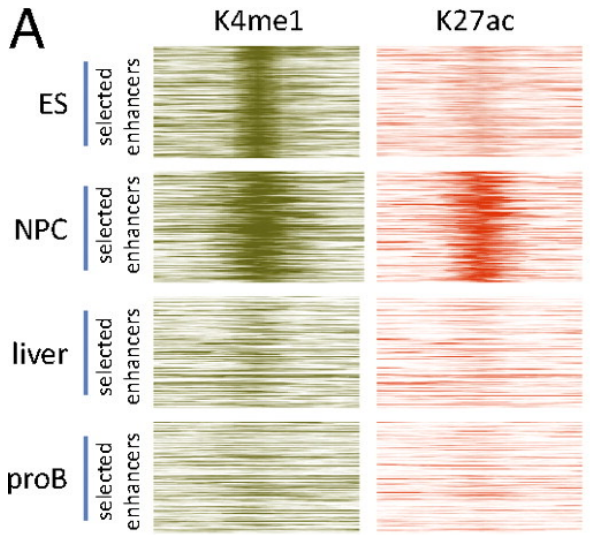
Contributed by Rudolf Jaenisch, October 26, 2010 (sent for review September 29, 2010)

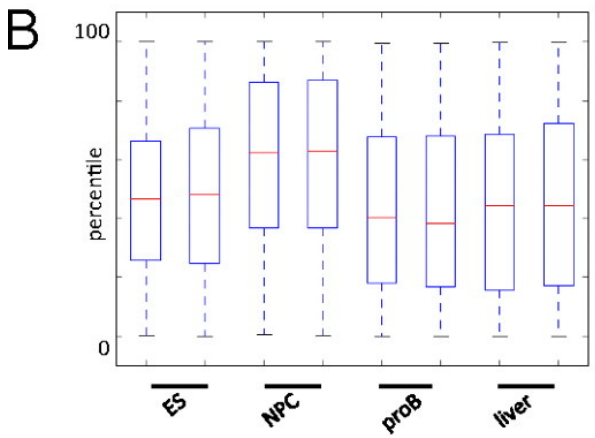
Developmental programs are controlled by transcription factors and chromatin regulators, which maintain specific gene expression programs through epigenetic modification of the genome. These regulatory events at enhancers contribute to the specific gene expression programs that determine cell state and the potential for differentiation into new cell types. Although enhancer elements are known to be associated with certain histone modifications and transcription factors, the relationship of these modifications to gene expression and developmental state has not been clearly defined. Here we interrogate the epigenetic landscape of enhancer elements in embryonic stem cells and several adult tissues in the mouse. We find that histone H3K27ac distinguishes active enhancers from inactive/poised enhancer elements containing H3K4me1 alone. This indicates that the amount of actively used enhancers is lower than previously anticipated. Furthermore, poised enhancer networks provide clues to unrealized developmental programs. Finally, we show that enhancers are reset during nuclear reprogramming.

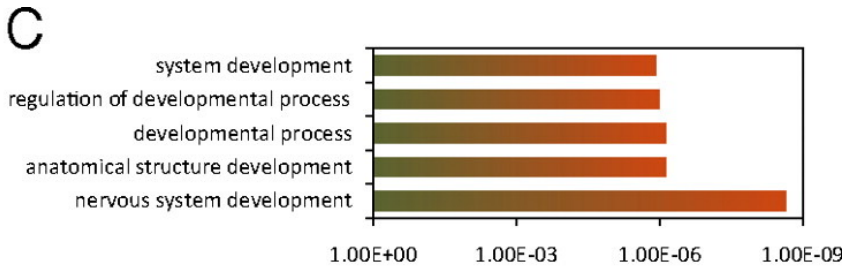
distal enhancers (13). Finally distal H3K4me1-enriched Stat1 binding sites become selectively activated upon INF-γ treatment of HeLa cells, suggesting that these regions maintain the potential to become active (5). This raises the intriguing possibility that enhancers contain information about the current and future developmental potential of a cell, as well as its ability to respond to external cues.

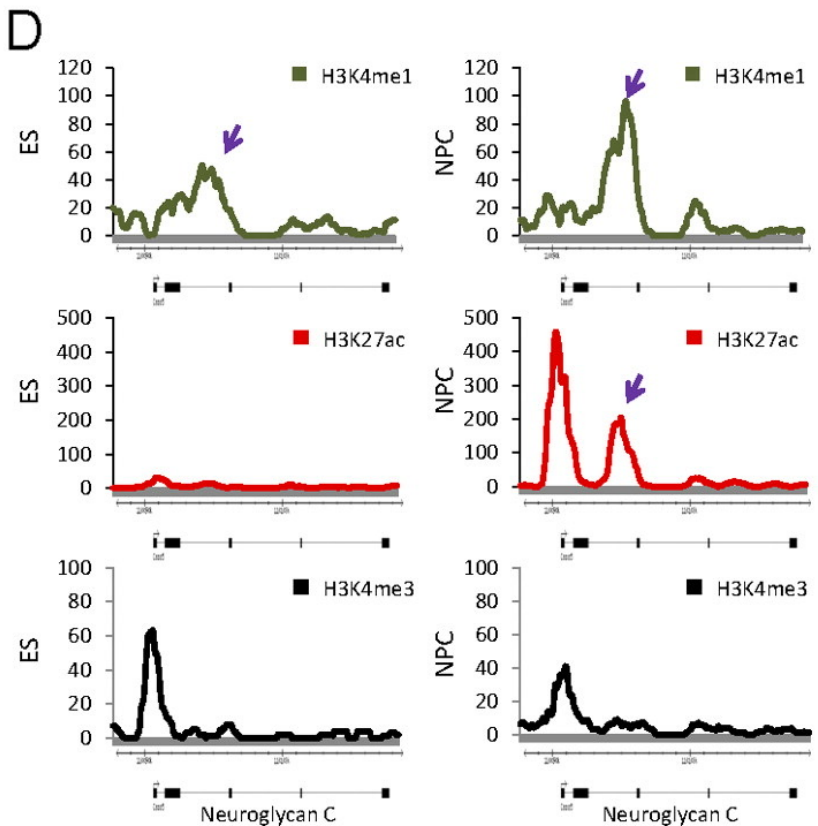
Here, we identify close to 135,000 candidate distal enhancer regions in five different cell types. In a candidate approach to identify factors that could potentially discriminate between the various possible enhancer states, we establish H3K27ac as an important enhancer mark that can distinguish between active and poised enhancer elements. This mark can be deposited by both p300 and CREB binding protein (CBP) (14) and is associated with active promoters in mammalian cells (15). Using this modification to distinguish poised from active enhancer networks allows us to make predictions on the developmental potency of the cell, as well as it current state. Furthermore, it suggests that the active global enhancer network of a cell is smaller than previously anticipated.

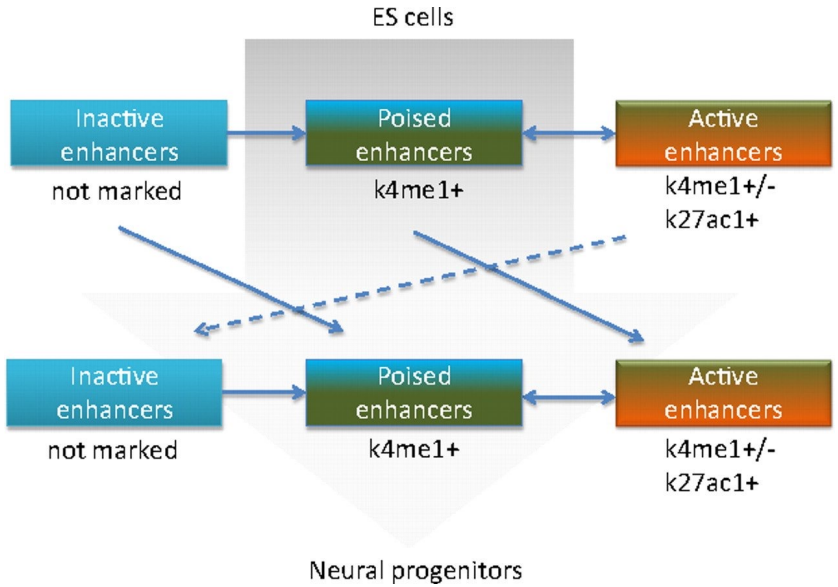
Histone modifications at enhancers





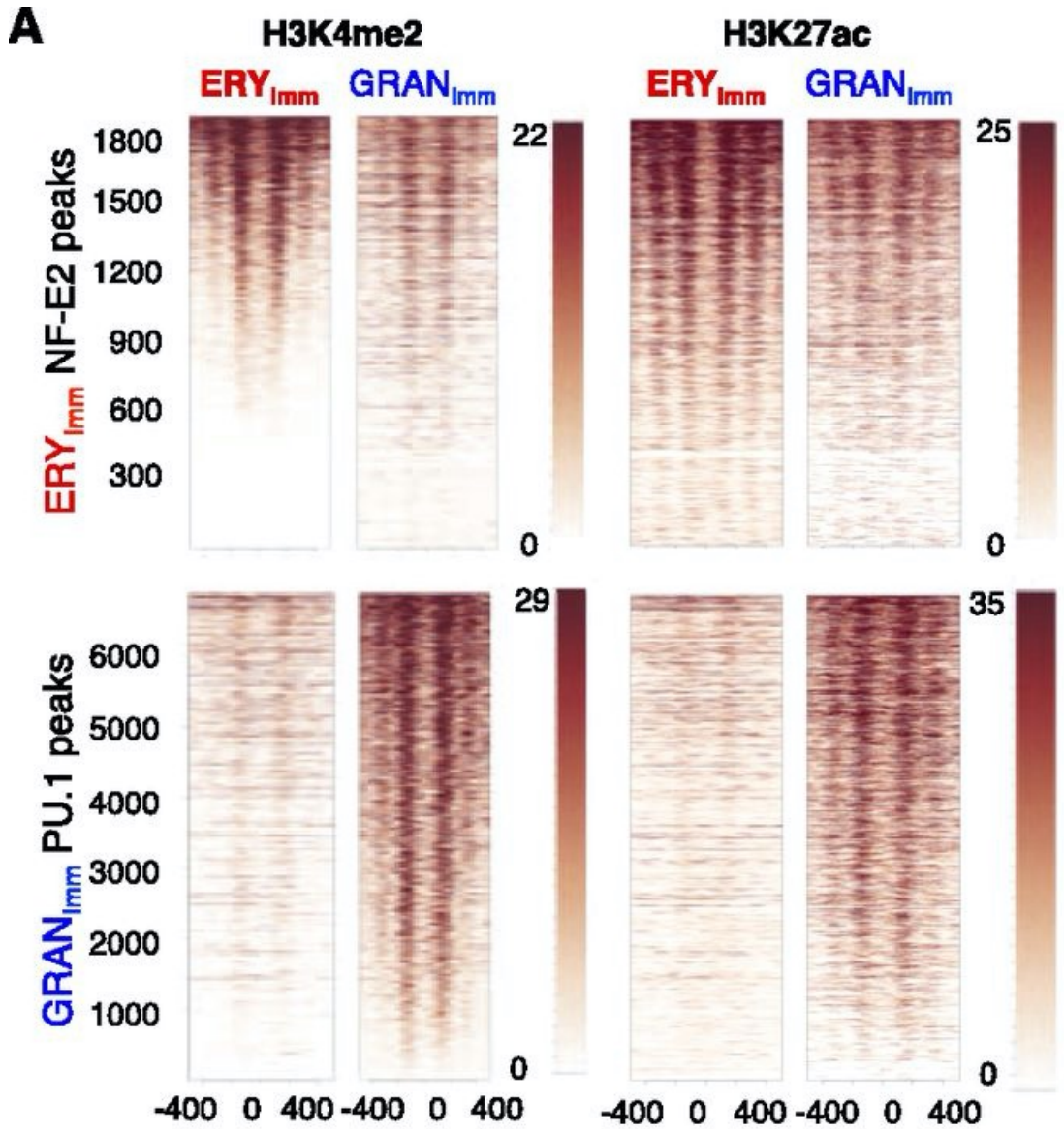


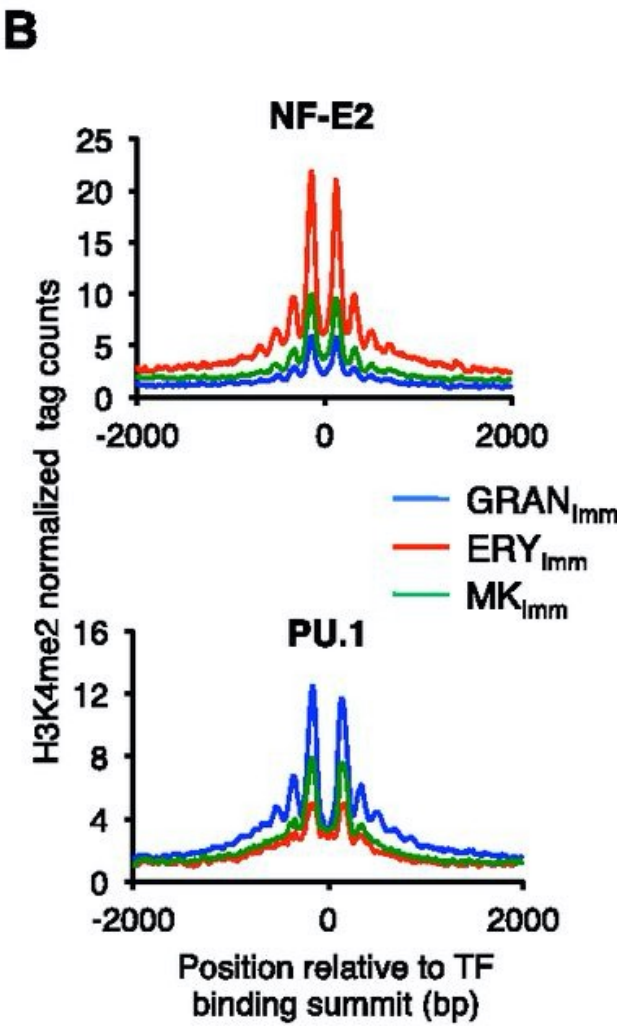




Creyghton et al. *PNAS* 2010

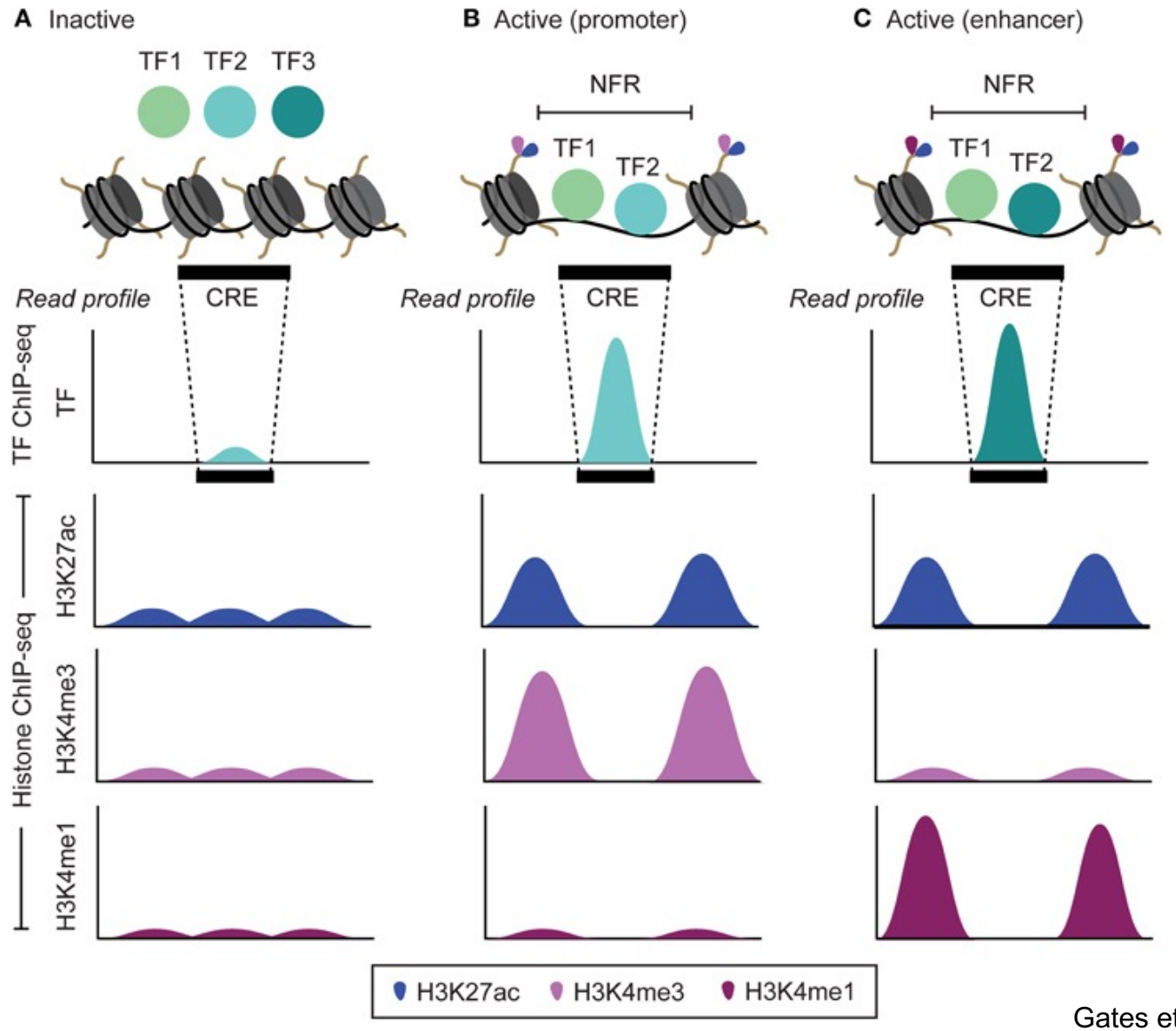
Histone modifications at enhancers





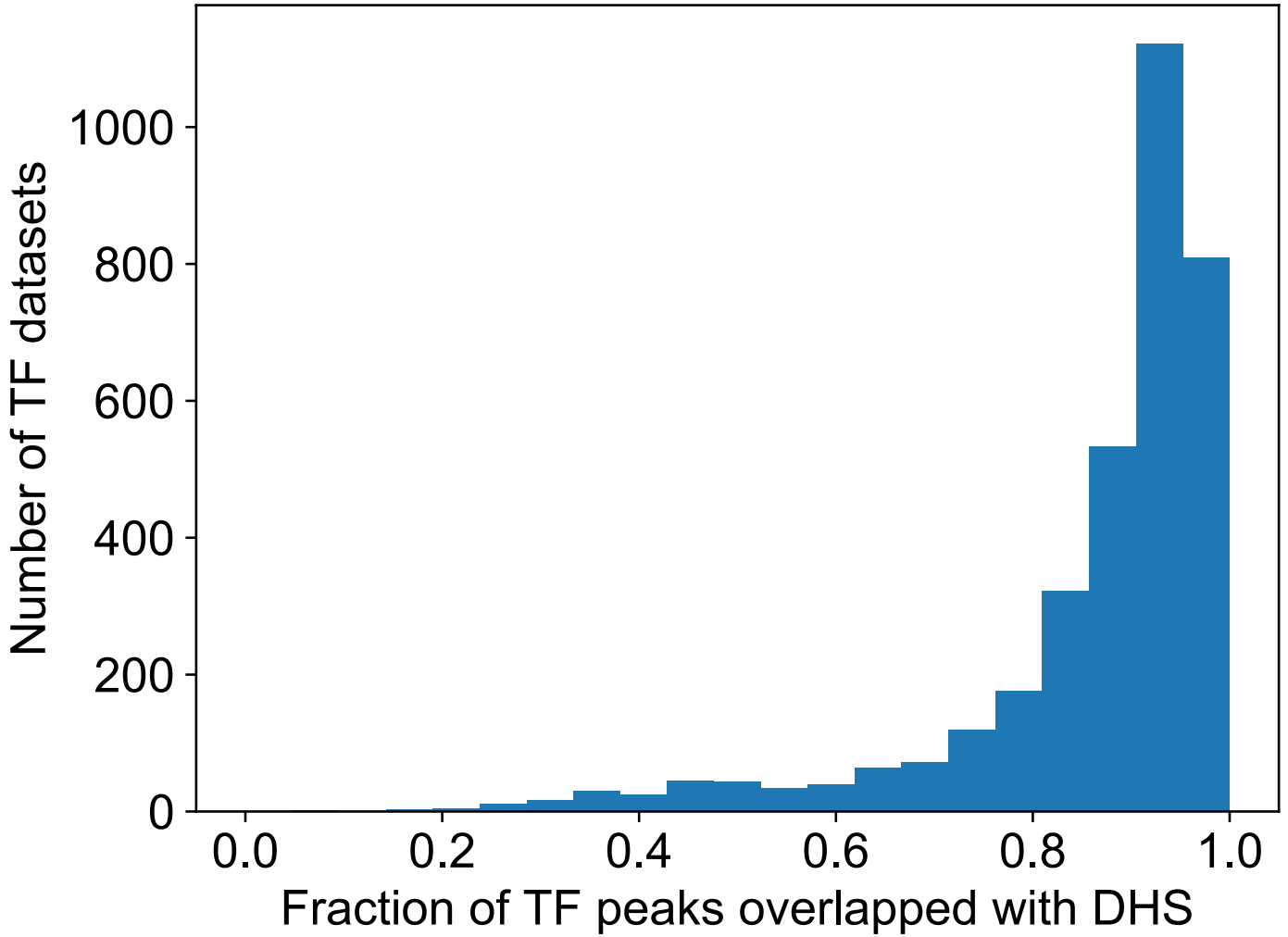
Luyten, Zang et al. Genes Dev 2014

Histone modifications at enhancers



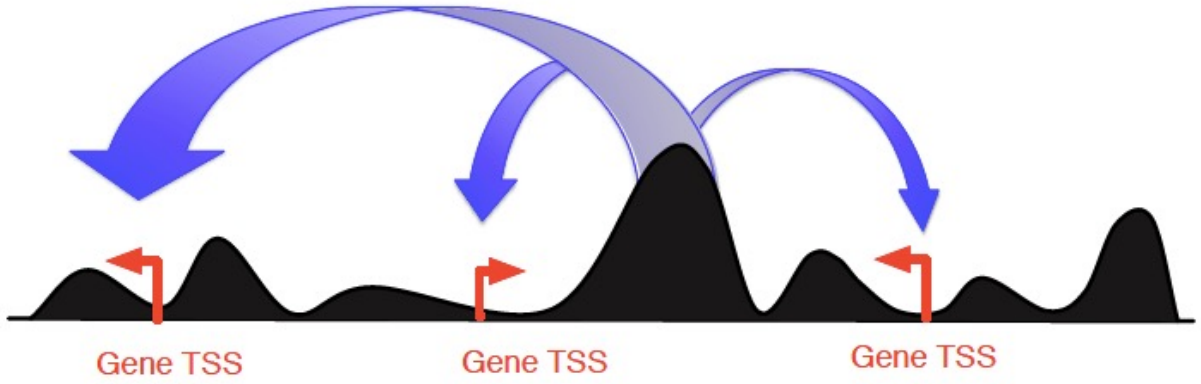
Gates et al. Trends in Biomed Sci *PNAS* 2017

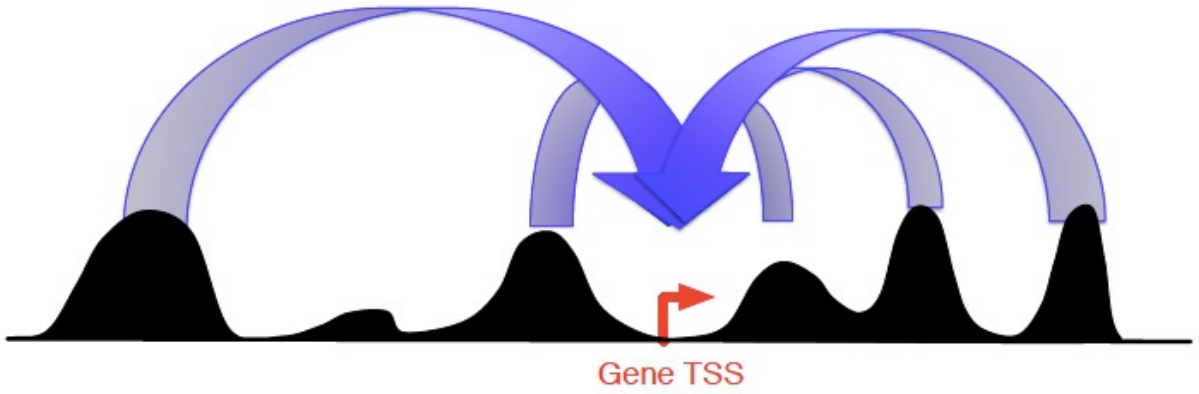
Most transcription factors primarily bind at enhancers with accessible chromatin



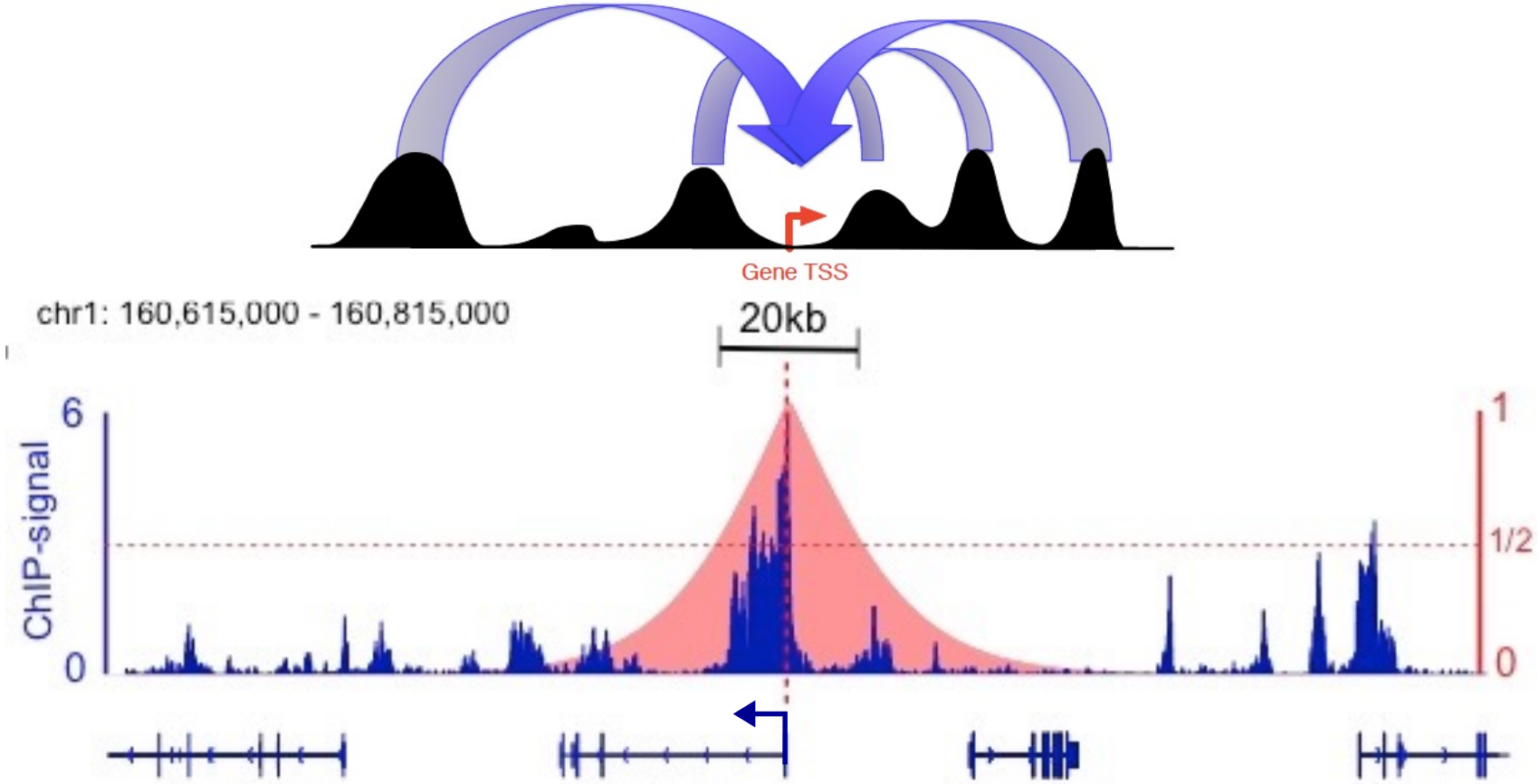
Wang et al. *Bioinformatics* 2018

Enhancers make a diverse and robust transcription program





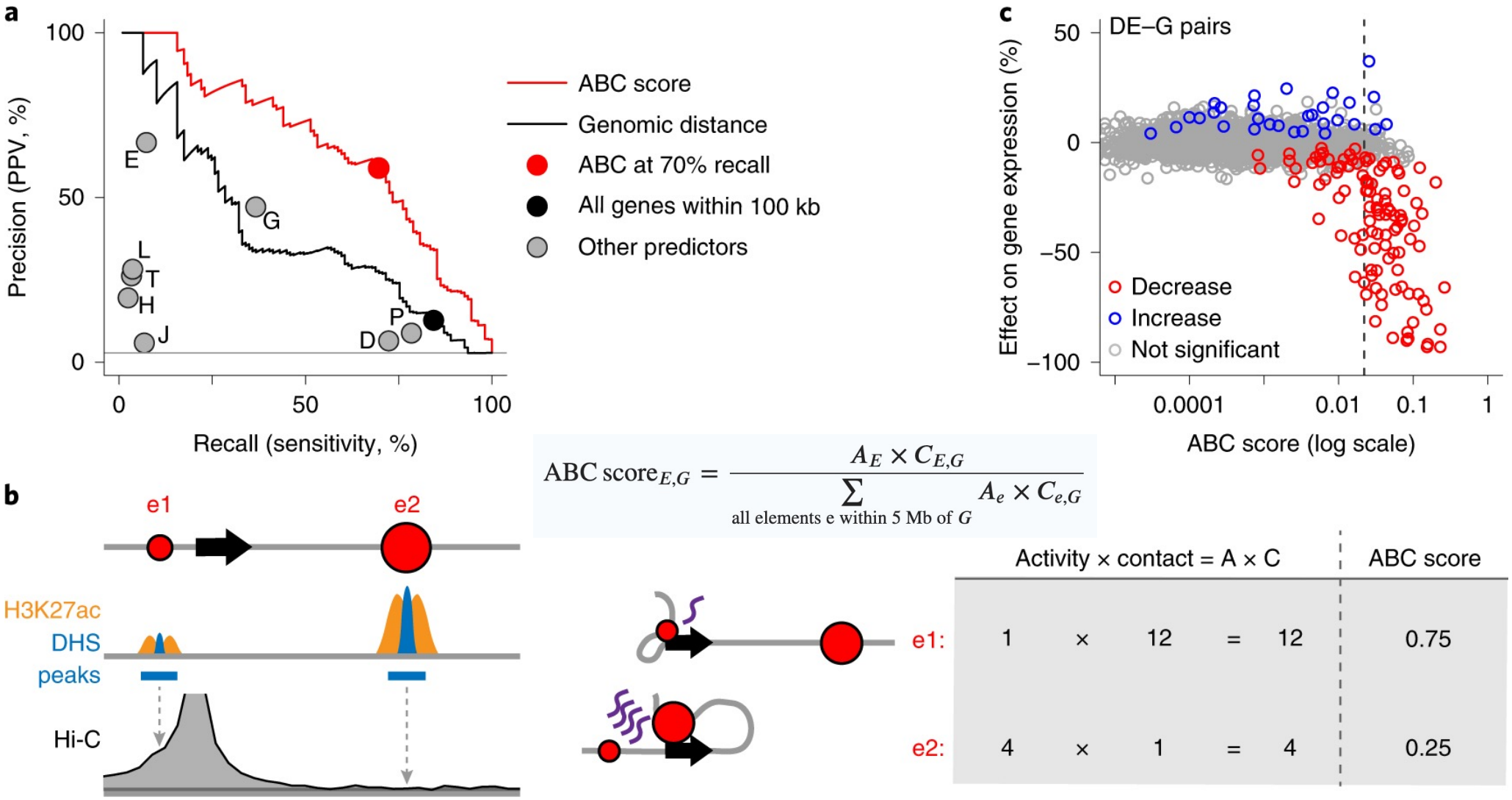
How to associate enhancers with target genes



$$P_i = \sum_{|j|<10^5} W_j Z_j$$

Wang et al. *Genome Res* 2016

How to associate enhancers with target genes



$$\text{ABC score}_{E,G} = \frac{A_E \times C_{E,G}}{\sum_{\text{all elements e within 5 Mb of } G} A_e \times C_{e,G}}$$

	Activity × contact = A × C					ABC score
e1:	1	×	12	=	12	0.75
e2:	4	×	1	=	4	0.25

Fulco et al. *Nat Genet* 2019

Super-enhancers (2013)

Cell

Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes

Warren A. Whyte,[1,4] David A. Orlando,[1,4] Denes Hnisz,[1,4] Brian J. Abraham,[1] Charles Y. Lin,[1,2] Michael H. Kagey,[1] Peter B. Rahl,[1] Tong Ihn Lee,[1] and Richard A. Young[1,3,*]
[1]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA
[2]Department of Medical Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02115, USA
[3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4]These authors contributed equally to this work
*Correspondence: young@wi.mit.edu
http://dx.doi.org/10.1016/j.cell.2013.03.035

SUMMARY

Master transcription factors Oct4, Sox2, and Nanog bind enhancer elements and recruit Mediator to activate much of the gene expression program of pluripotent embryonic stem cells (ESCs). We report here that the ESC master transcription factors form unusual enhancer domains at most genes that control the pluripotent state. These domains, which we call super-enhancers, consist of clusters of enhancers that are densely occupied by the master regulators and Mediator. Super-enhancers differ from typical enhancers in size, transcription factor density and content, ability to activate transcription, and sensitivity to perturbation. Reduced levels of Oct4 or Mediator cause preferential loss of expression of super-enhancer-associated genes relative to other genes, suggesting how changes in gene expression programs might be accomplished during development. In other more differentiated cells, super-enhancers containing cell-type-specific master transcription factors are also found at genes that define cell identity. Super-enhancers thus play key roles in the control of mammalian cell identity.

INTRODUCTION

Transcription factors typically regulate gene expression by binding *cis*-acting regulatory elements known as enhancers and by recruiting coactivators and RNA polymerase II (RNA Pol II) to target genes (Lelli et al., 2012; Ong and Corces, 2011). Enhancers are segments of DNA that are generally a few hundred base pairs in length and are typically occupied by multiple transcription factors (Carey, 1998; Levine and Tjian, 2003; Panne, 2008; Spitz and Furlong, 2012).

Much of the transcriptional control of mammalian development is due to the diverse activity of transcription-factor-bound enhancers that control cell-type-specific patterns of gene expression (Bulger and Groudine, 2011; Hawrylycz et al., 2012; Maston et al., 2006). Between 400,000 and 1.4 million putative enhancers have been identified in the mammalian genome by using a variety of high-throughput techniques that detect features of enhancers such as specific histone modifications (Dunham et al., 2012; Thurman et al., 2012). The number of enhancers that are active in any one cell type has been estimated to be in the tens of thousands, and enhancer activity is largely cell-type specific (Dunham et al., 2012; Heintzman et al., 2009; Shen et al., 2012; Visel et al., 2009; Yip et al., 2012).

In embryonic stem cells (ESCs), control of the gene expression program that establishes and maintains ESC state is dependent on a remarkably small number of master transcription factors (Ng and Surani, 2011; Orkin and Hochedlinger, 2011; Young, 2011). These transcription factors, which include Oct4, Sox2, and Nanog (OSN), bind to enhancers together with the Mediator coactivator complex (Kagey et al., 2010). The Mediator complex facilitates the ability of enhancer-bound transcription factors to recruit RNA Pol II to the promoters of target genes (Borggrefe and Yue, 2011; Conaway and Conaway, 2011; Kornberg, 2005; Malik and Roeder, 2010) and is essential for maintenance of ESC state and embryonic development (Ito et al., 2000; Kagey et al., 2010; Risley et al., 2010).

ESCs are highly sensitive to reduced levels of Mediator. Indeed, reductions in the levels of many subunits of Mediator cause the same rapid loss of ESC-specific gene expression as loss of Oct4 and other master transcription factors (Kagey et al., 2010). It is unclear why reduced levels of Mediator, a general coactivator, can phenocopy the effects of reduced levels of Oct4 in ESCs.

Interest in further understanding the importance of Mediator in ESCs led us to further investigate enhancers bound by the master transcription factors and Mediator in these cells. We found that much of enhancer-associated Mediator occupies exceptionally large enhancer domains and that these domains are associated with genes that play prominent roles in ESC biology. These large domains, or super-enhancers, were found to contain high levels of the key ESC transcription factors Oct4, Sox2, Nanog, Klf4, and Esrrb to stimulate higher transcriptional activity than typical enhancers and to be exceptionally sensitive to

Cell

Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers

Jakob Lovén,[1,7] Heather A. Hoke,[1,2,7] Charles Y. Lin,[1,3,5,7] Ashley Lau,[1,2] David A. Orlando,[1] Christopher R. Vakoc,[4] James E. Bradner,[5,6] Tong Ihn Lee,[1] and Richard A. Young[1,2,*]
[1]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA
[2]Department of Biology
[3]Computational and Systems Biology Program
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4]Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA
[5]Department of Medical Oncology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA
[6]Department of Medicine, Harvard Medical School, 25 Shattuck Street, Boston, MA 02115, USA
[7]These authors contributed equally to this work
*Correspondence: young@wi.mit.edu
http://dx.doi.org/10.1016/j.cell.2013.03.036

SUMMARY

Chromatin regulators have become attractive targets for cancer therapy, but it is unclear why inhibition of these ubiquitous regulators should have gene-specific effects in tumor cells. Here, we investigate how inhibition of the widely expressed transcriptional coactivator BRD4 leads to selective inhibition of the *MYC* oncogene in multiple myeloma (MM). BRD4 and Mediator were found to co-occupy thousands of enhancers associated with active genes. They also co-occupied a small set of exceptionally large super-enhancers associated with genes that feature prominently in MM biology, including the *MYC* oncogene. Treatment of MM tumor cells with the BET-bromodomain inhibitor JQ1 led to preferential loss of BRD4 at super-enhancers and consequent transcription elongation defects that preferentially impacted genes with super-enhancers, including *MYC*. Super-enhancers were found at key oncogenic drivers in many other tumor cells. These observations have implications for the discovery of cancer therapeutics directed at components of super-enhancers in diverse tumor types.

INTRODUCTION

Chromatin regulators are attractive as therapeutic targets for cancer because they are deregulated in numerous cancers (Baylin and Jones, 2011; Elsässer et al., 2011; Esteller, 2008; Feinberg and Tycko, 2004; You and Jones, 2012) and are amenable to small-molecule inhibition (Cole, 2008; Dawson and Kouzarides, 2012; Geutjes et al., 2012). Inhibition of some chromatin regulators has already proven to be efficacious for treatment of certain cancers (Issa and Kantarjian, 2009; Marks and Xu, 2009). Most chromatin regulators, however, are expressed in a broad range of healthy cells and contribute generally to gene expression, so inhibition of these important genome-associated proteins might be expected to adversely affect global gene expression in healthy cells and thus produce highly toxic effects. Nonetheless, inhibitors of some chromatin regulators, such as BRD4, have been shown to selectively inhibit transcription of key oncogenic drivers such as *c-MYC* (hereafter referred to as *MYC*) in multiple tumor types (Dawson et al., 2011; Delmore et al., 2011; Mertz et al., 2011; Zuber et al., 2011). It is important to understand how inhibition of a widely expressed, general regulator such as BRD4 can exert a selective effect on the expression of a small number of genes in specific cells.

BRD4 is a member of the bromodomain and extraterminal (BET) subfamily of human bromodomain proteins, which includes BRDT, BRD2, BRD3, and BRD4. These proteins associate with acetylated chromatin and facilitate transcriptional activation (LeRoy et al., 2008; Rahman et al., 2011). BRD4 was first identified as an interaction partner of the murine Mediator coactivator complex (Jiang et al., 1998) and was subsequently shown to associate with Mediator in a variety of human cells (Dawson et al., 2011; Wu and Chiang, 2007). BRD4 is involved in the control of transcriptional elongation by RNA polymerase II (RNA Pol II) through its recruitment of the positive transcription elongation factor P-TEFb (Jang et al., 2005; Yang et al., 2005). Almost all human cells express the *BRD4* gene, based on analysis of human tissue expression data across 90 distinct tissue types (human body index - transcriptional profiling, see Extended Experimental Procedures), and BRD4 is found to be associated with a large population of active genes in $CD4^+$ T cells (Zhang et al., 2012). It is not yet clear whether the BRD4 protein is generally involved in the transcription of active genes in tumor cells or if it is selectively associated with a subset of these genes.

Cell

Resource

Super-Enhancers in the Control of Cell Identity and Disease

Denes Hnisz,[1,3] Brian J. Abraham,[1,3] Tong Ihn Lee,[1,3] Ashley Lau,[1,2] Violaine Saint-André,[1] Alla A. Sigova,[1] Heather A. Hoke,[1,2] and Richard A. Young[1,2,*]
[1]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA
[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3]These authors contributed equally to this work
*Correspondence: young@wi.mit.edu
http://dx.doi.org/10.1016/j.cell.2013.09.053

SUMMARY

Super-enhancers are large clusters of transcriptional enhancers that drive expression of genes that define cell identity. Improved understanding of the roles that super-enhancers play in biology would be afforded by knowing the constellation of factors that constitute these domains and by identifying super-enhancers across the spectrum of human cell types. We describe here the population of transcription factors, cofactors, chromatin regulators, and transcription apparatus occupying super-enhancers in embryonic stem cells and evidence that super-enhancers are highly transcribed. We produce a catalog of super-enhancers in a broad range of human cell types and find that super-enhancers associate with genes that control and define the biology of these cells. Interestingly, disease-associated variation is especially enriched in the super-enhancers of disease-relevant cell types. Furthermore, we find that cancer cells generate super-enhancers at oncogenes and other genes important in tumor pathogenesis. Thus, super-enhancers play key roles in human cell identity in health and in disease.

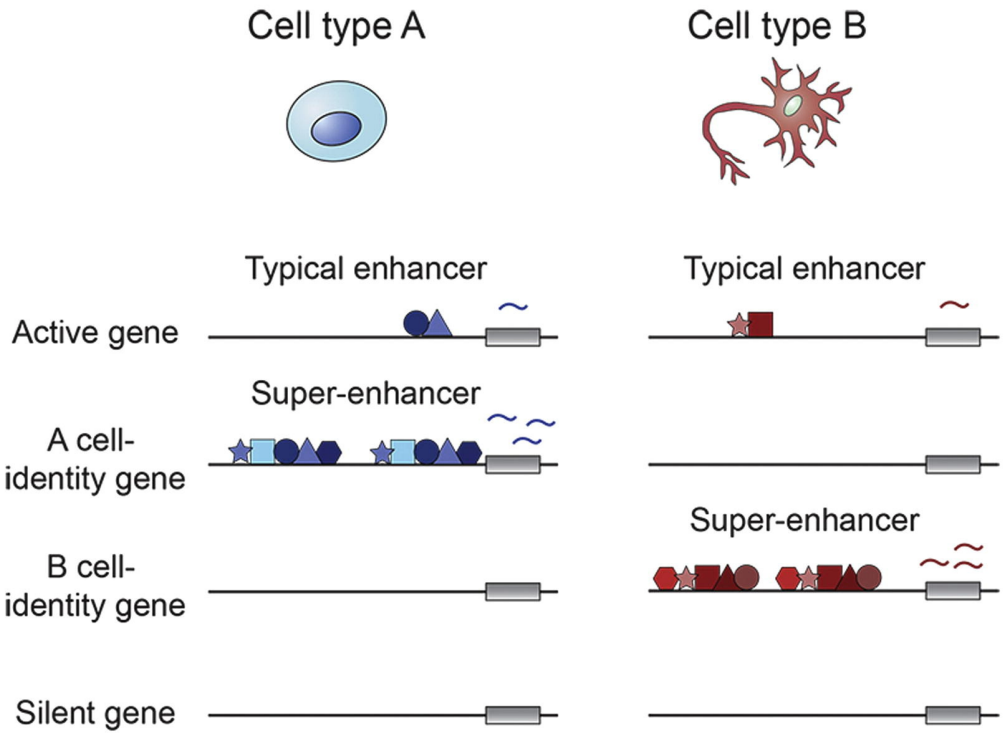
INTRODUCTION

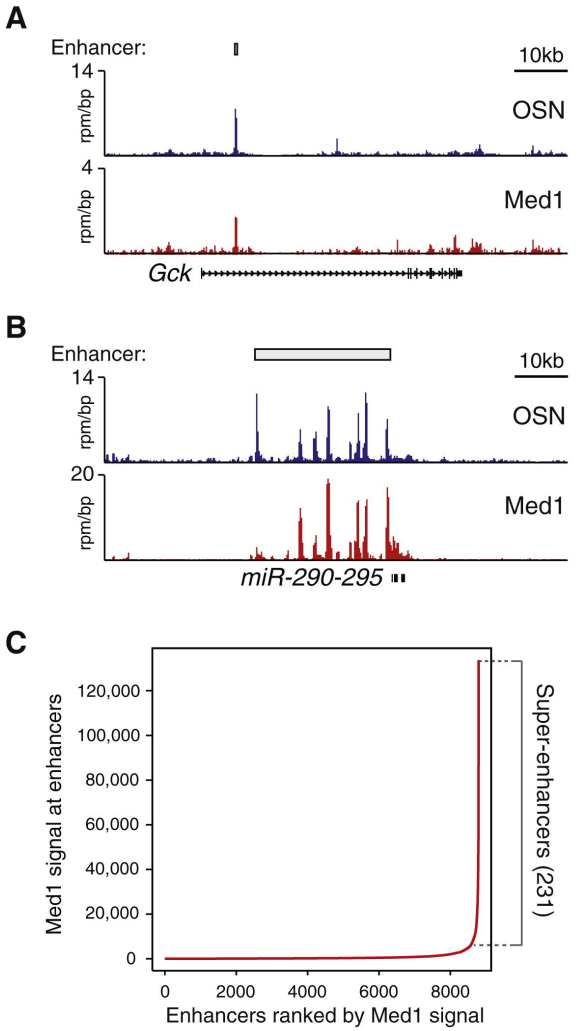
Transcription factors bind DNA regulatory elements called enhancers, which play key roles in the control of cell-type-specific gene expression programs (Bulger and Groudine, 2011; Calo and Wysocka, 2013; Carey, 1998; Lelli et al., 2012; Levine and Tjian, 2003; Maston et al., 2006; Ong and Corces, 2011; Panne, 2008; Spitz and Furlong, 2012; Xie and Ren, 2013). A typical mammalian cell contains thousands of active enhancers, and it has been estimated that there may be ~1 million enhancers active in all human cells (Bernstein et al., 2012; Heintzman et al., 2009; Thurman et al., 2012). It is important to further understand enhancers and their components because they control specific gene expression programs, and much disease-associated sequence variation occurs in these regulatory elements (Grossman et al., 2013; Hindorff et al., 2009; Lee and Young, 2013; Maurano et al., 2012).

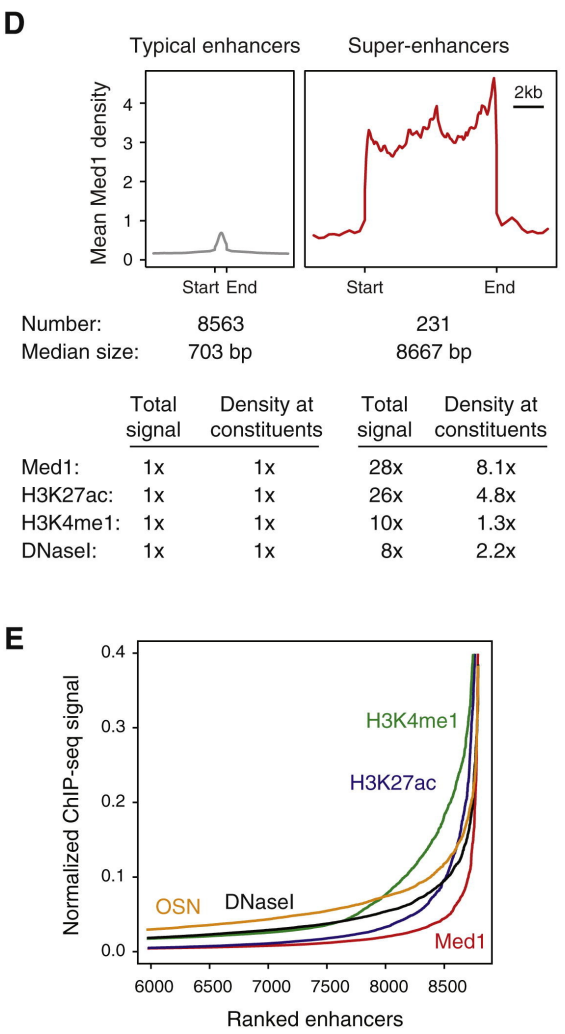
The set of enhancers that control any one cell's gene expression program is probably best defined in murine embryonic stem cells (ESCs). Co-occupancy of murine ESC genomic sites by the master transcription factors Oct4, Sox2, and Nanog is highly predictive of enhancer activity (Chen et al., 2008), and 8,794 enhancers have been identified in ESCs by using ChIP-seq data sets for Oct4, Sox2, and Nanog (Whyte et al., 2013). A subset of these enhancers forms 231 unusual enhancer domains at most genes that control the pluripotent state; these super-enhancers consist of clusters of enhancers that are densely occupied by five key ESC transcription factors and the Mediator coactivator (Whyte et al., 2013). There are many additional transcription factors, cofactors, and chromatin regulators that contribute to the control of ESCs (Ng and Surani, 2011; Orkin and Hochedlinger, 2011; Young, 2011), and it would be instructive to know how these occupy enhancers and super-enhancers in ESCs. Similarly, it would be useful to know if super-enhancers are transcribed, as enhancer RNAs (eRNAs) have been proposed to contribute to enhancer activity (Lai et al., 2013; Lam et al., 2013; Li et al., 2013; Ling et al., 2004; Mousavi et al., 2013; Ørom et al., 2010).

Super-enhancers are associated with key genes that control cell state in cells where they have been identified thus far, so identification of these domains in additional cell types could provide a valuable resource for further study of cellular control. We have generated a catalog of super-enhancers in 86 human cell and tissue types. These super-enhancers are associated with genes encoding cell-type-specific transcription factors and thus identify candidate master transcription factors for many cell types that should prove useful for further understanding transcriptional control of cell state and for reprogramming studies. Using this catalog, we find that DNA sequence variation associated with specific diseases is especially enriched in the super-enhancers of disease-relevant cells, suggesting that hypotheses regarding the role of specific cell types and genes in many diseases might be guided by knowledge of super-enhancers. Furthermore, tumor cells acquire super-enhancers at key oncogenes and at genes that function in the acquisition of hallmark capabilities in cancer, suggesting that these domains provide biomarkers for tumor-specific pathologies that may be valuable for diagnosis and therapeutic intervention.

Super-enhancers regulate cell-identity genes



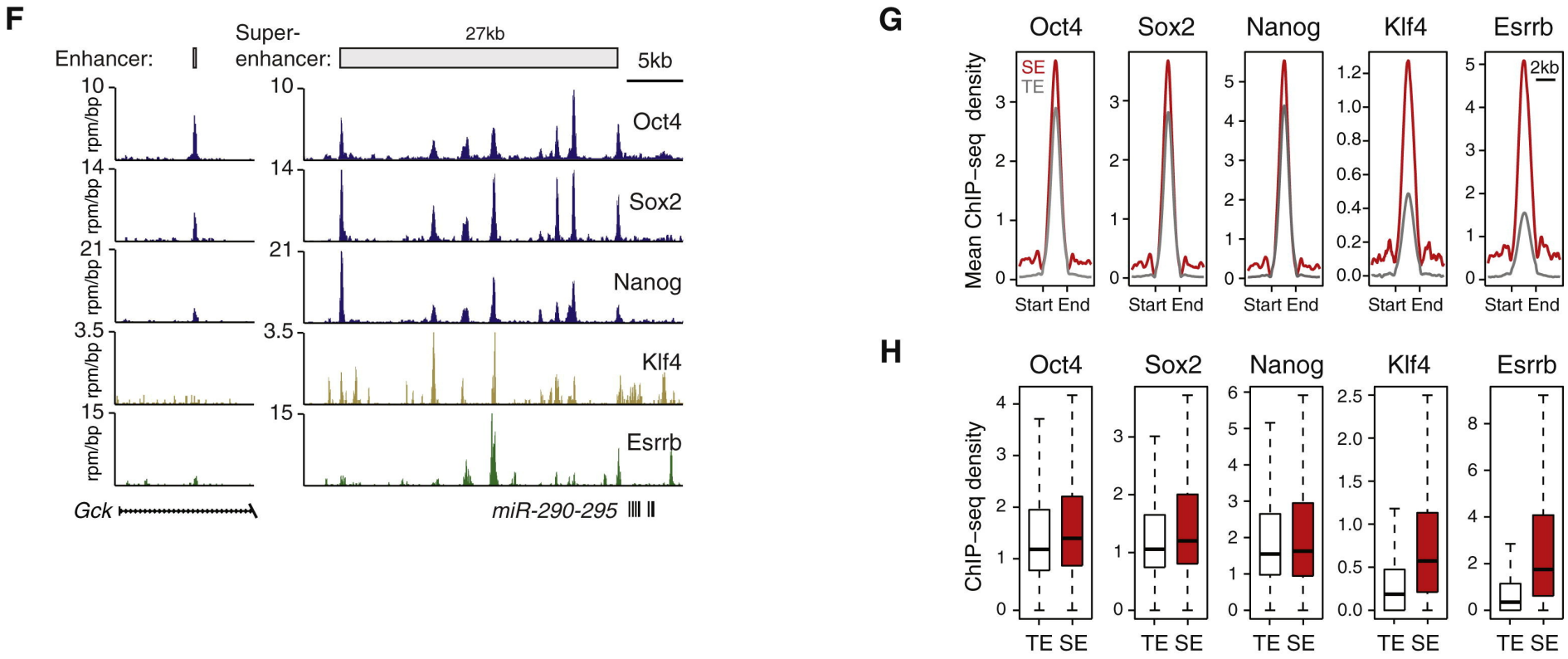




Number:	8563		231	
Median size:	703 bp		8667 bp	
	Total signal	Density at constituents	Total signal	Density at constituents
Med1:	1x	1x	28x	8.1x
H3K27ac:	1x	1x	26x	4.8x
H3K4me1:	1x	1x	10x	1.3x
DNaseI:	1x	1x	8x	2.2x

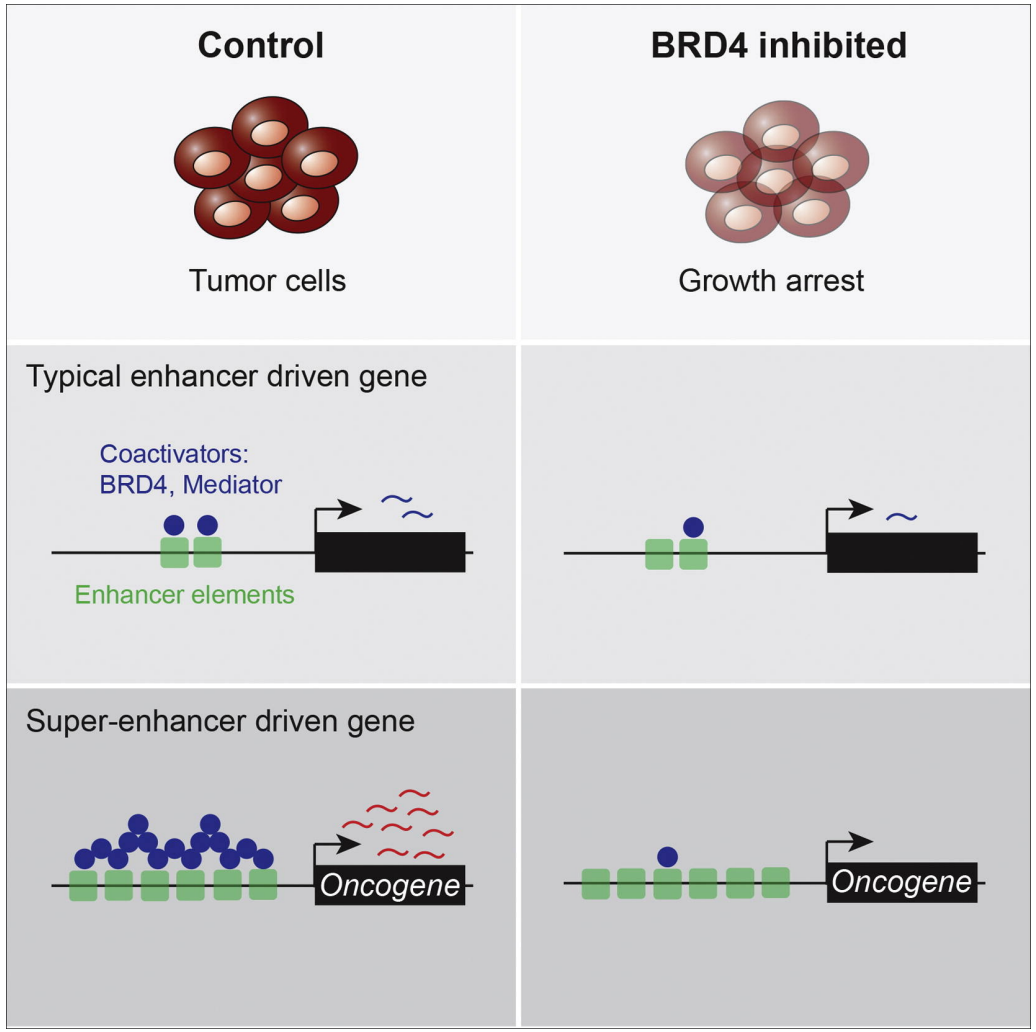
Whyte et al, *Cell* 2013

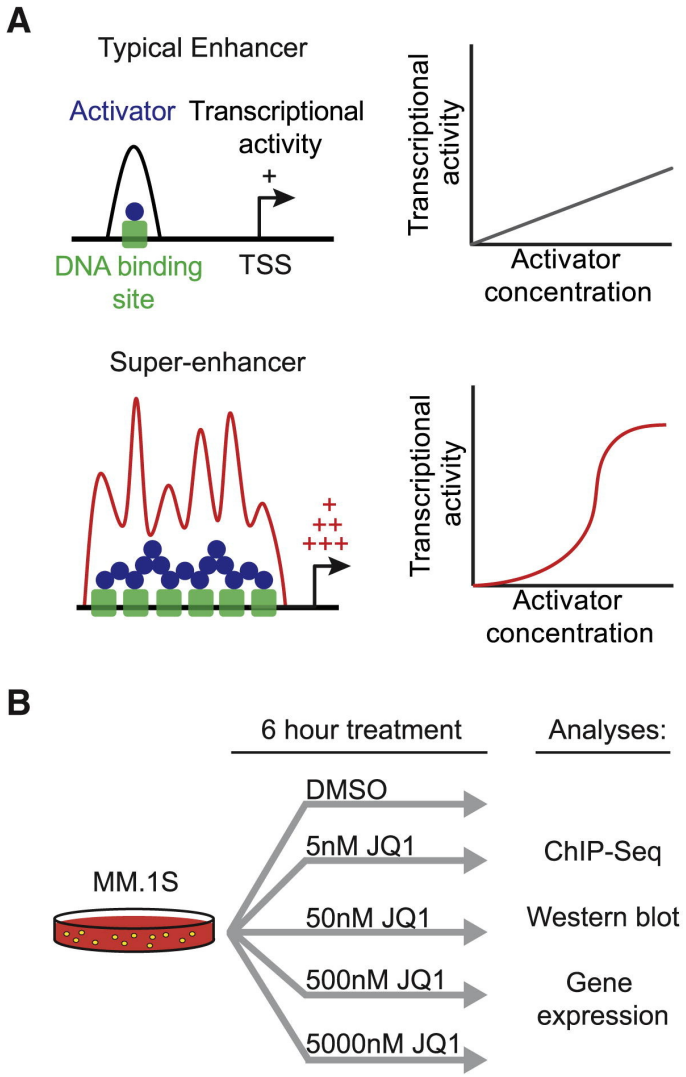
Super-enhancers regulate cell-identity genes



Whyte et al, *Cell* 2013

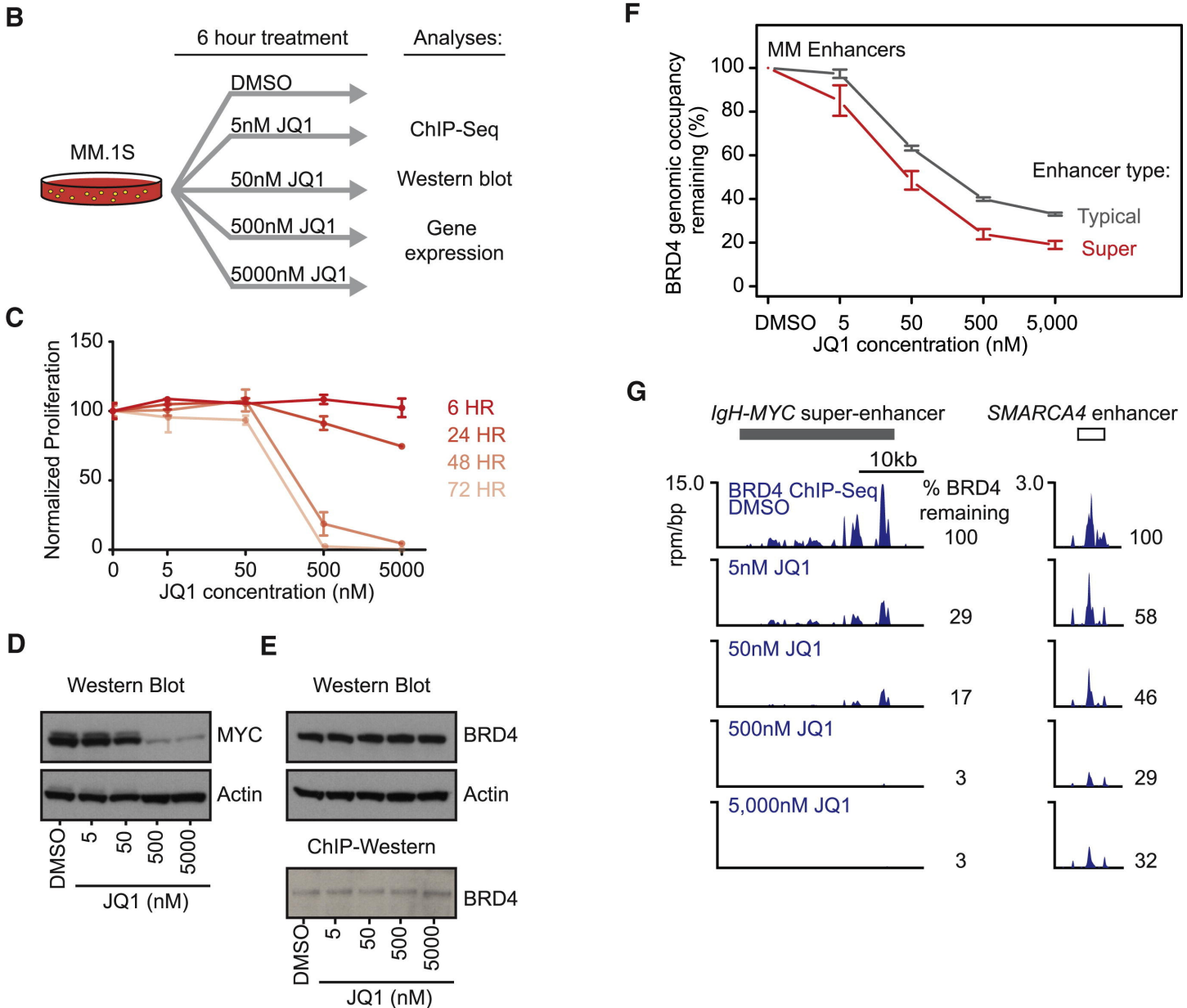
Super-enhancers in cancer





Loven et al, *Cell* 2013

Super-enhancers in cancer are sensitive to bromodomain inhibition



Loven et al, *Cell* 2013

Stretch Enhancers

Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants

Stephen C. J. Parker[a,1], Michael L. Stitzel[a,1], D. Leland Taylor[a], Jose Miguel Orozco[a], Michael R. Erdos[a], Jennifer A. Akiyama[b], Kelly Lammerts van Bueren[c], Peter S. Chines[a], Narisu Narisu[a], NISC Comparative Sequencing Program[a], Brian L. Black[c], Axel Visel[b,d], Len A. Pennacchio[b,d], and Francis S. Collins[a,2]

[a]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892; [b]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; [c]Cardiovascular Research Institute, University of California, San Francisco, CA 95158; and [d]Department of Energy Joint Genome Institute, Walnut Creek, CA 94598

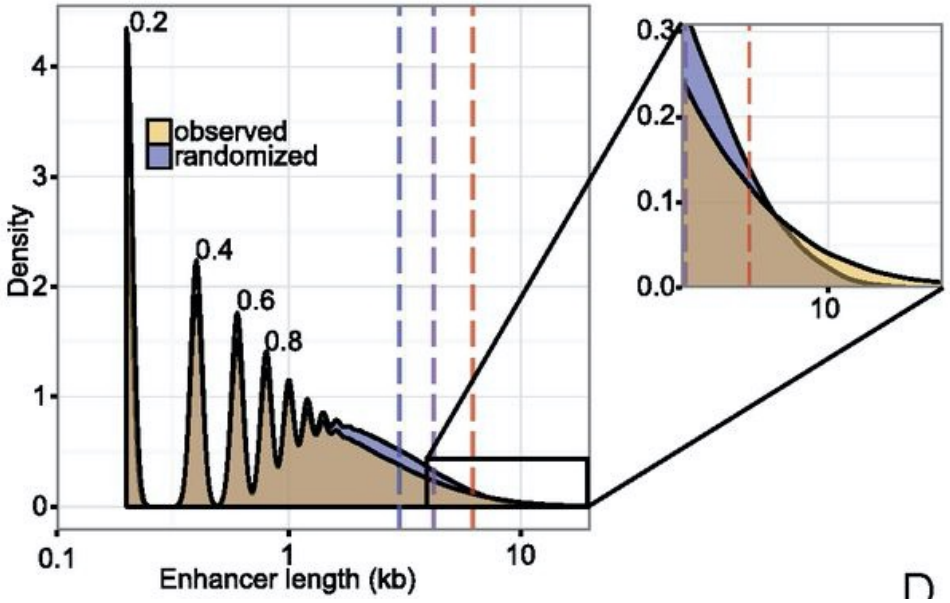
Contributed by Francis S. Collins, September 16, 2013 (sent for review August 2, 2013)

Chromatin-based functional genomic analyses and genomewide association studies (GWASs) together implicate enhancers as critical elements influencing gene expression and risk for common diseases. Here, we performed systematic chromatin and transcriptome profiling in human pancreatic islets. Integrated analysis of islet data with those from nine cell types identified specific and significant enrichment of type 2 diabetes and related quantitative trait GWAS variants in islet enhancers. Our integrated chromatin maps reveal that most enhancers are short (median = 0.8 kb). Each cell type also contains a substantial number of more extended (≥3 kb) enhancers. Interestingly, these stretch enhancers are often tissue-specific and overlap locus control regions, suggesting that they are important chromatin regulatory beacons. Indeed, we show that (*i*) tissue specificity of enhancers and nearby gene expression increase with enhancer length; (*ii*) neighborhoods containing stretch enhancers are enriched for important cell type–specific genes; and (*iii*) GWAS variants associated with traits relevant to a particular cell type are more enriched in stretch enhancers compared with short enhancers. Reporter constructs containing stretch enhancer sequences exhibited tissue-specific activity in cell culture experiments and in transgenic mice. These results suggest that stretch enhancers are critical chromatin elements for coordinating cell type–specific regulatory programs and that sequence variation in stretch enhancers affects risk of major common human diseases.

DNA Elements (ENCODE) cell types to generate consistent chromatin state assignments across all 10 cell types. We anchored these assignments based on overlap with previously published chromatin states (2) in the nine ENCODE cell types to produce a consistent annotation of promoter, enhancer, insulator, transcribed, and repressed chromatin states (*SI Appendix*, Fig. S1). In parallel, we integrated our human islet RNA-seq data with ENCODE RNA-seq data, resulting in a unified set of chromatin state and mRNA maps for islets and the nine ENCODE cell types (Fig. 1*A*). After subsampling to normalize the amount of ChIP-seq reads, the fraction of the genome covered by select chromatin states remained relatively constant across any given cell type (Fig. 1*B*, *Upper*). However, we observed that additional read depth identified additional signal-enriched enhancer regions (*SI Appendix*, Fig. S2), a finding consistent with other studies (10, 11). Thus, in subsequent analyses, we used chromatin states identified using all reads (Fig. 1*B*, *Lower*) and note that the trends reported herein are consistently observed even when normalized read chromatin states are used. As shown in Fig. 1*A*, our integrative approach identified both common (e.g.,

Significance

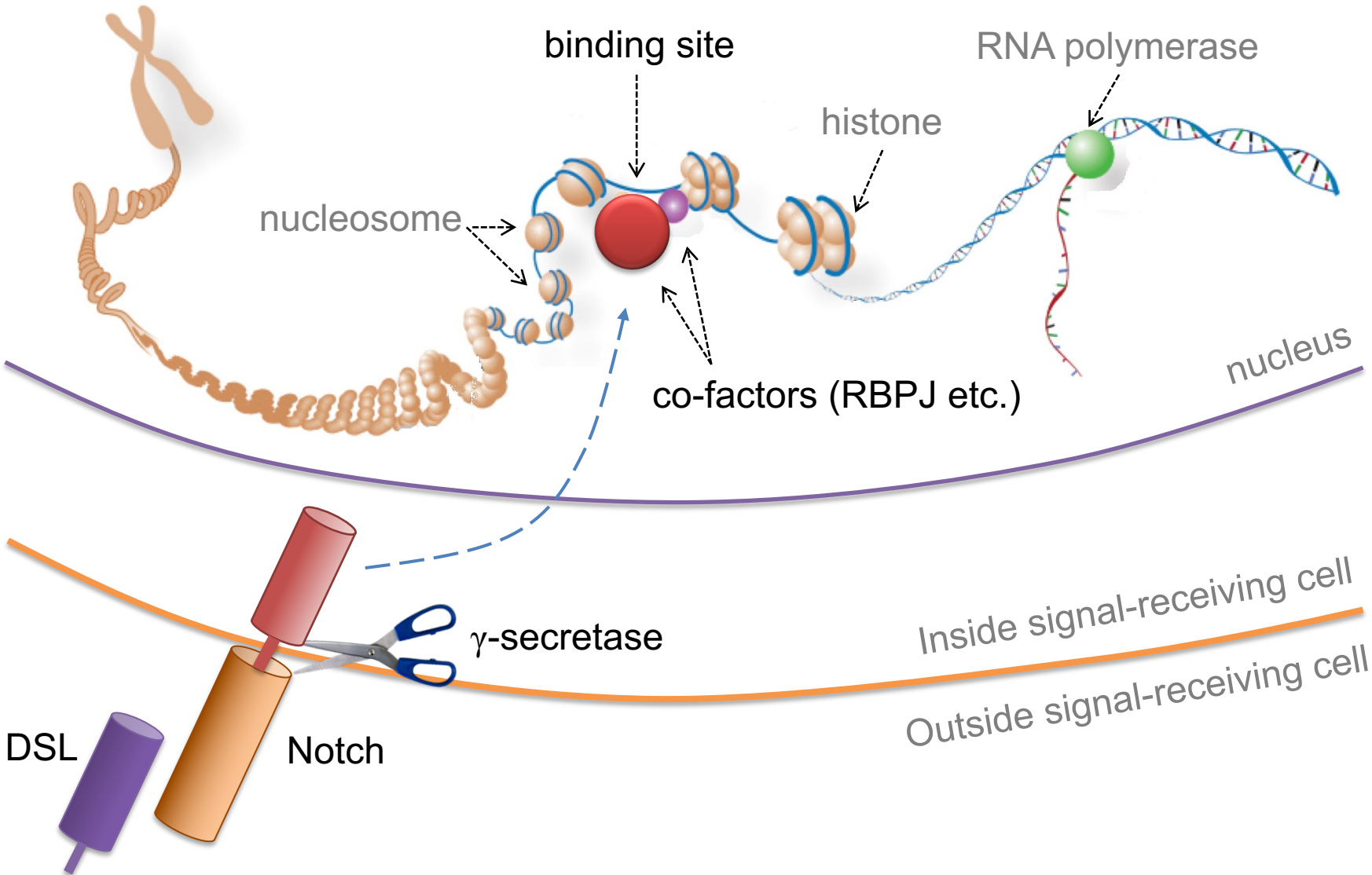
Using high-throughput experiments, we determined the func-



Parker et al. PNAS 2013

PNAS PNAS PNAS
om IP address 128.143.0.10.

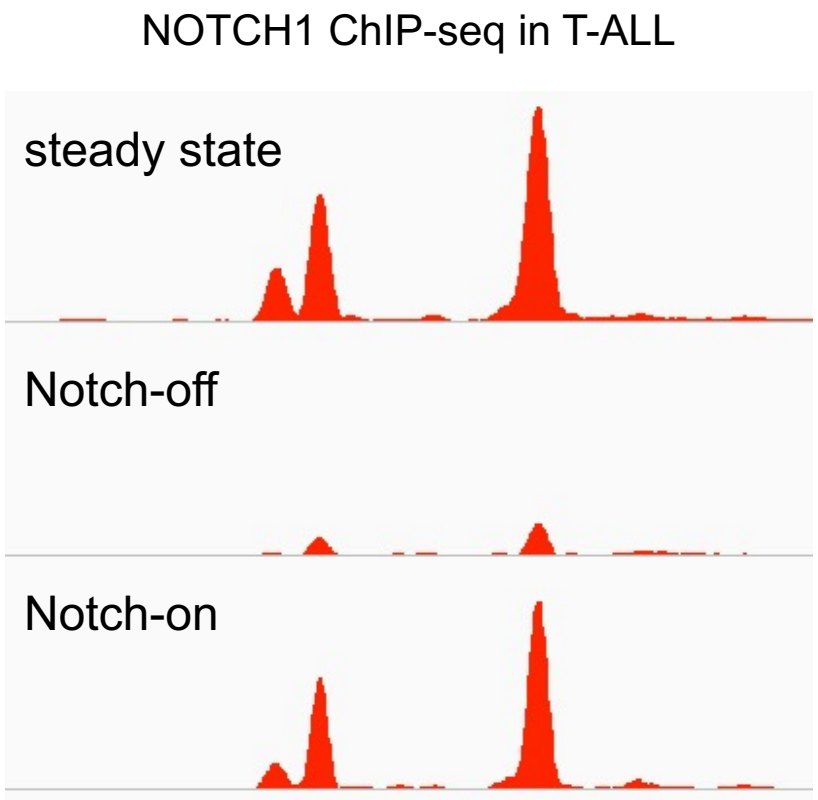
Nuclear Notch acts as a transcription factor



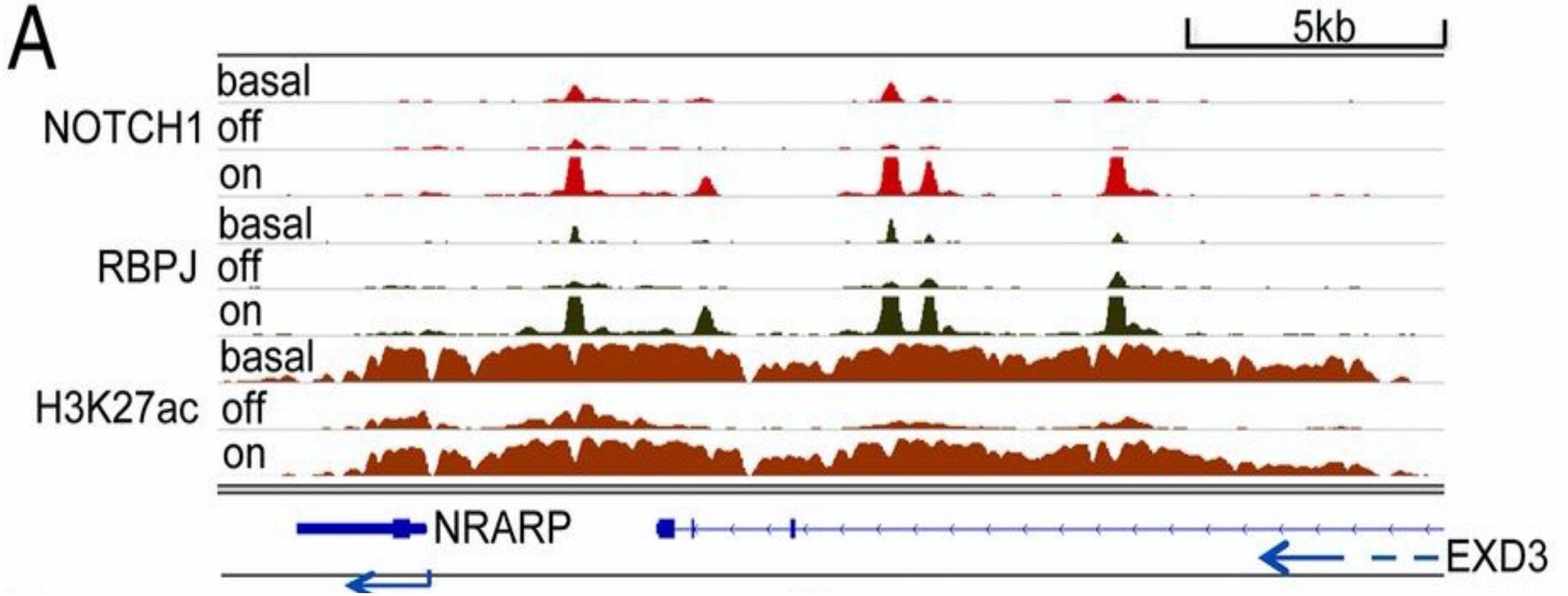
Gamma secretase inhibitor (GSI) to perturb Notch activity

GSI treatment ➔

GSI wash out ➔

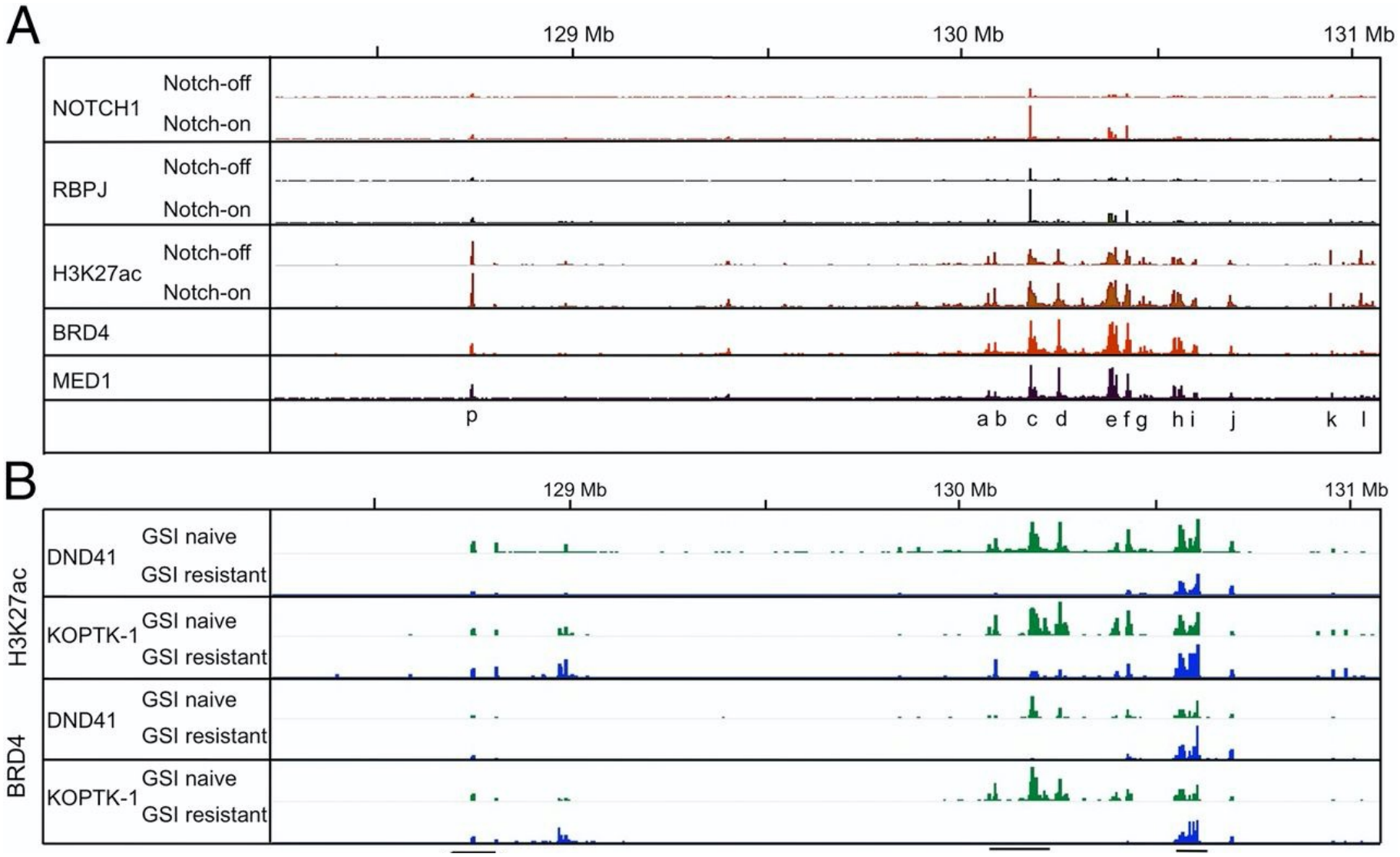


Notch induces super-enhancers in T-ALL



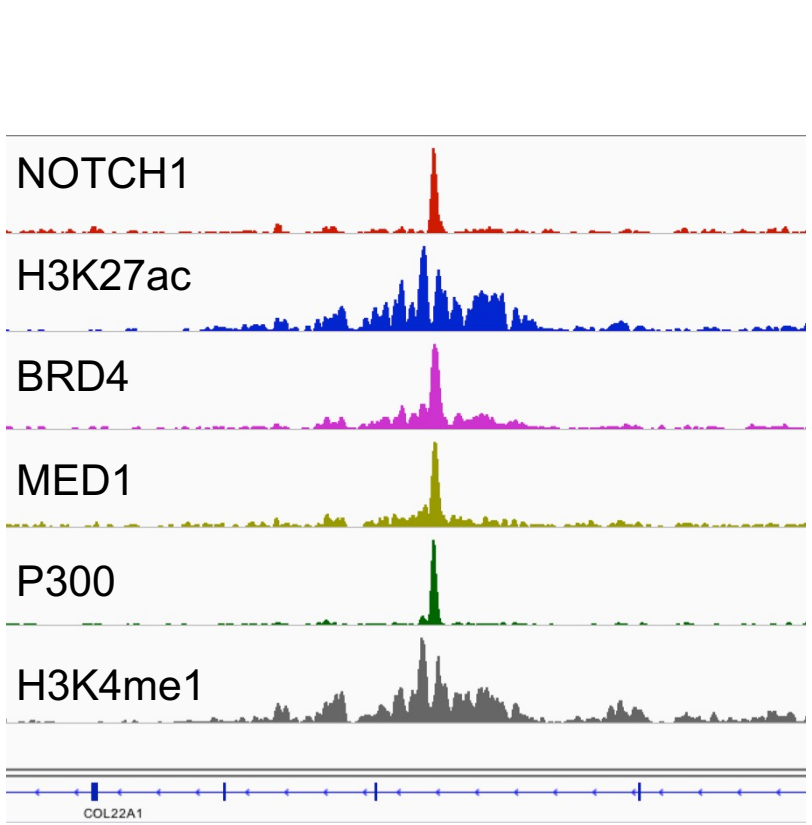
Wang, Zang et al. PNAS 2014

A super-enhancer cluster near MYC locus in T-ALL

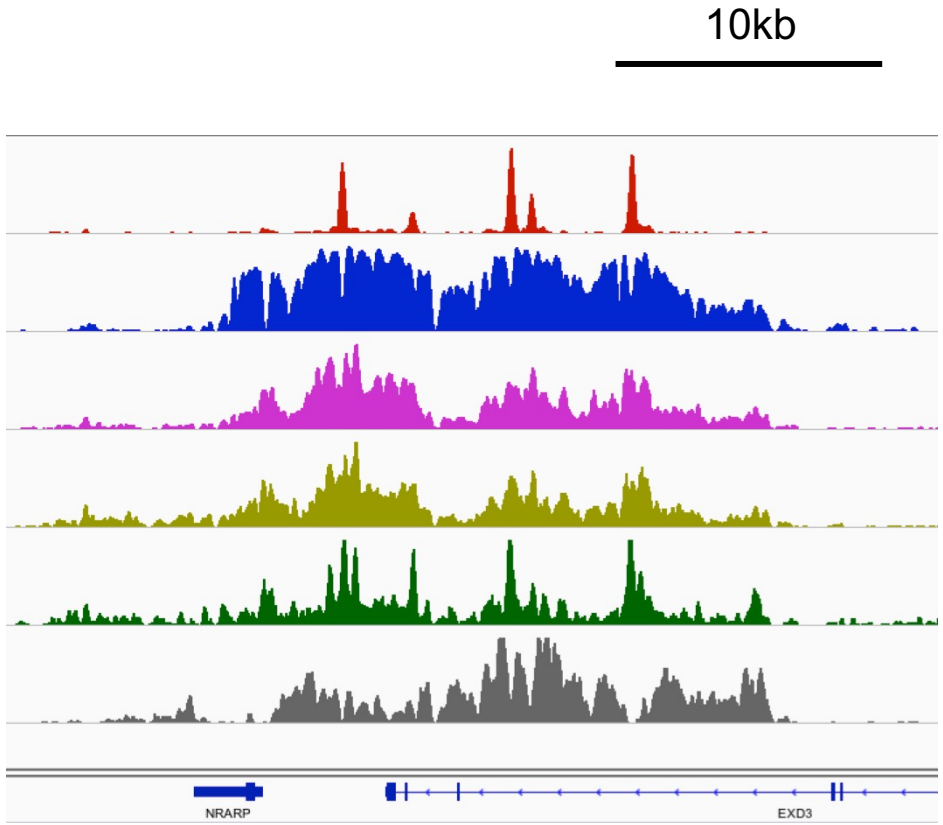


Yashiro-Ohtahi et al. PNAS 2014

Typical enhancers vs. super-enhancers in T-ALL

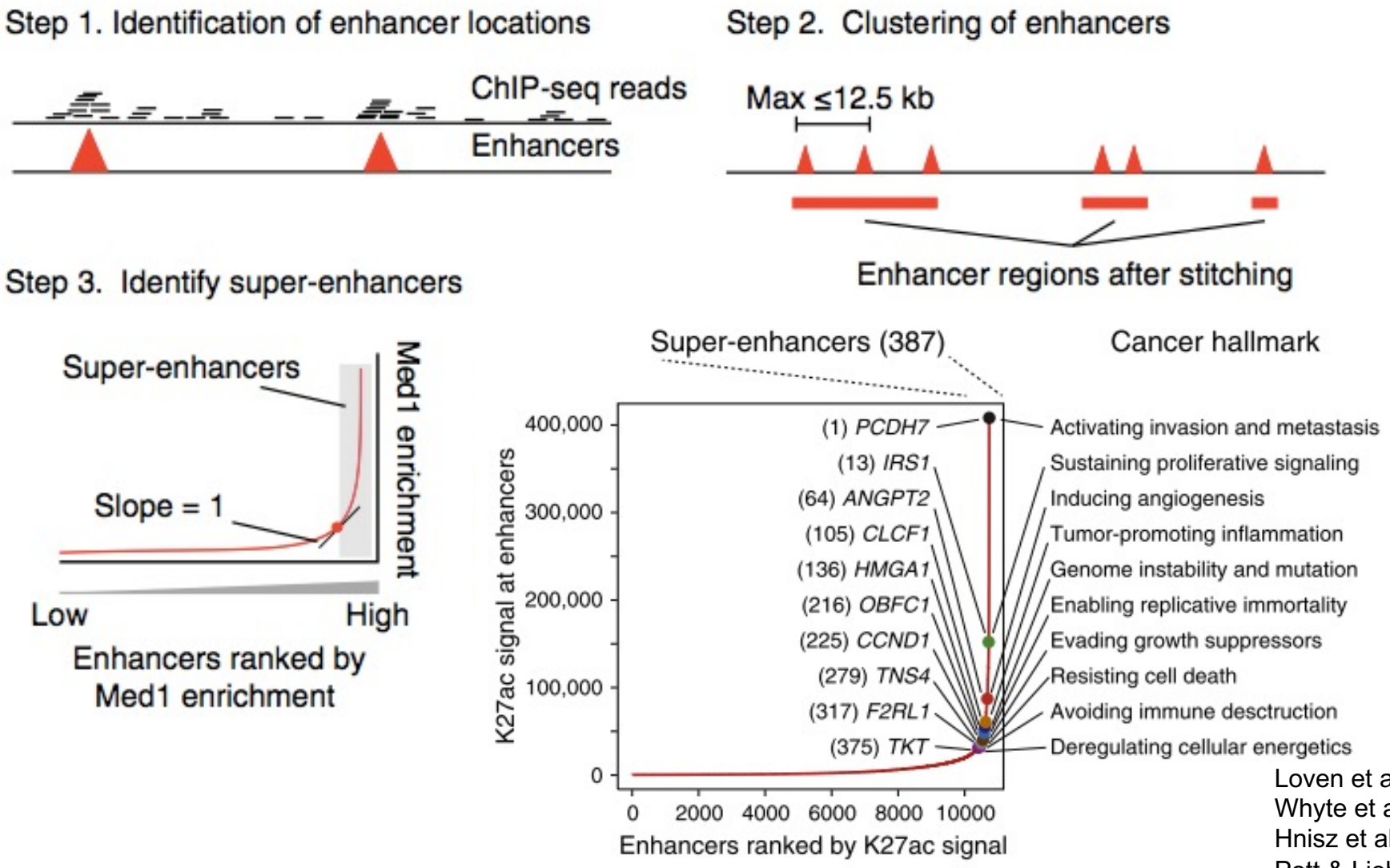


Typical enhancer



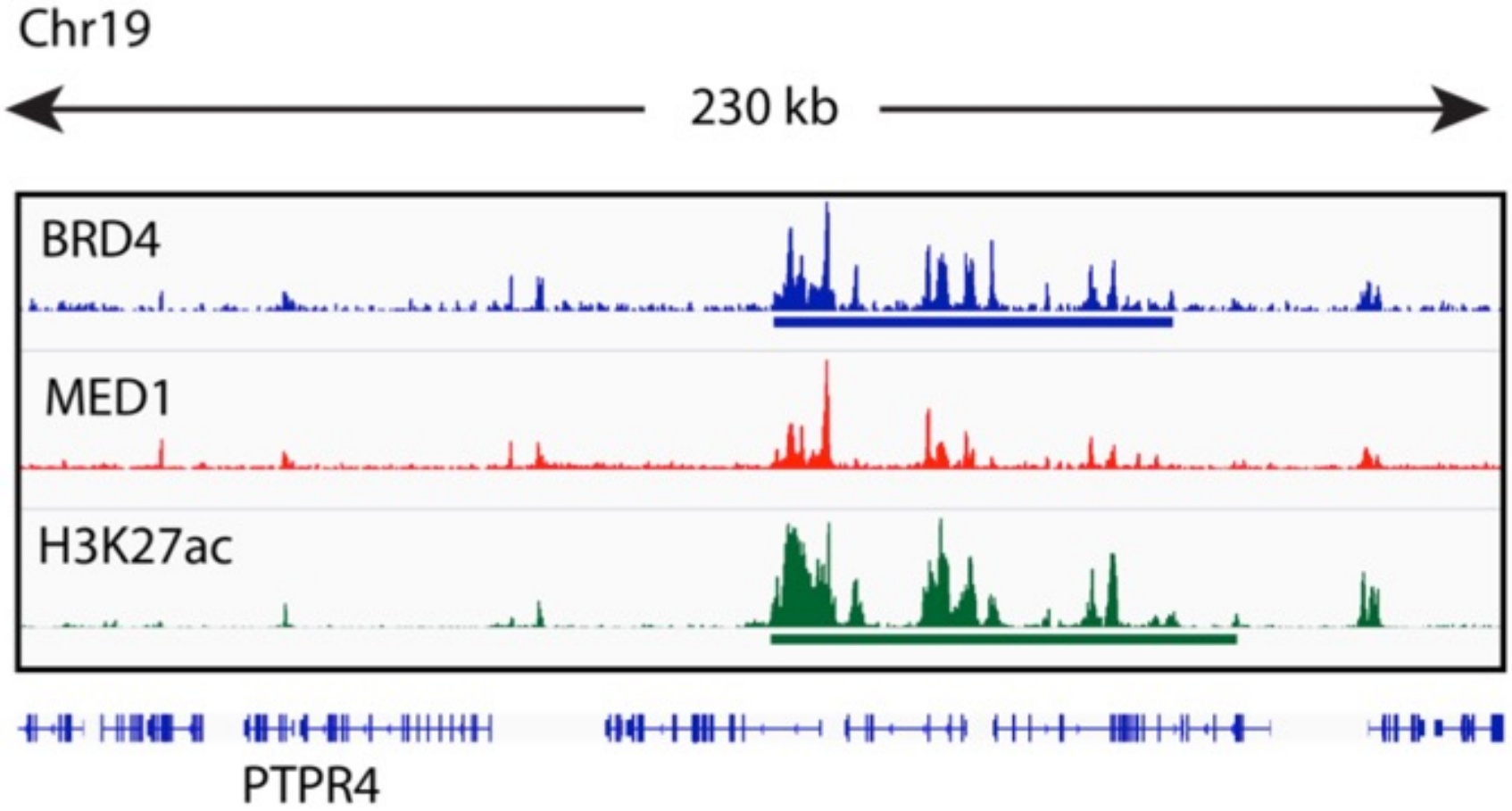
Super-enhancer

ROSE model for super-enhancer identification

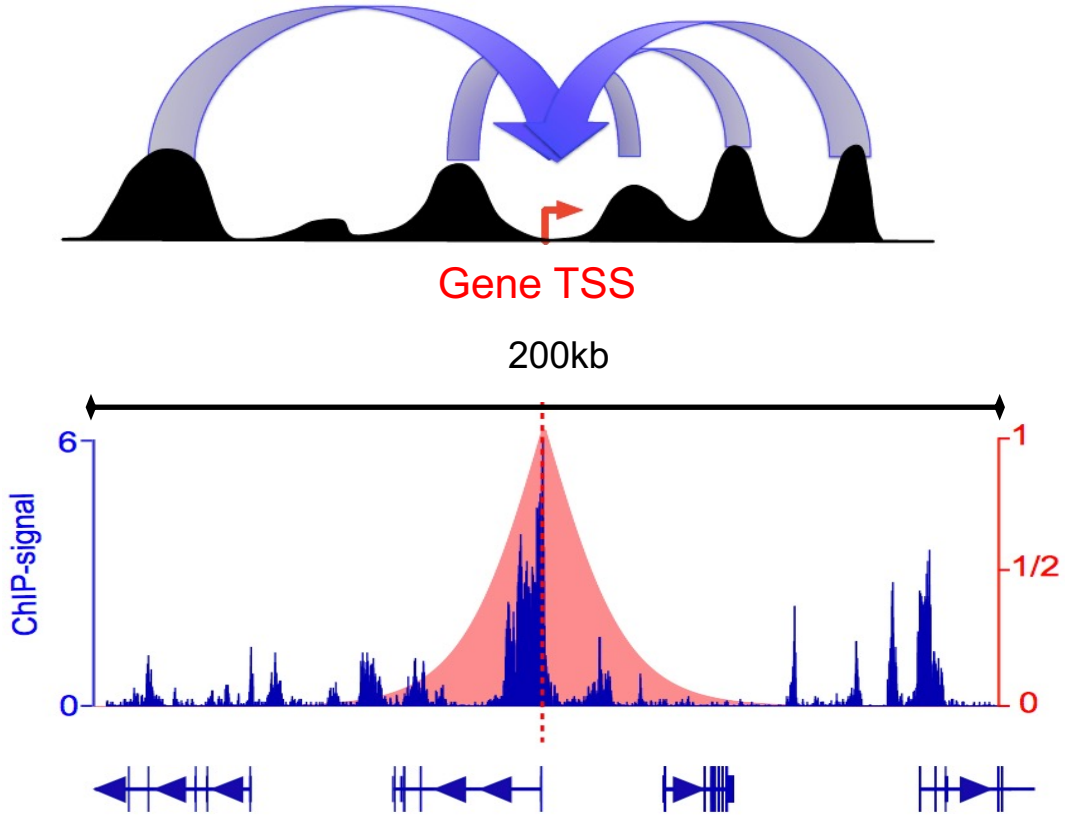


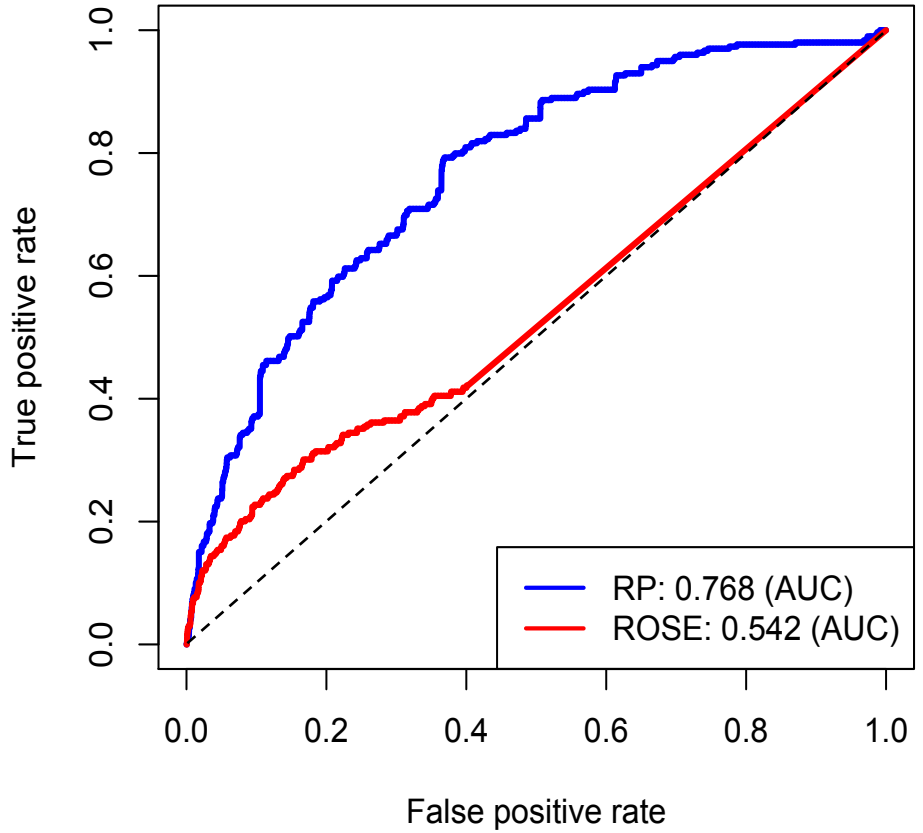
Loven et al, *Cell* 2013
Whyte et al, *Cell* 2013
Hnisz et al, *Cell* 2013
Pott & Lieb, *Nat Genet* 2015

Super-enhancer target genes can be far away



Computational approach to identify (super)enhancer target genes

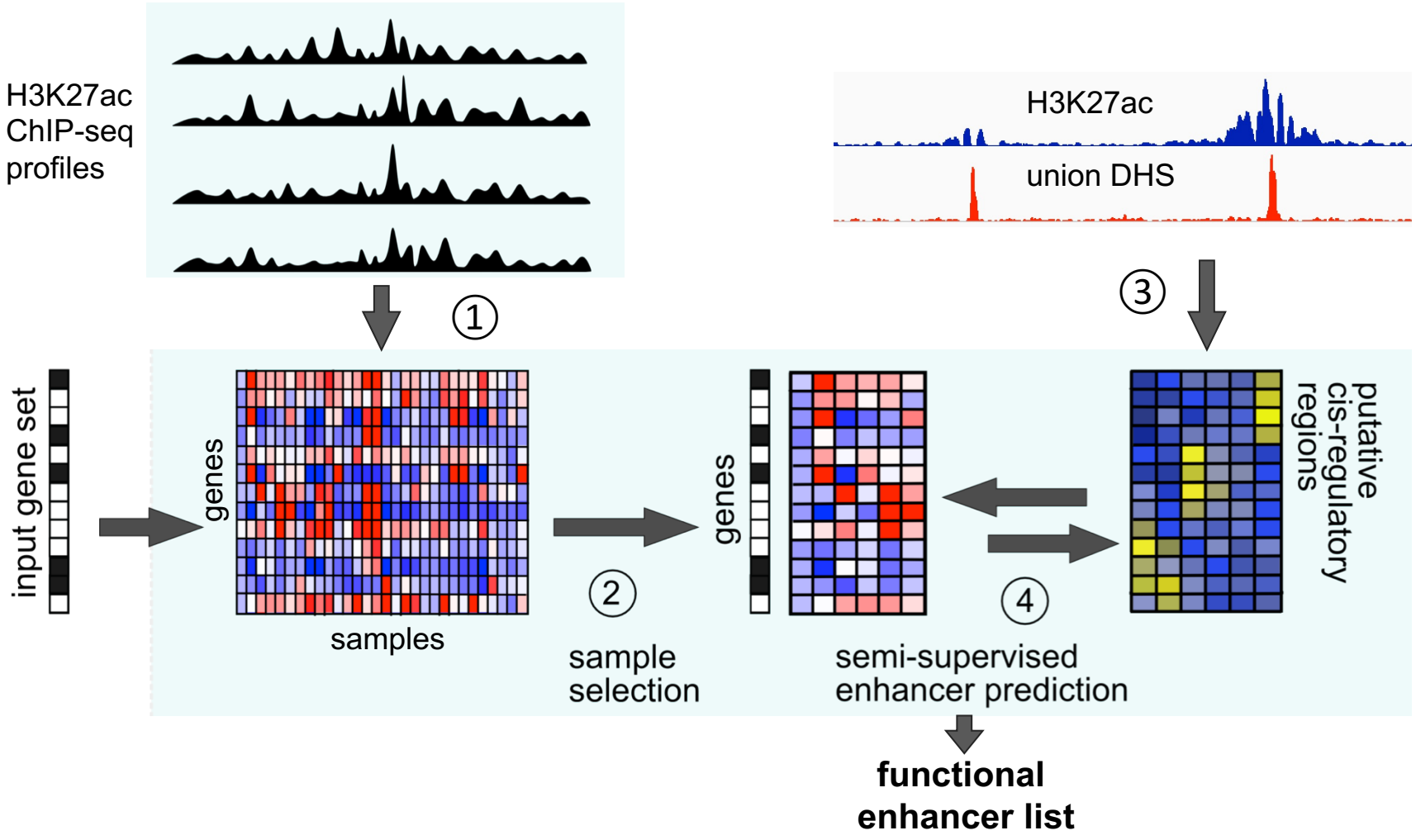




$$P_i = \sum_{|j|<10^5} W_j Z_j \qquad W_j = \frac{2\exp(-\alpha|j|)}{1+\exp(-\alpha|j|)} \qquad \alpha = \frac{\log 3}{10^4}$$

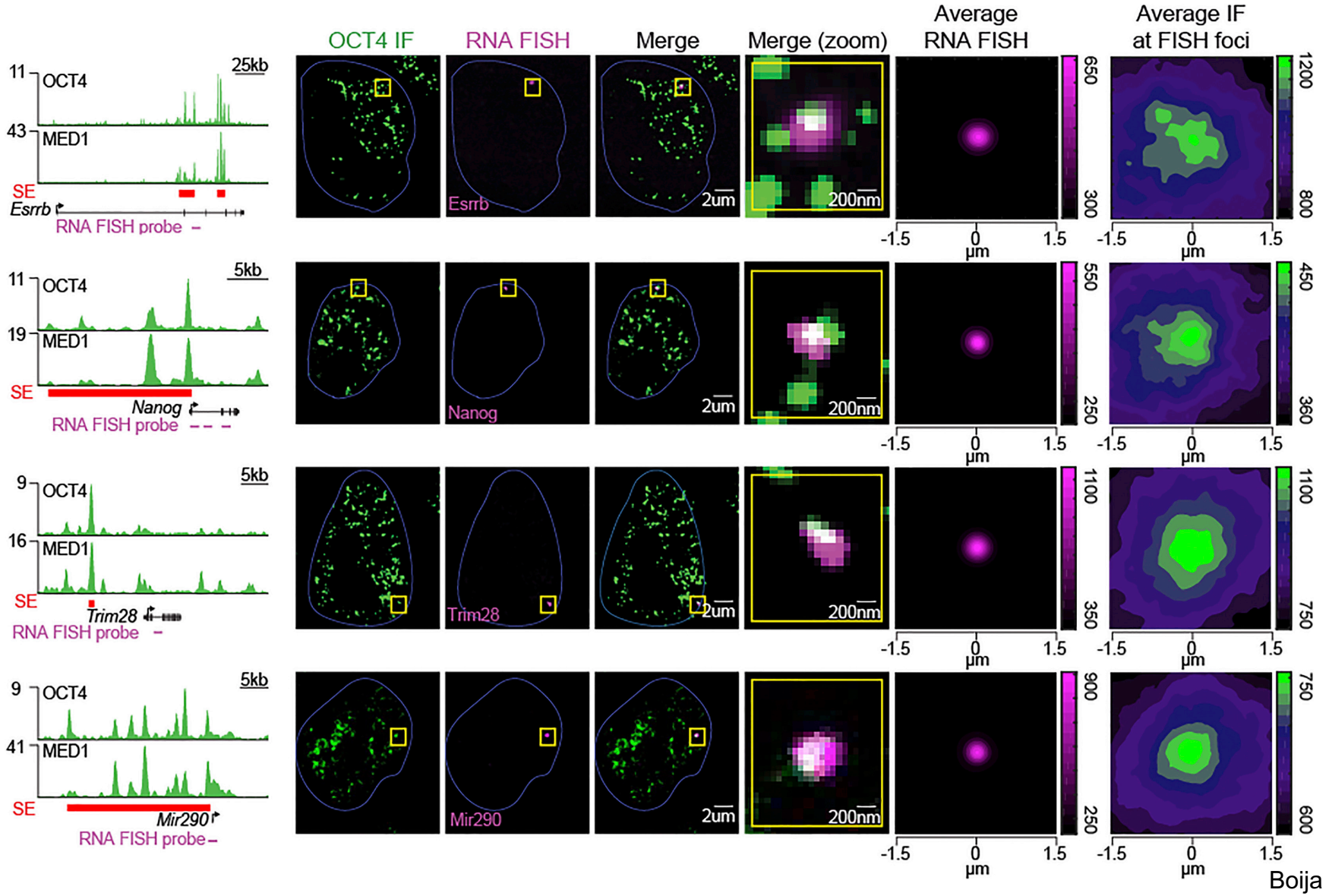
Wang, Zang et al. *Genome Res* 2016

MARGE: Model-based Analysis of Regulation of Gene Expression



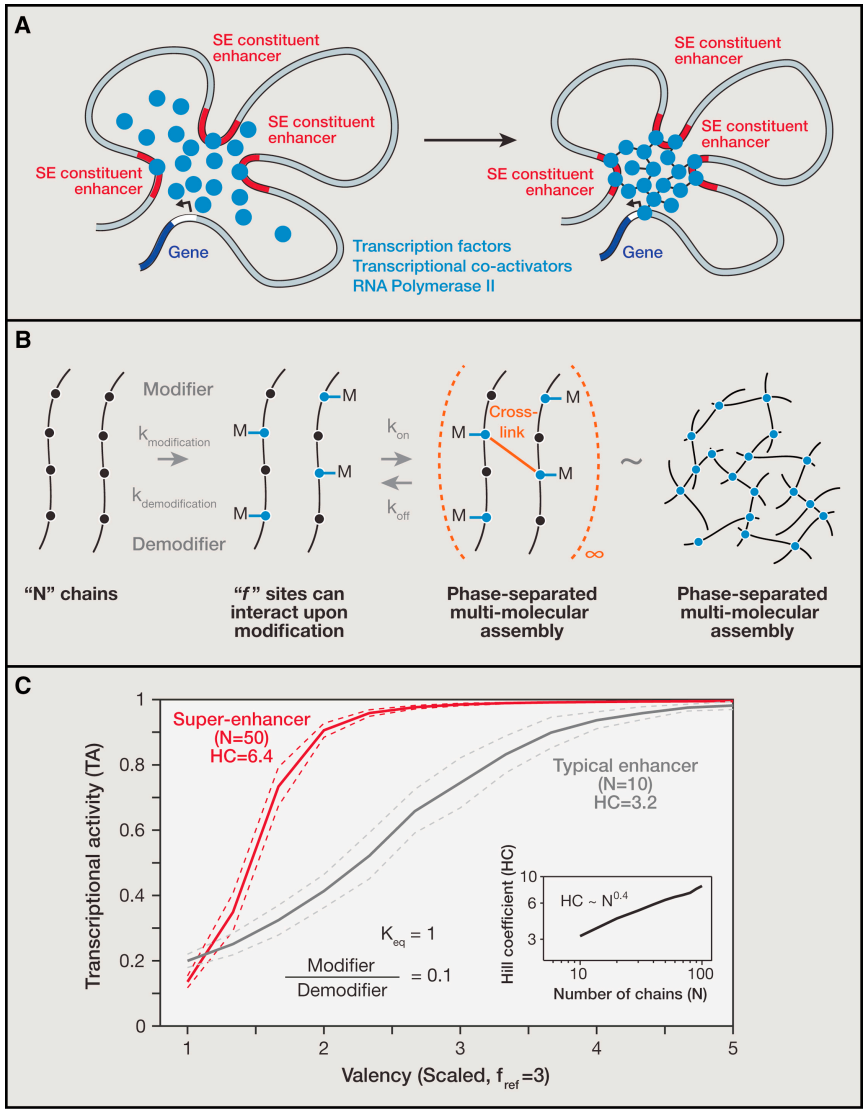
Wang, Zang et al. *Genome Res* 2016

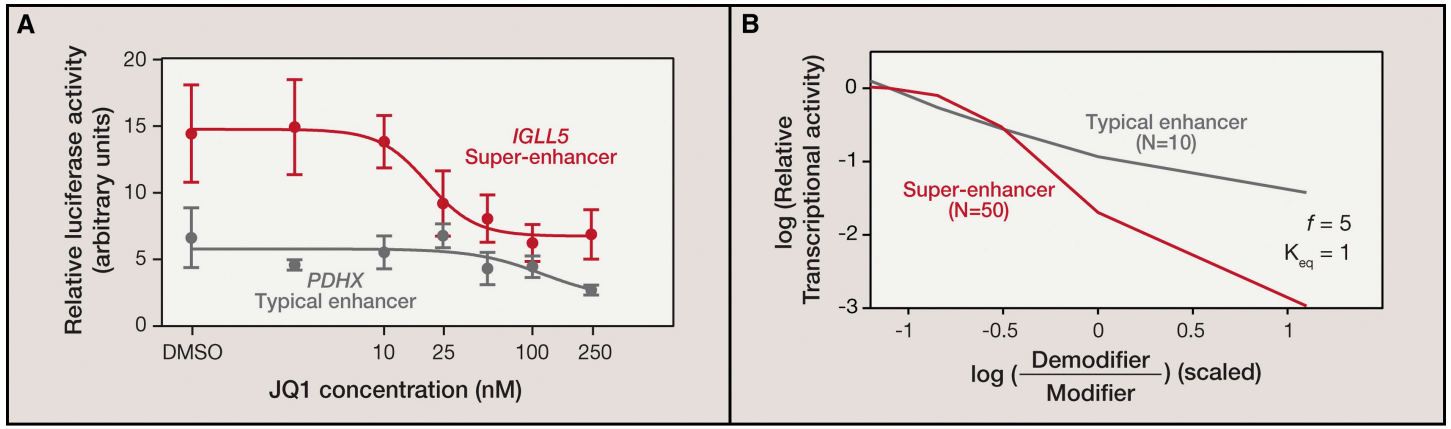
Super-enhancers and transcriptional condensates

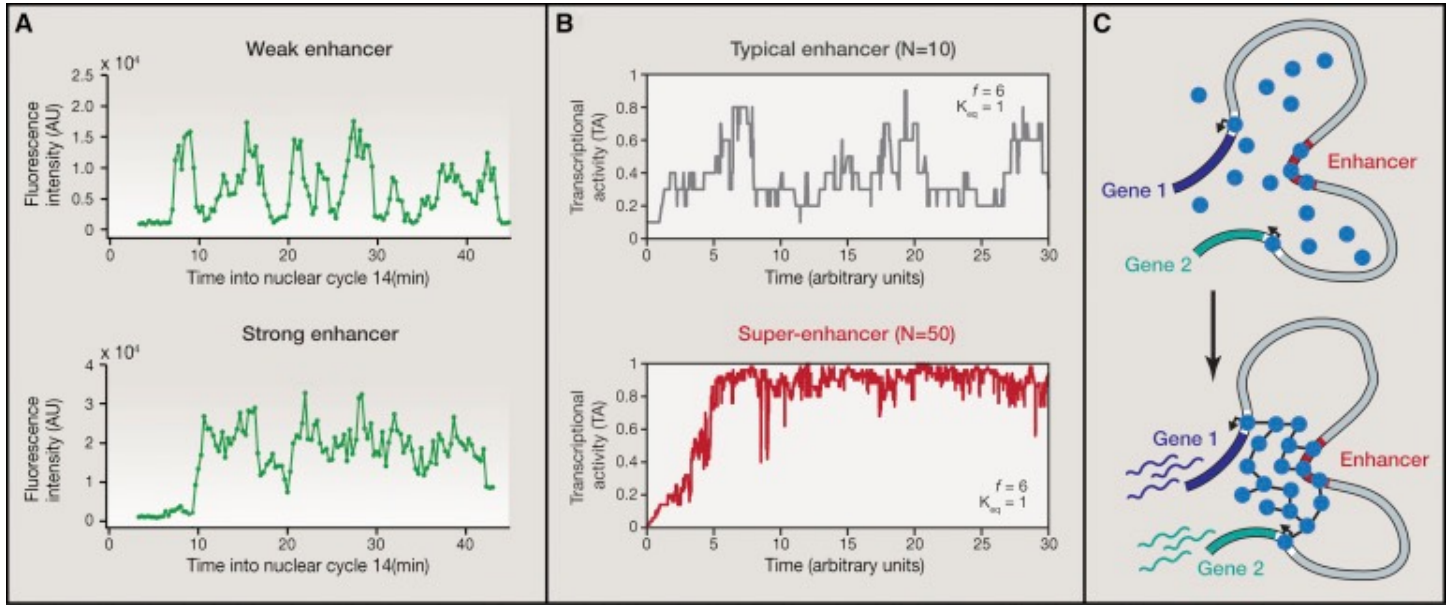


Boija et al. *Cell* 2018

A phase separation model for transcriptional control

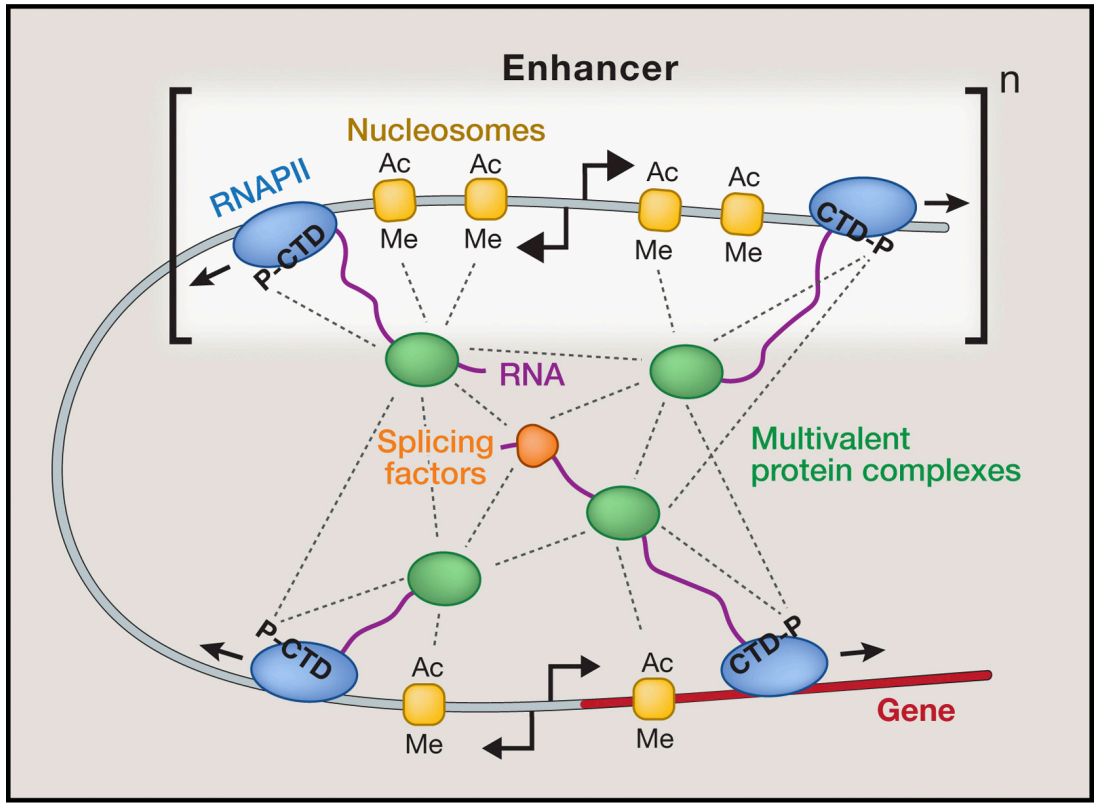


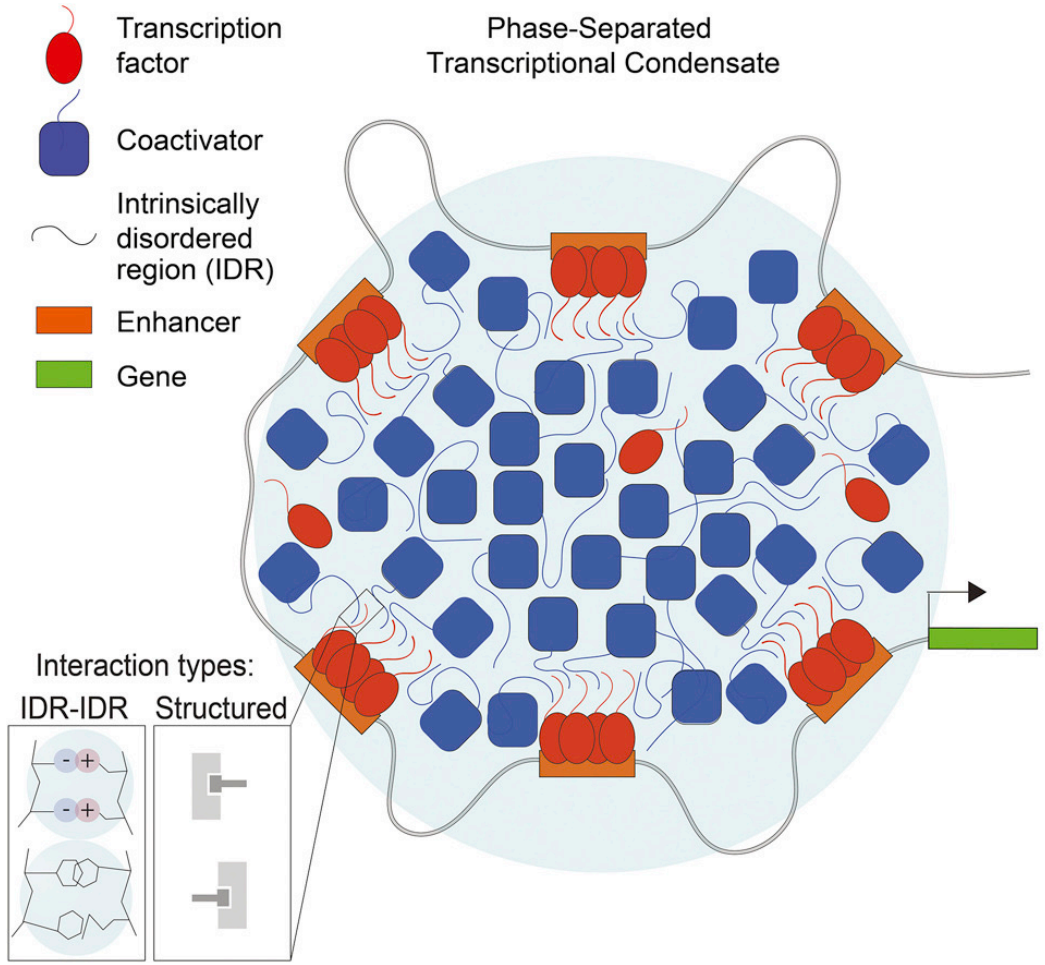




Hnisz et al. Cell 2017

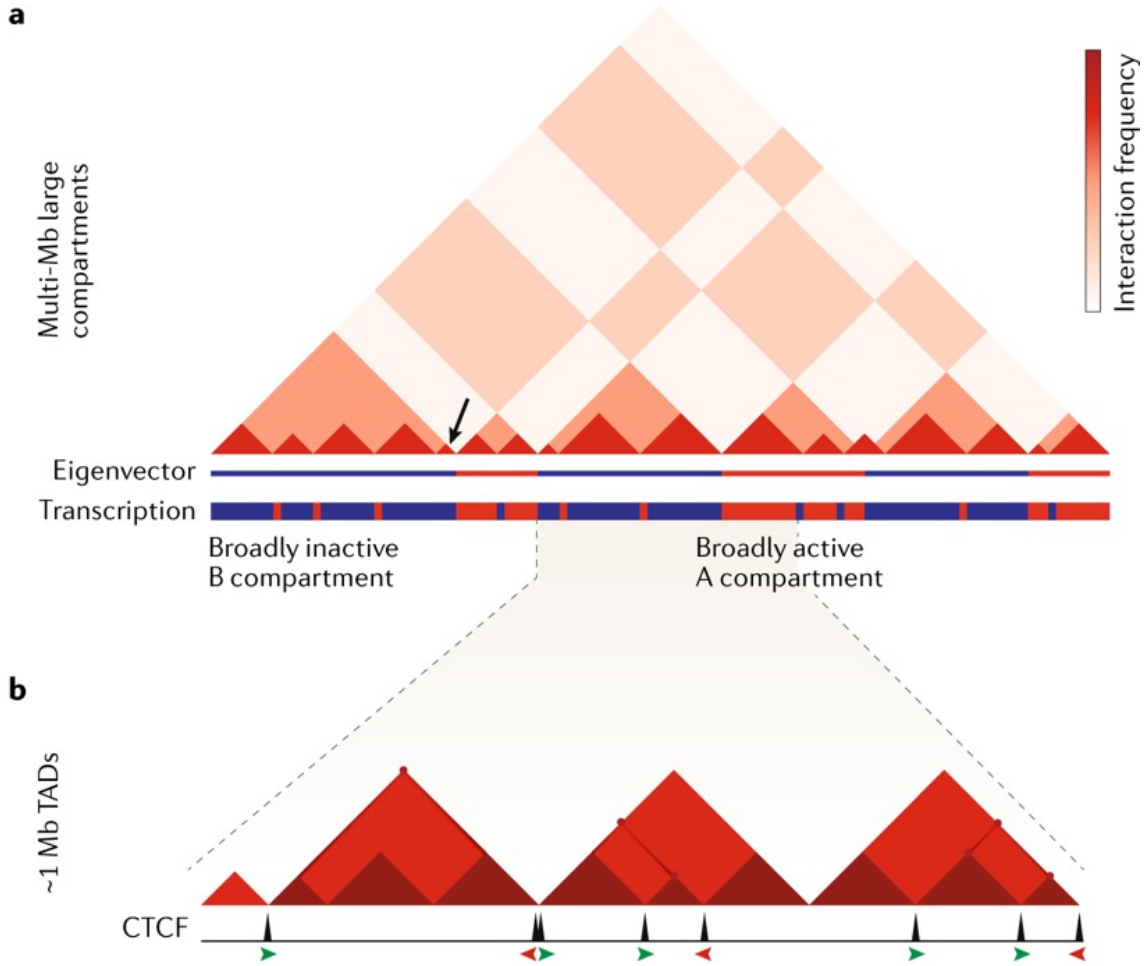
A phase separation model for transcriptional control

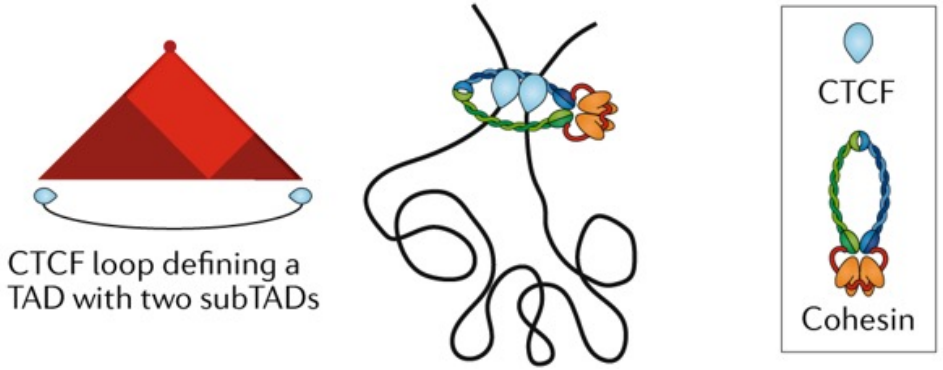




Hnisz et al. Cell 2017

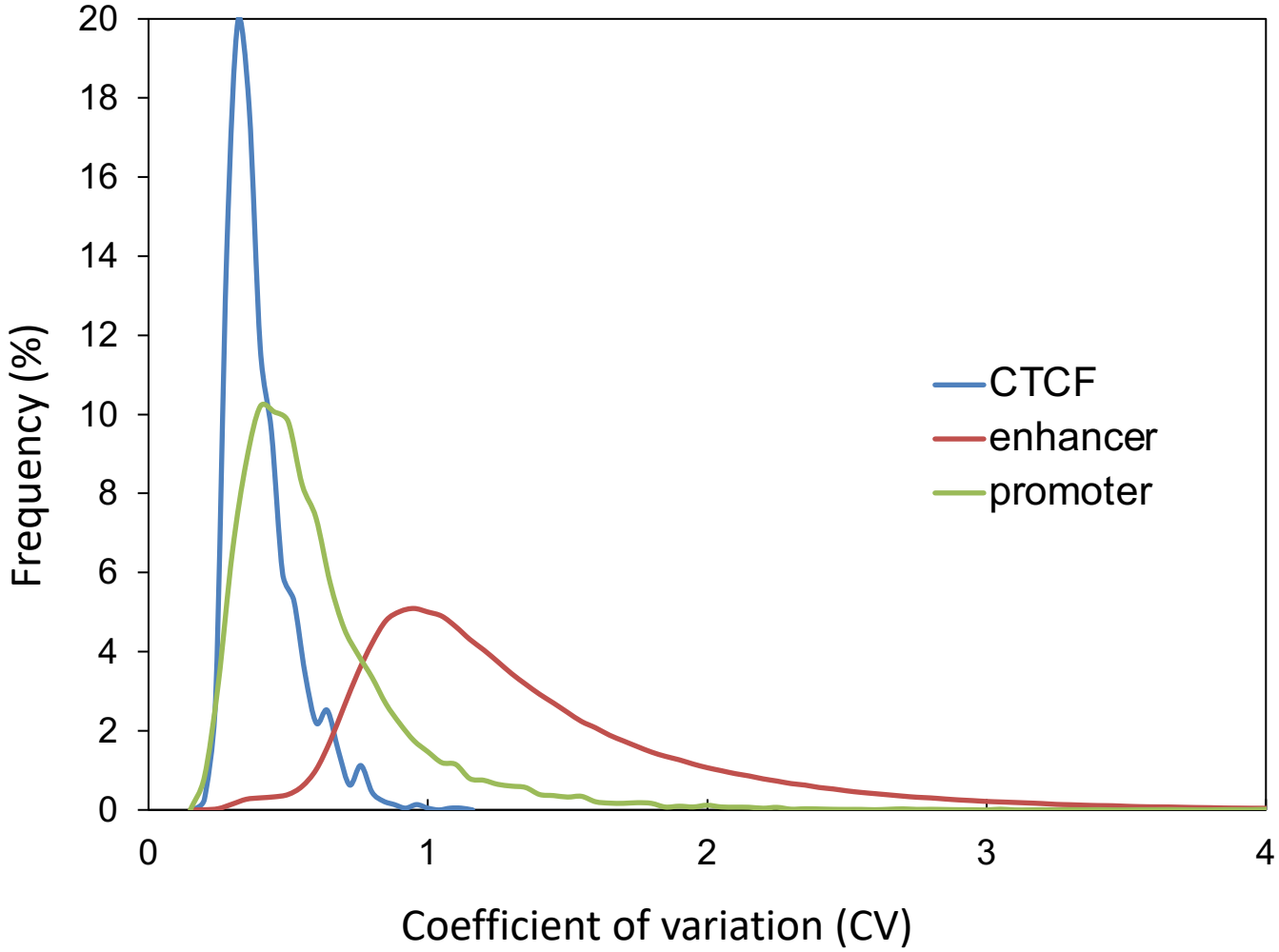
CTCF binds at chromatin domain boundaries and mediate 3D genome organization





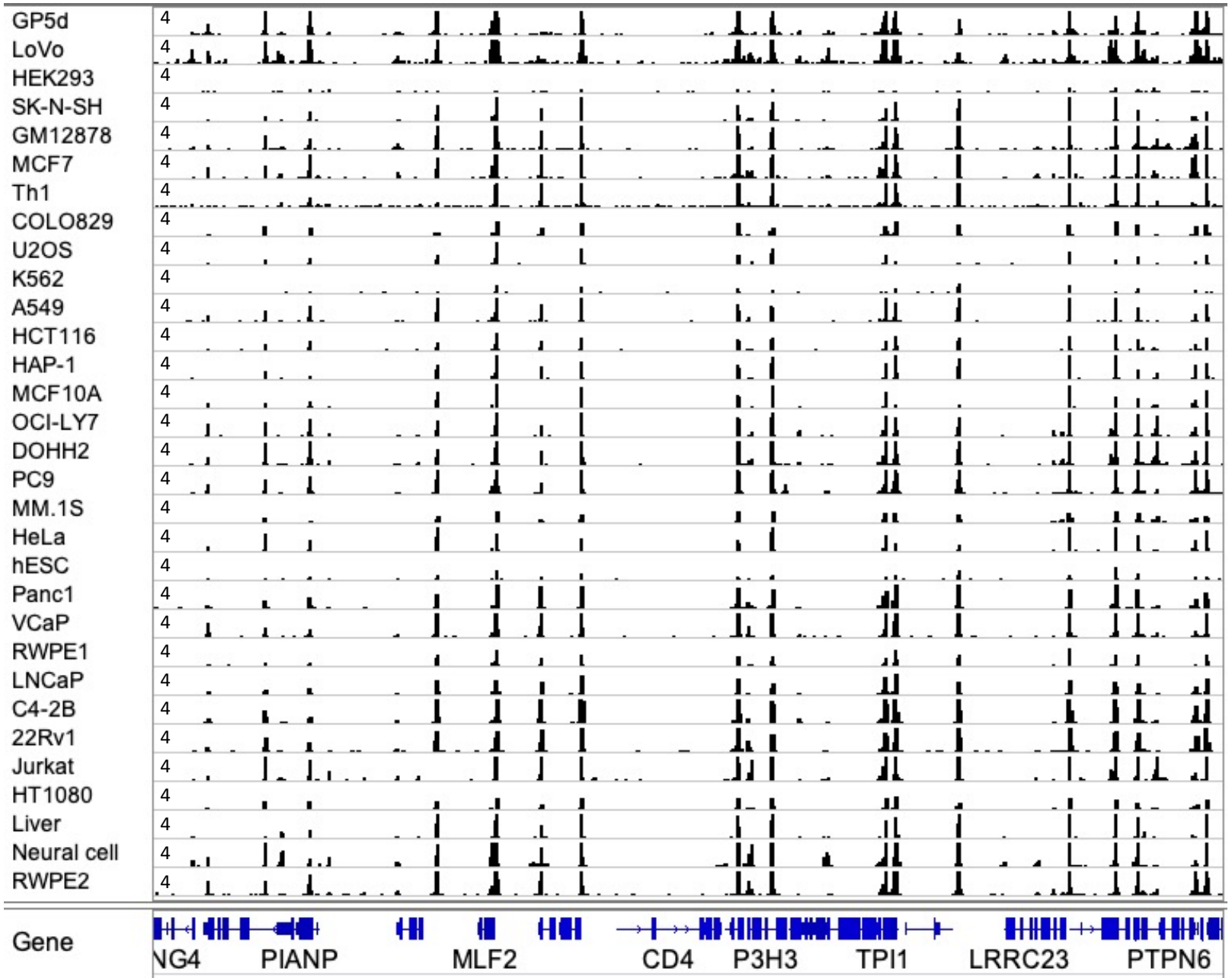
Rowley & Corces, *Nat Rev Genet* (2018)

Chromatin accessibility at CTCF binding sites are more stable than other regulatory regions in the human genome

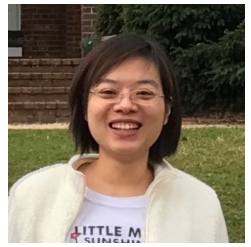


Fang, Wang *et al.*, *Genome Biology* 2020

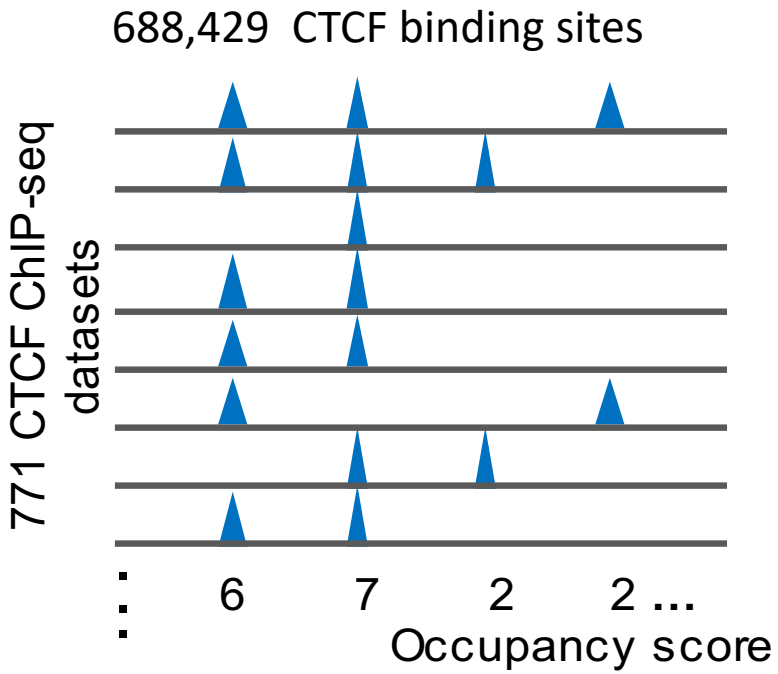
ChIP-seq reveals global landscape of CTCF binding in the genome

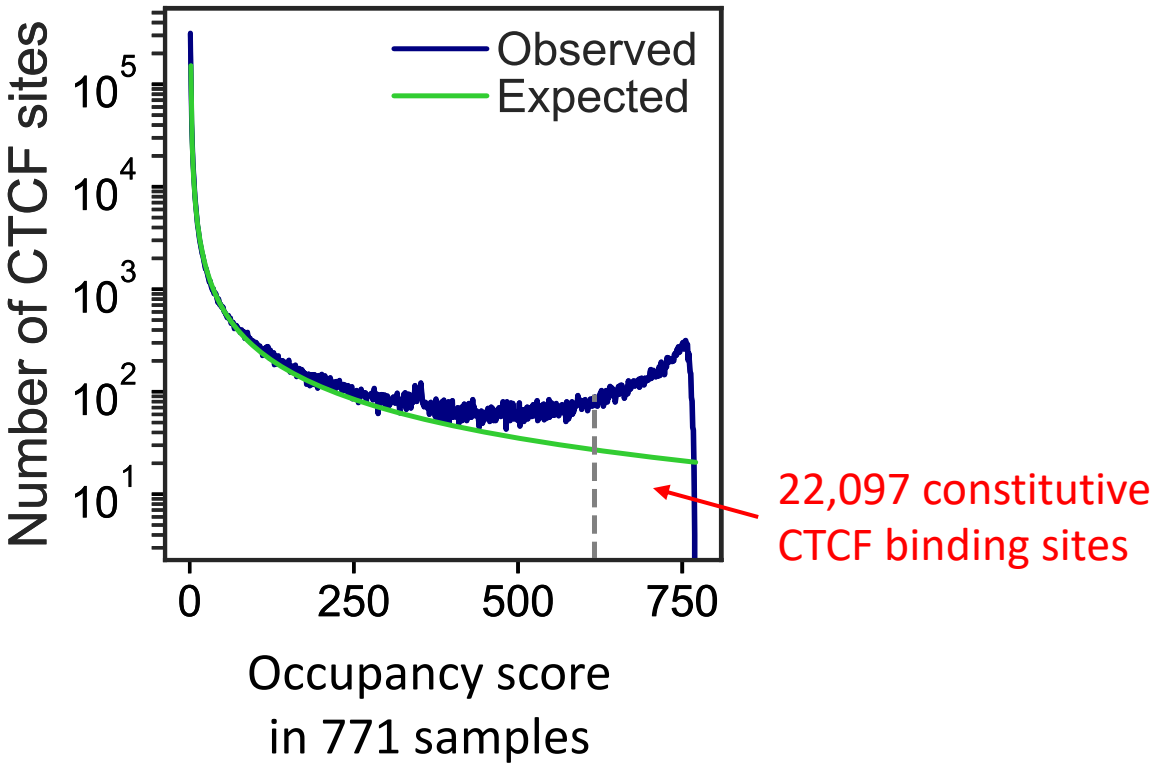


ChIP-seq reveal cross-cell-type CTCF binding patterns



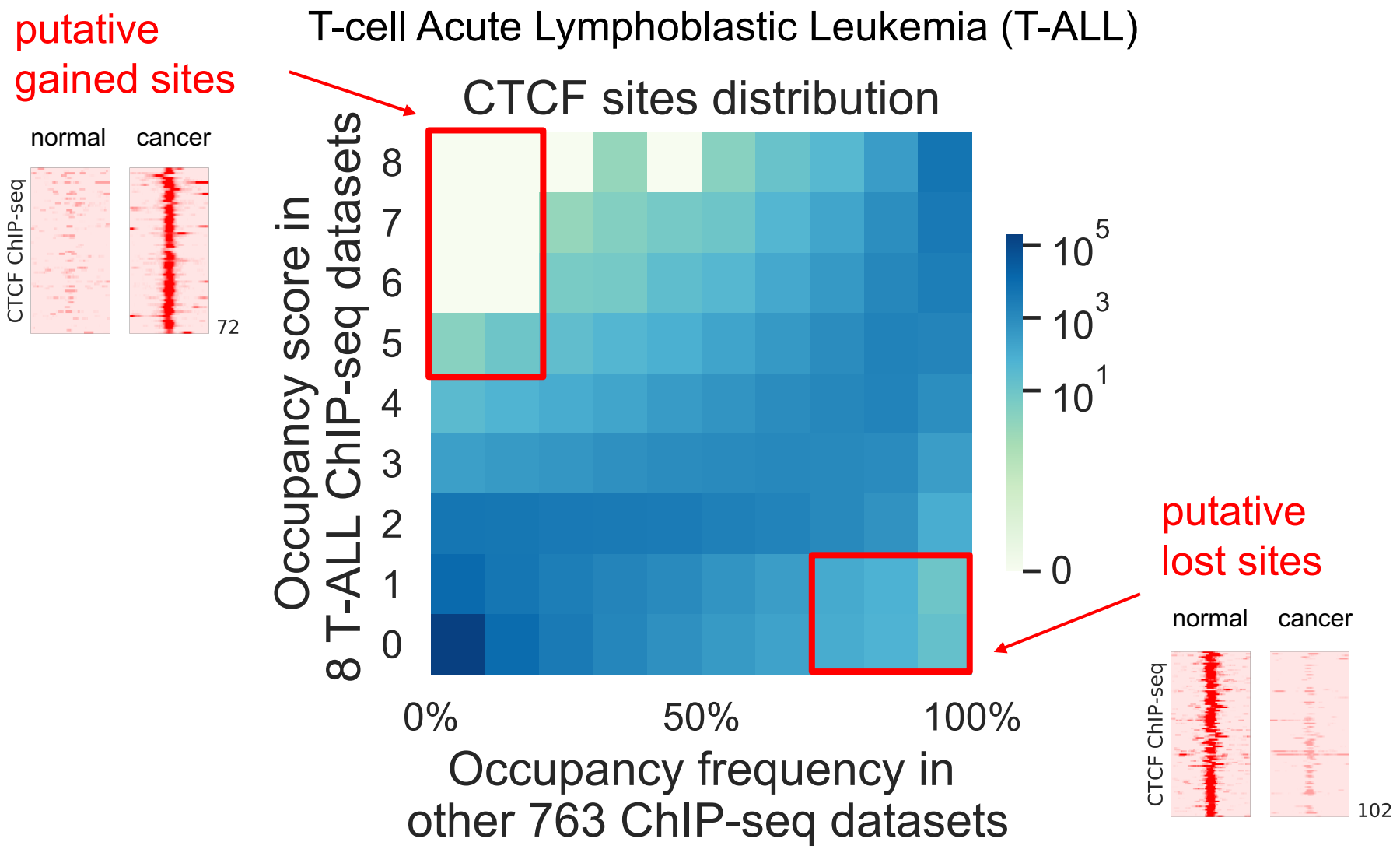
Zhenjia Wang





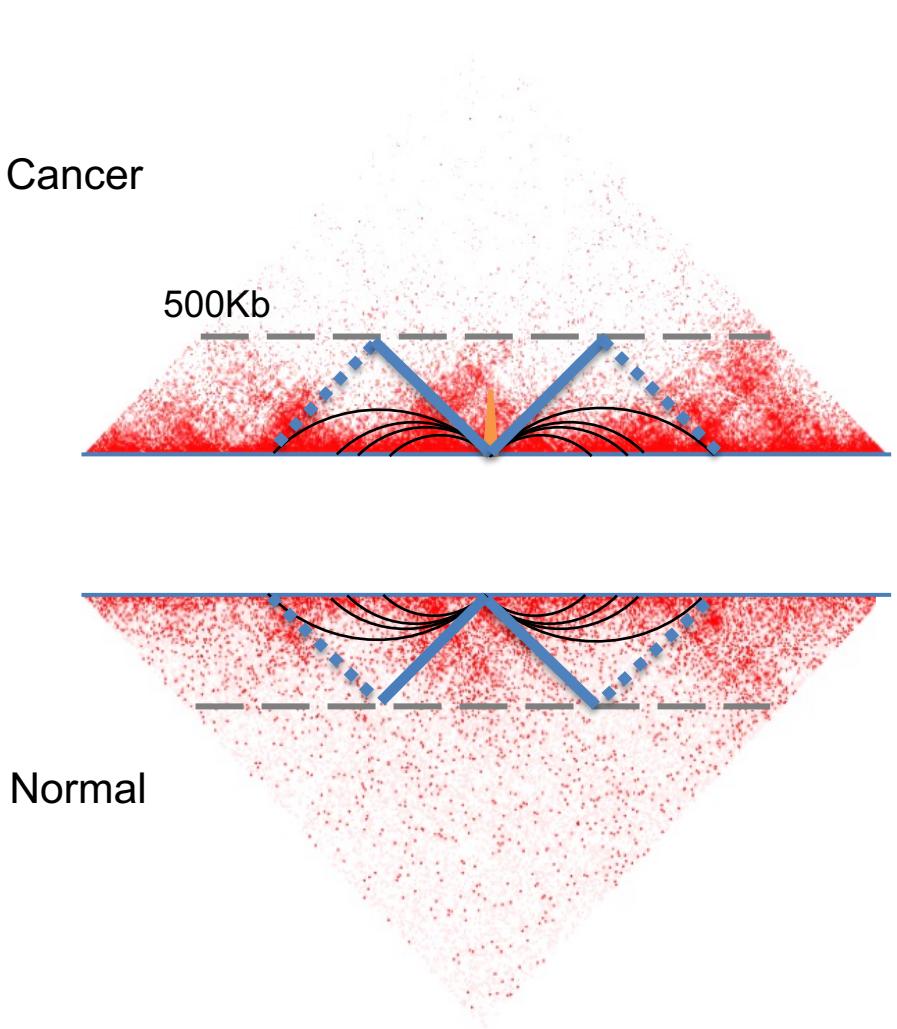
Fang, Wang *et al.*, *Genome Biology* 2020

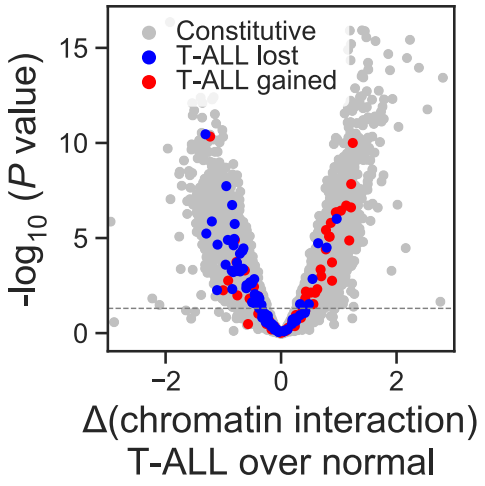
Identification of cancer-specific loss/gain of CTCF bindings

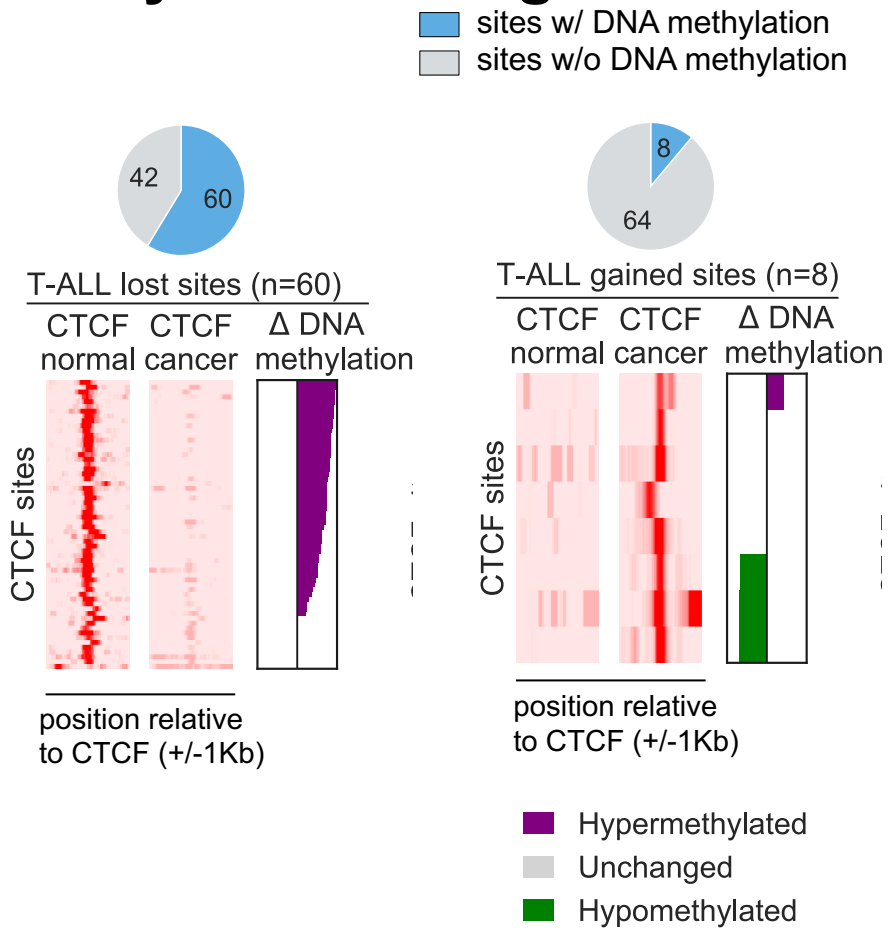


Fang, Wang *et al.*, *Genome Biology* 2020

Cancer-specific gained/lost CTCF binding partially associates with chromatin interaction changes and DNA methylation changes



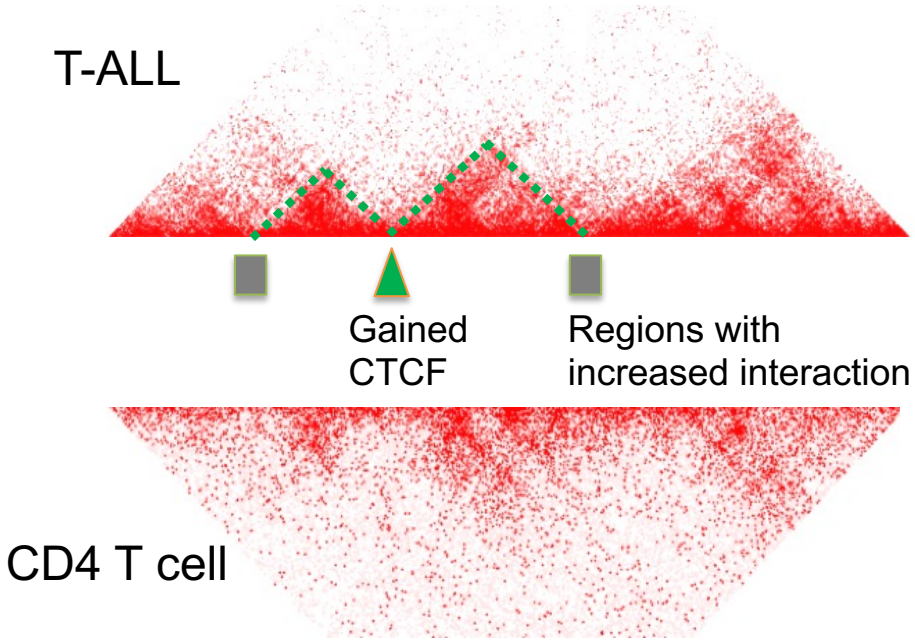




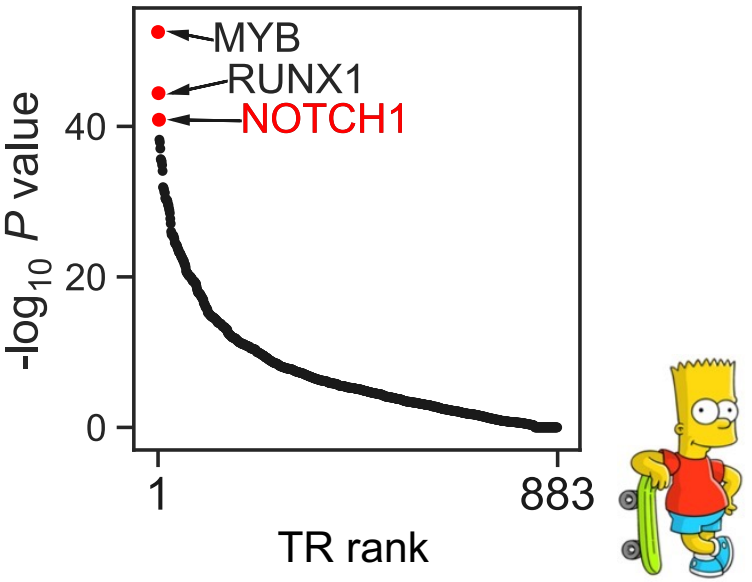
Wang *et al.*, *Bioinformatics* 2021

Fang, Wang *et al.*, *Genome Biology* 2020

Gained CTCF associating regions with increased chromatin interaction show oncogenic TF binding in T-ALL

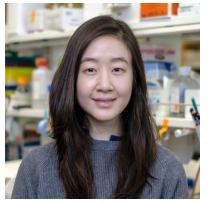


BART analysis

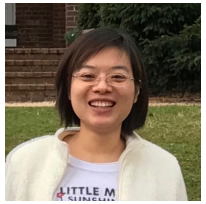


Fang, Wang *et al.*, *Genome Biology* 2020

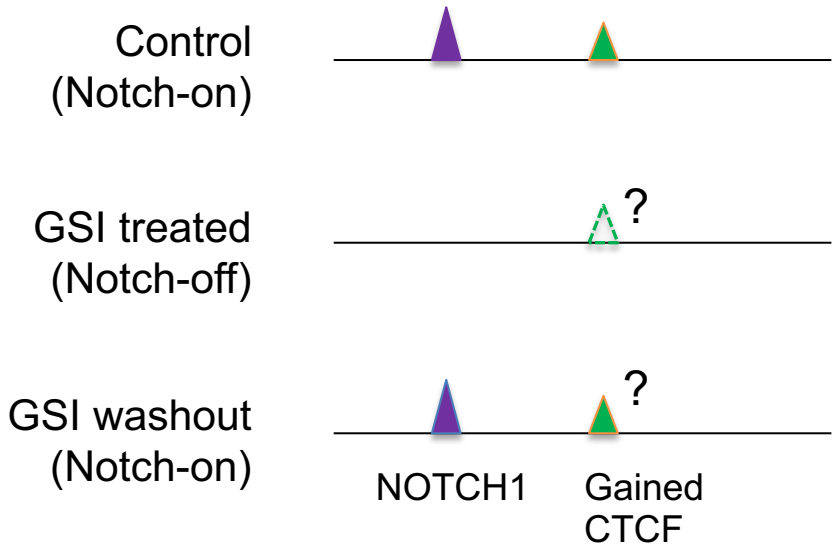
Perturbation of Notch results in alterations of T-ALL gained CTCF binding

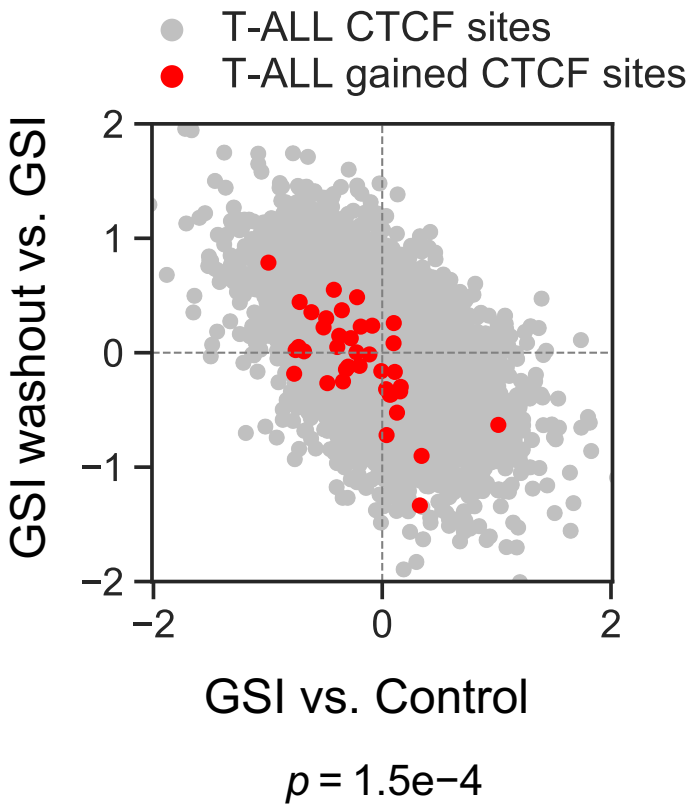


Celestia Fang
Ntziachristos Lab



Zhenjia Wang

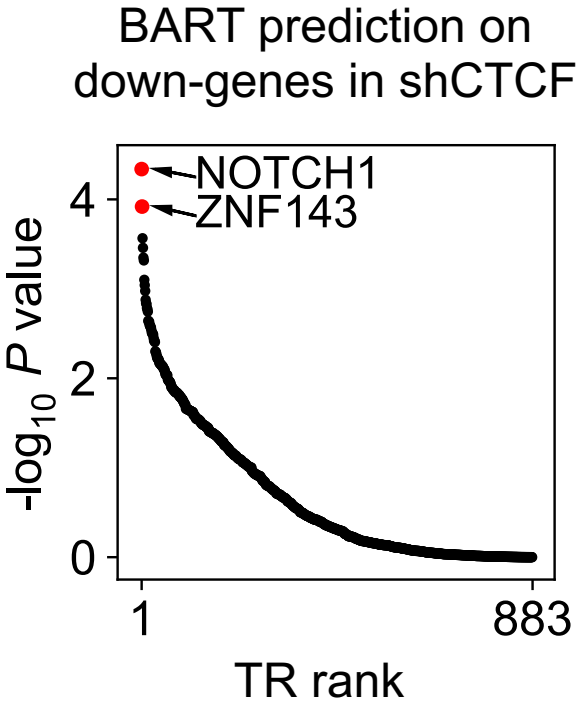


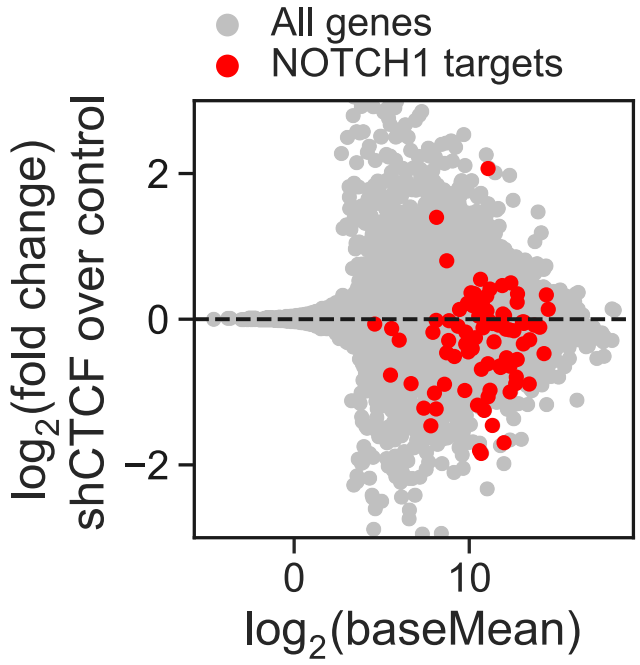


$p = 1.5e{-4}$

Fang, Wang *et al.*, *Genome Biology* 2020

T-ALL gained CTCF binding is required for oncogenic NOTCH1 to activate target genes



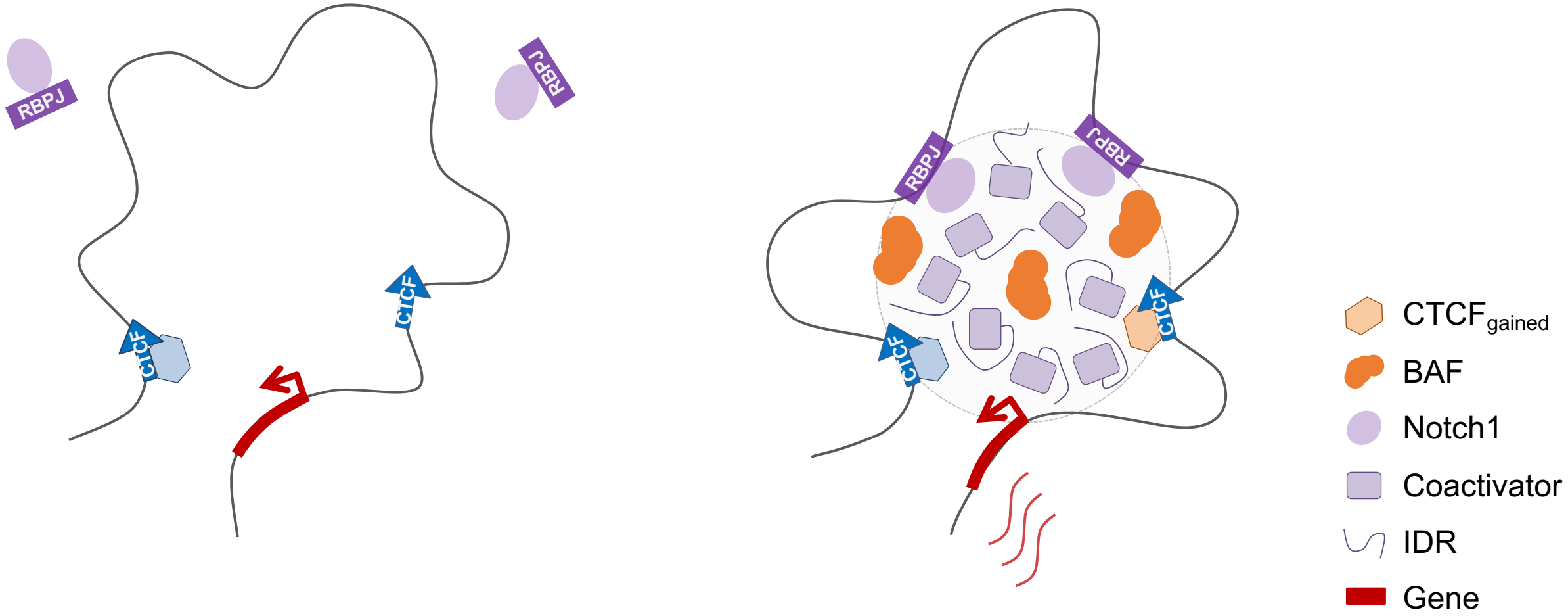


Fang, Wang *et al.*, *Genome Biology* 2020

CTCF facilitates formation of transcriptional condensates at super-enhancers for transcriptional activation

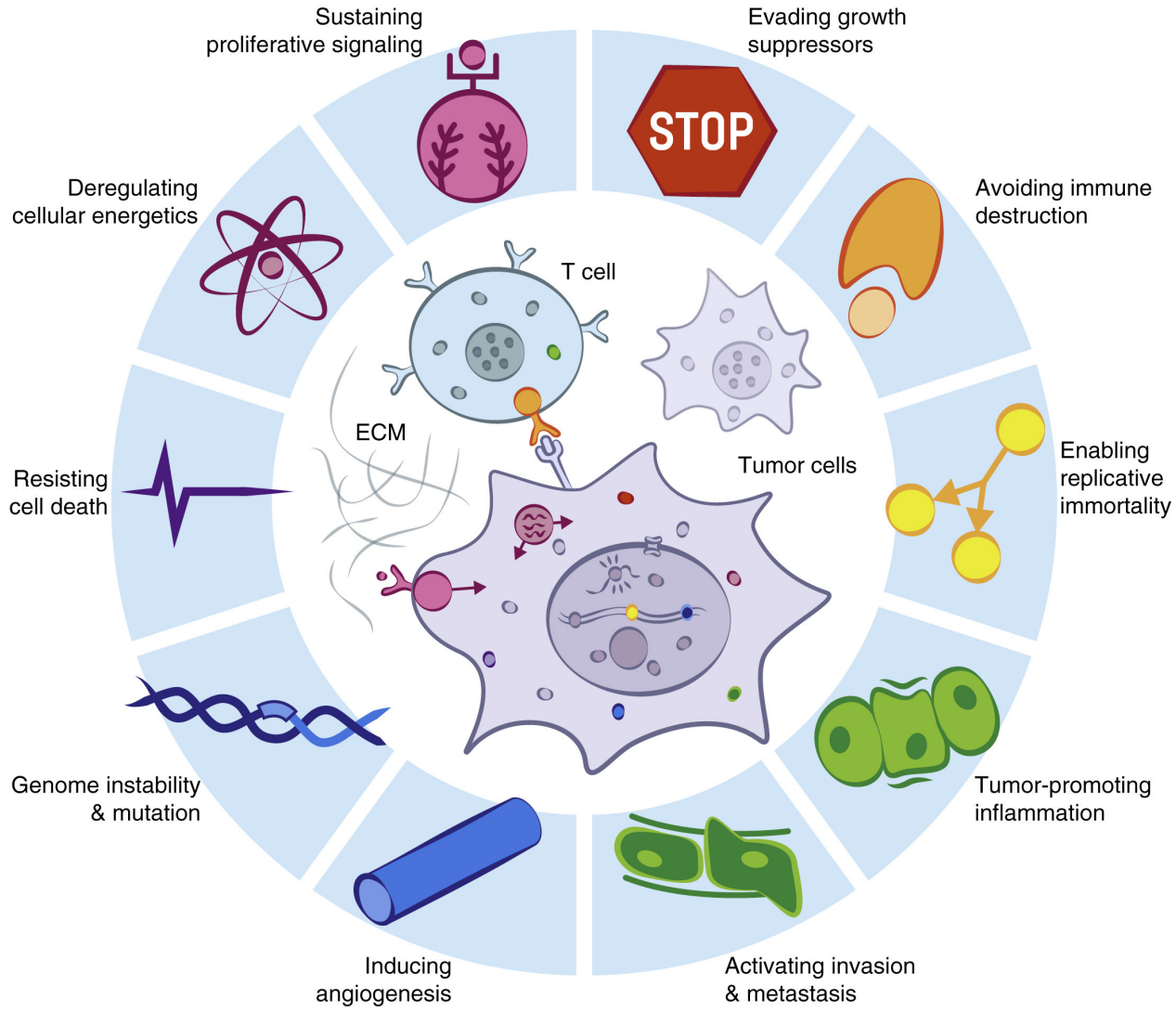
Without gained CTCF

With gained CTCF



Fang, Wang *et al.*, *Genome Biology* 2020

Biomolecular condensates occur in cancer hallmarks



Boija *et al.*, Cancer Cell 2021