L4.7 – Transcriptional regulation (Research Papers and Literature Review)

AGENDA

- Research Paper
  (including discussion about promoter/enhancer similarities)

- Literature Review on Transcriptional Regulation during Adipogenesis

Architectural and Functional Commonalities between Enhancers and Promoters

Tae-Kyung Kim[1],* and **Ramin Shiekhattar[2],***
[1]Department of Neuroscience, University of Texas Southwestern Medical Center, Dallas, TX 75390-9111, USA
[2]University of Miami Miller School of Medicine, Sylvester Comprehensive Cancer Center, Department of Human Genetics, Biomedical Research Building, Room 719, 1501 NW 10th Avenue, Miami, FL 33136, USA
*Correspondence: taekyung.kim@utsouthwestern.edu (T.-K.K.), rshiekhattar@med.miami.edu (R.S.)
http://dx.doi.org/10.1016/j.cell.2015.08.008

With the explosion of genome-wide studies of regulated transcription, it has become clear that traditional definitions of enhancers and promoters need to be revisited. These control elements can now be characterized in terms of their local and regional architecture, their regulatory components, including histone modifications and associated binding factors, and their functional contribution to transcription. This Review discusses unifying themes between promoters and enhancers in transcriptional regulatory mechanisms.

Enhancers share many characteristics with the promoters they regulate
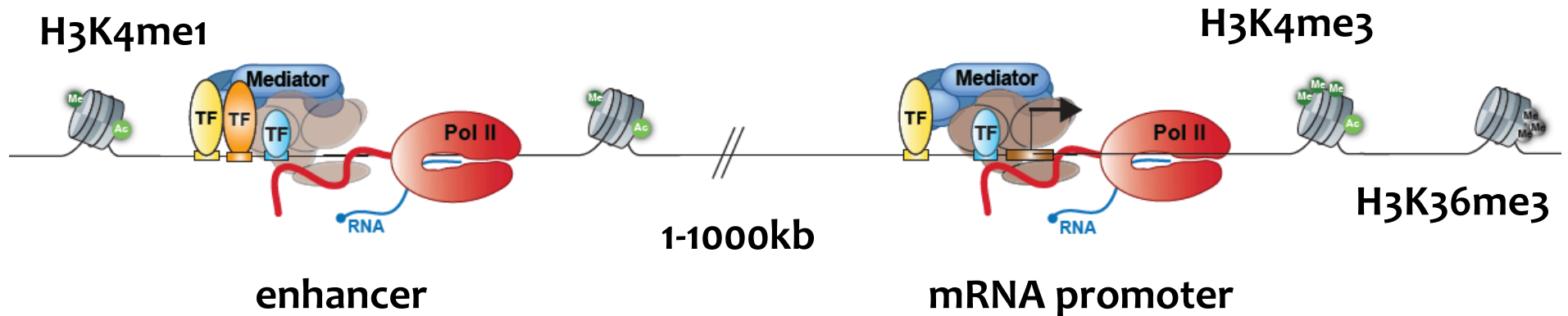
1. DNA accessibility (both marked by HS)
2. Transcription factor and Mediator binding
3. Architecture of Core promoter elements)
4. Transcription by Pol II (including presence of Ser-5P)

Similar to a promoter, an enhancer can direct RNA transcription from a defined site by independent RNAPII transcription machinery assembled with general TFs.

Apparent Differences
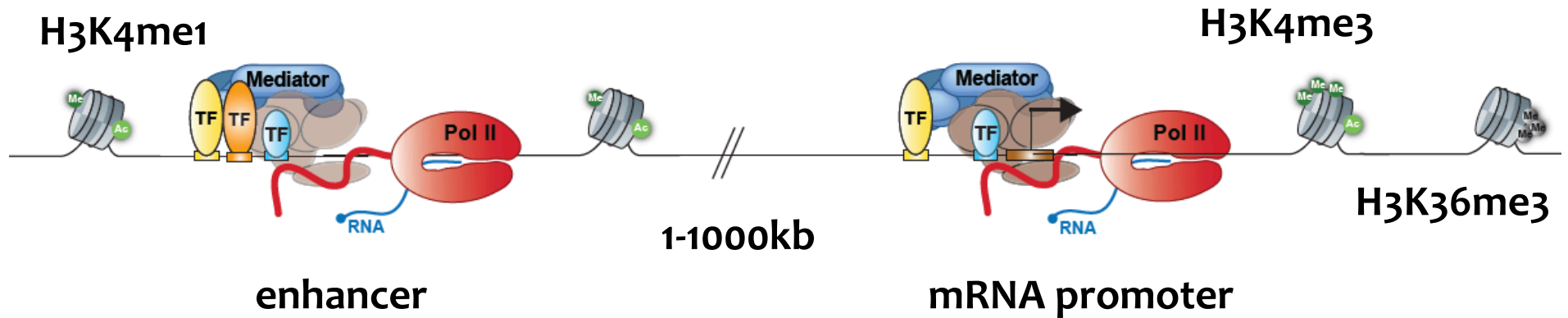
1. Lack of elongating POLII (Ser-2P) and H3K36me3

2. H3K4me1 versus H3K4me3

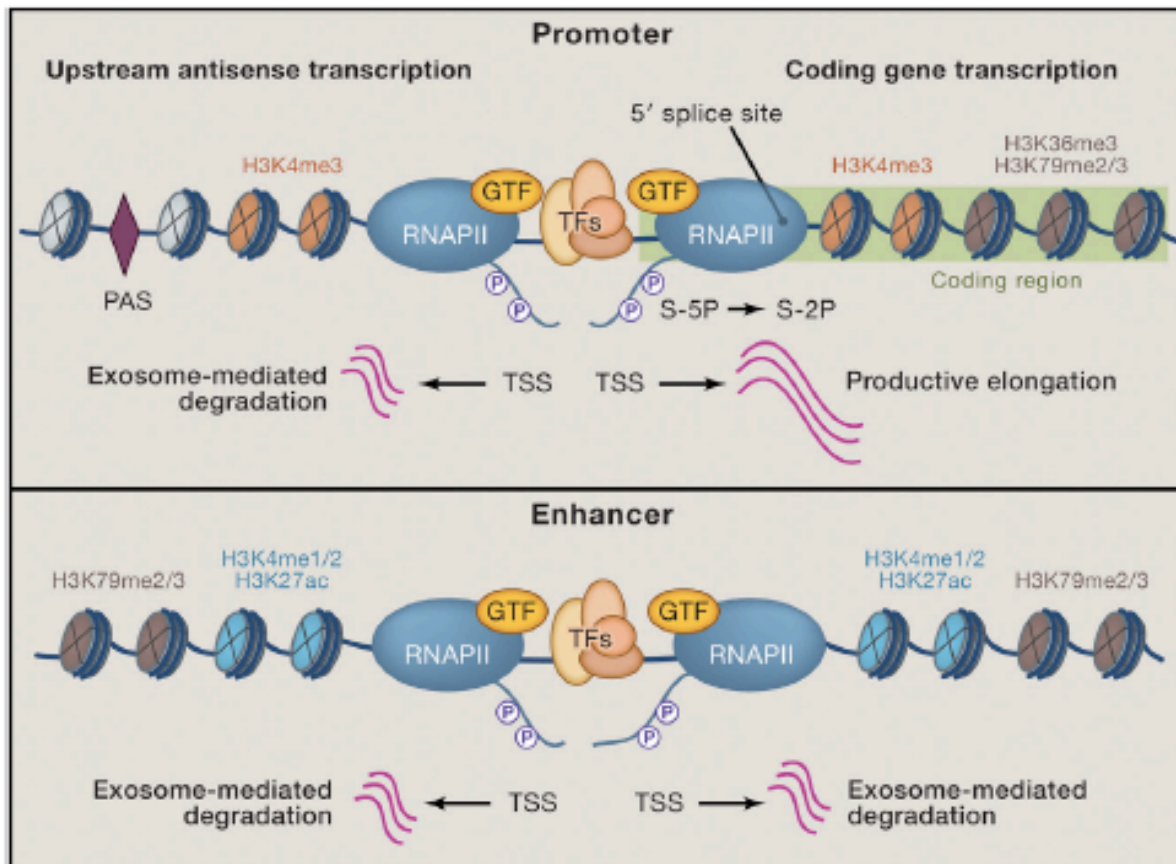3. Bidirectional transcription

Figure 1. A Contemporary View on Promoters and Enhancers

Features of promoters include transcription initiation in the sense and anti-sense direction being mediated by the transcription machinery assembled independently onto its own core promoter. Although not shown here, convergent transcription has been observed at the promoters of weakly expressed genes. H3K4me3 is highly enriched at the promoter regions. Enhancer-like chromatin signatures (H3K4me1 and H3K27ac) and the Tyr-1P form of RNAPII have also been observed near the upstream anti-sense TSSs. Polyadenylation sites are enriched near the 3' end of the upstream anti-sense RNAs and mediate the exosome-dependent degradation of the antisense RNAs. 5' splice sites are only present in the coding gene and might contribute to the productive elongation of sense mRNA transcripts through the binding of the U1 splicing complex, which blocks PAS-mediated early termination. The Ser-5P form of RNAPII is engaged in upstream anti-sense transcription, but it is not known whether Ser-2P of RNAPII occurs during the elongation of anti-sense RNA.

Features of enhancers include, as with the promoter, recruitment of the general transcription factors (GTF), including RNAPII, and initiation of transcription at defined sites. Enhancer-driven transcription typically exhibits more prominent bi-directionality than that stemming from the promoter. H3K4me1/2 is commonly enriched at enhancers. Functionally active enhancers also exhibit a high level of H3K27 acetylation, whereas poised or inactive enhancers are marked by H3K27me3. Ser-5P and Tyr-1P forms of the RNAPII have been observed. It is not clear whether or not Ser-2P RNAPII and H3K36me3 marks are present at active enhancers. 5' splice site sequences are not enriched near the regions surrounding enhancers. Both strands of enhancer RNAs appear to be degraded by the exosome, although it is not known whether it is mediated by the PAS-dependent mechanism.

Open questions

- Are promoter and enhancers functionally interchangeable?

- How is transcription of eRNAs regulated?

- Are there specific features that are required for eRNA function?

Widespread transcriptional pausing and elongation control at enhancers

Telmo Henriques,[1,3] Benjamin S. Scruggs,[1] Michiko O. Inouye,[1,3] Ginger W. Muse,[1] Lucy H. Williams,[1] Adam B. Burkholder,[2] Christopher A. Lavender,[2] David C. Fargo,[2] and Karen Adelman[1,3]

[1]Epigenetics and Stem Cell Biology Laboratory, [2]Center for Integrative Bioinformatics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, USA; [3]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA

- Is enhancer RNA transcription a good marker for enhancer identification? (Basic and clinical relevance)

- How is transcription of enhancer RNA regulated?

Start-seq defines positions and levels of engaged Pol II
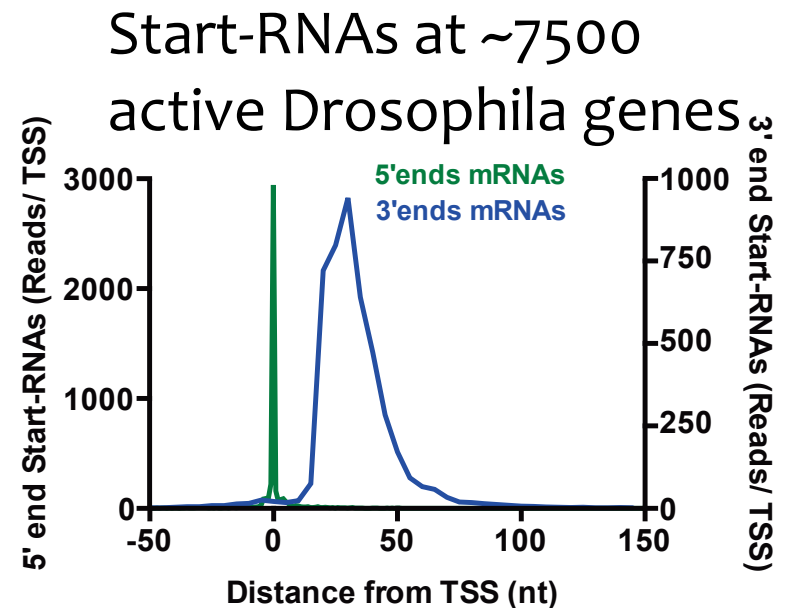
Isolation and sequencing of RNAs that are:
- Associated with stably paused Pol II
- Nuclear/ chromatin associated
- Short (<100 nt)
- Capped

Start-RNAs at ~7500 active Drosophila genes
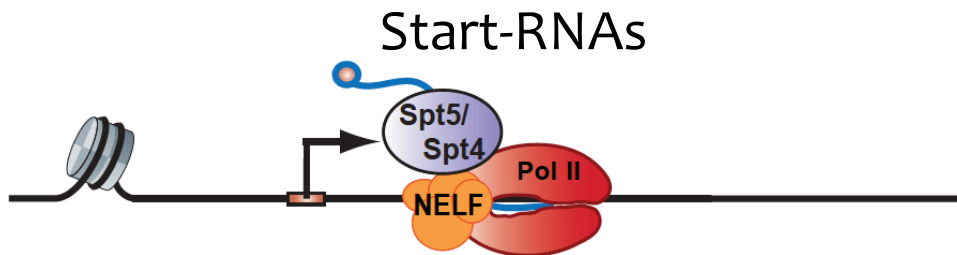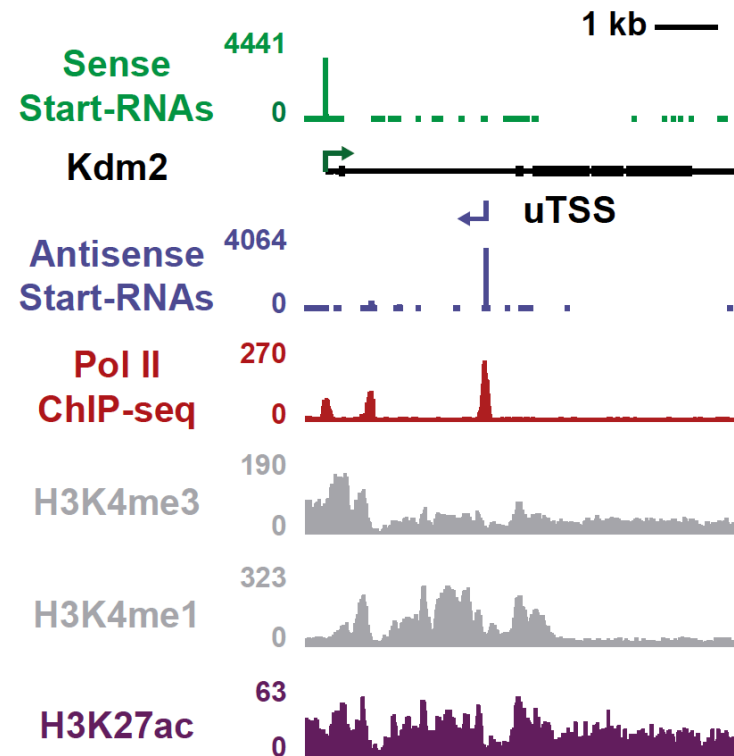

Start-RNAs:
5' ends – indicate Transcription Start Site
3' ends – reveal the position of paused polymerase

This technique is independent of the stability of RNA transcript

Start-seq in Drosophila S2 cells identifies ~10,000 annotated TSS and ~12,000 unannotated TSSs

TSScall is the newly developed unbiased algorithm used to define individual TSSs and clusters of TSSs (by subtraction of the annotated set)

Henriques et al. (2018) *Genes & Dev..*

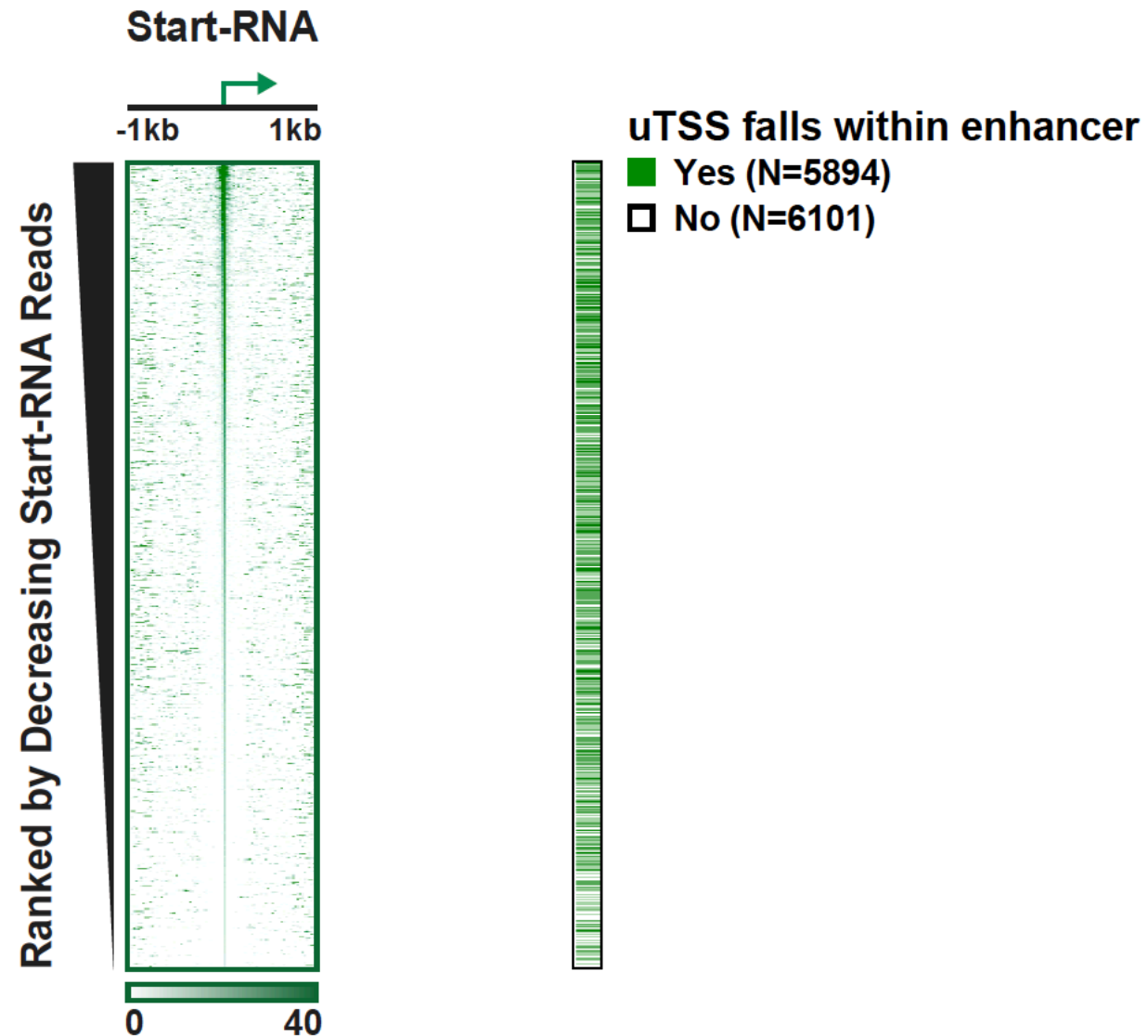# Compare set of uTSS with list of functionally defined enhancers

Previous work has identified 11,364 functional enhancers in this same cell type using high-throughput enhancer-reporter assays.
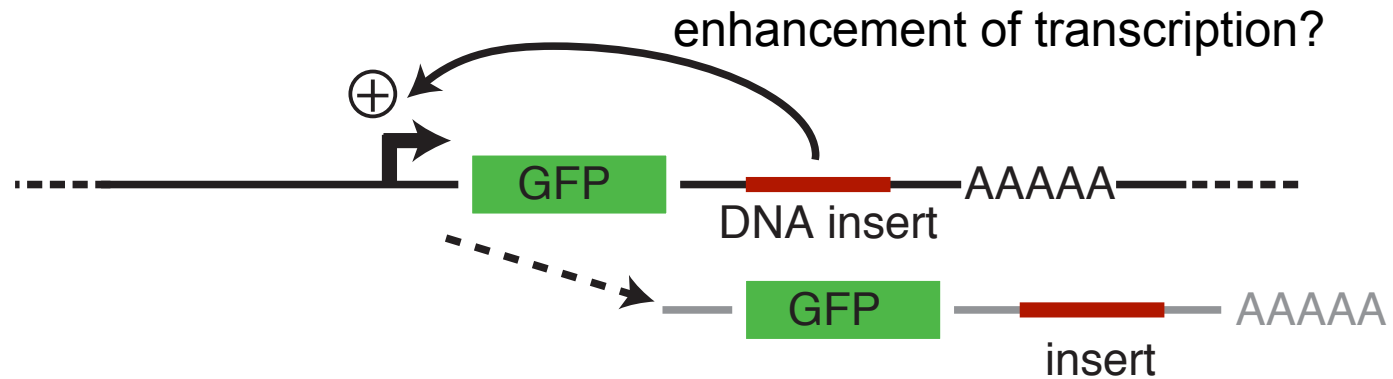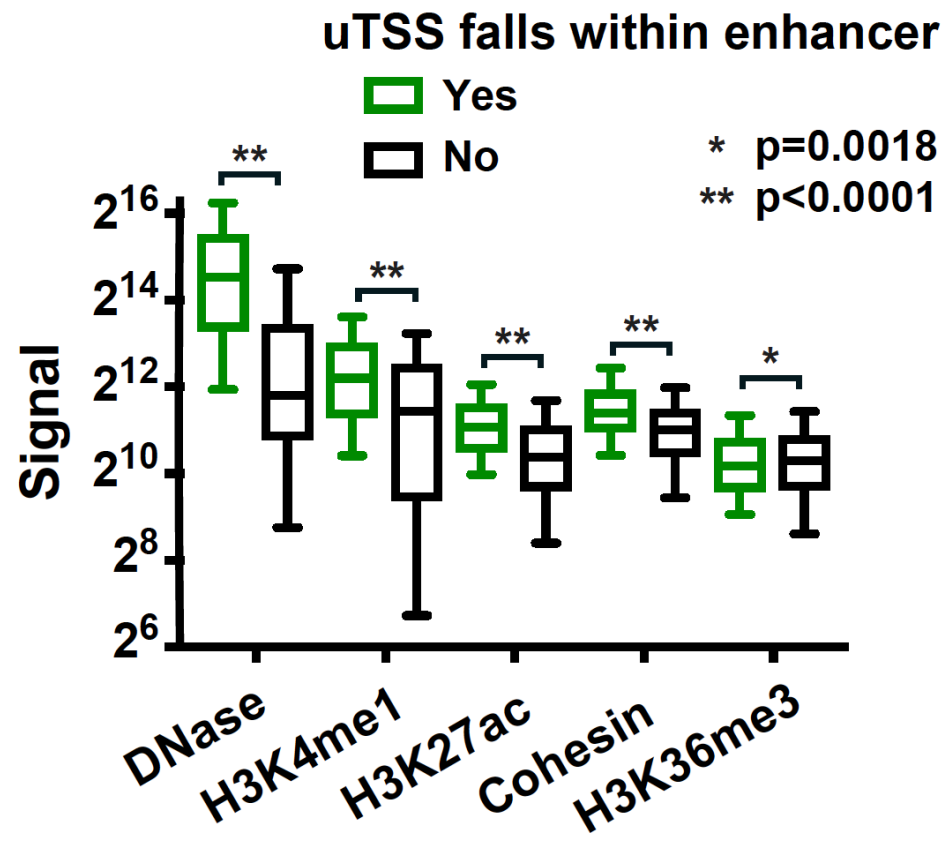

STARR-seq= Self Transcribing Active regulatory Region

Arnold et al. (2013) Science
Zabidi et al. (2015) Nature

Of ~12,000 unannotated TSSs identified in Drosophila S2 cells 50% fall within functionally-defined enhancers

Start-RNA

-1kb 1kb

Ranked by Decreasing Start-RNA Reads

0 40

uTSS falls within enhancer
■ Yes (N=5894)
□ No (N=6101)

Unannotated TSSs within enhancers exhibit enhancer-like features

# Clusters of unannotated TSSs are reminiscent of super-enhancers
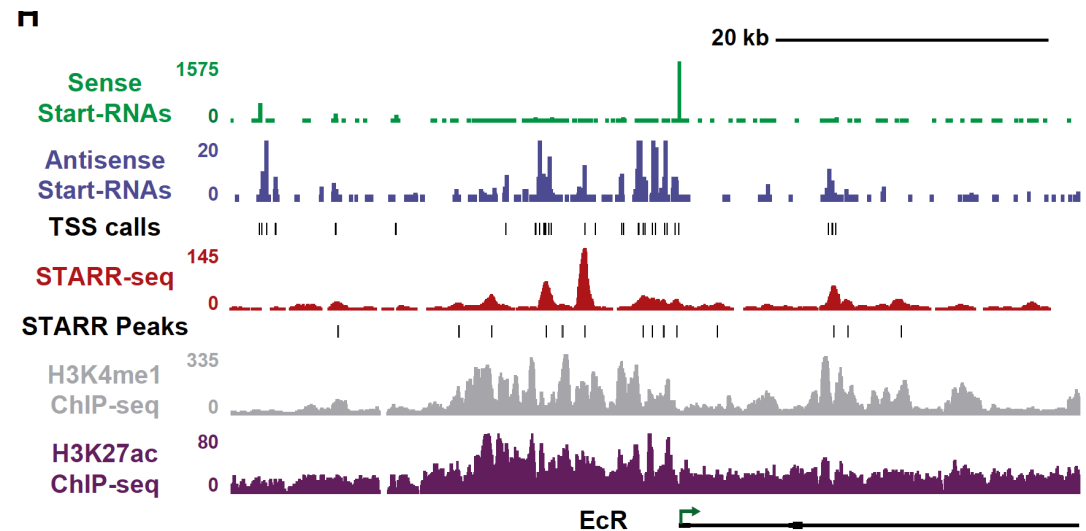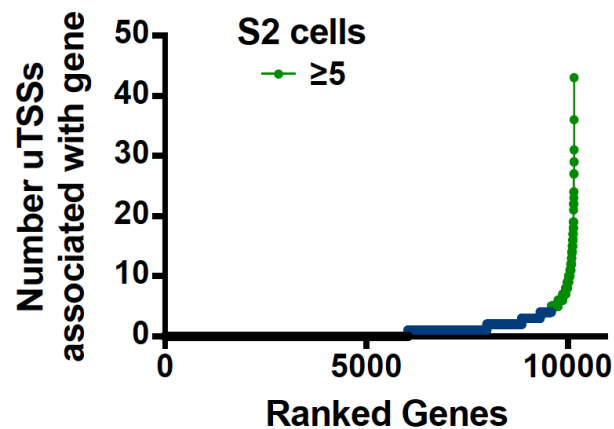
Ecdysone Receptor
(Fly steroid hormone that works across development)

Finding unannotated TSSs could be a very effective way of identifying enhancers and super-enhancers

- ~50% of unannotated TSSs fall in enhancers
- Super-enhancer clusters are readily detected
- Not every transcription initiation event identifies a regulatory element with enhancer activity, but many do!

- By comparison, ~25% of ENCODE (ChromHMM) predicted enhancers have enhancer activity
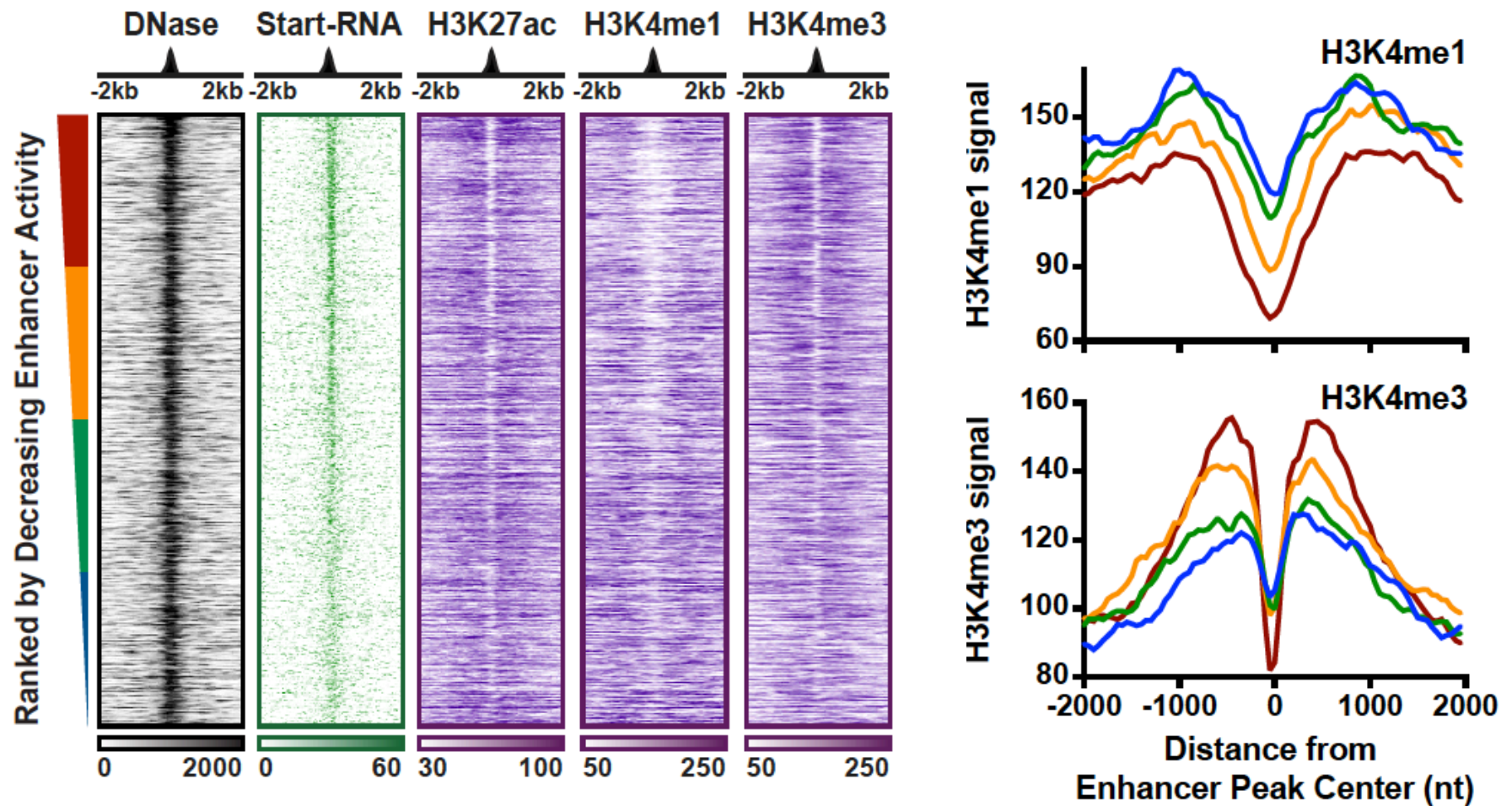
~95% of functionally-defined Drosophila enhancers generate short nascent RNAs
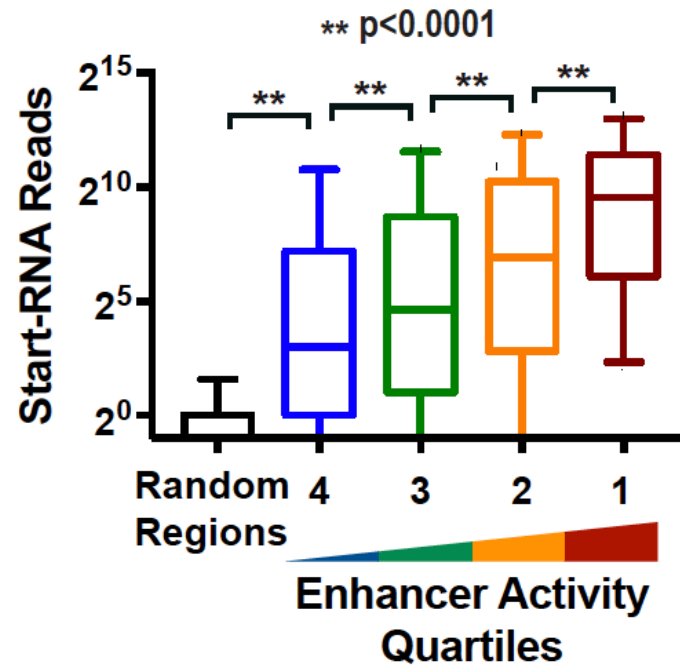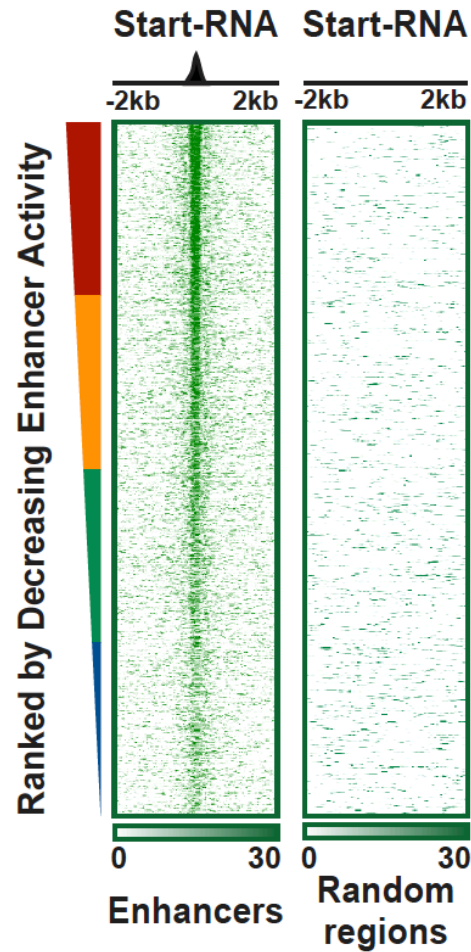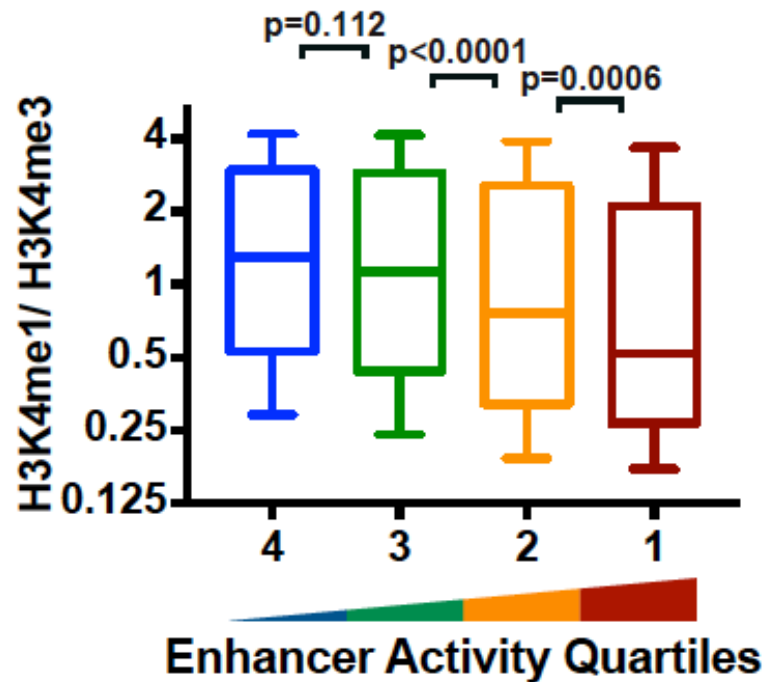

Nascent RNAs identify enhancer location and activity

N=11,364 regions with enhancer activity

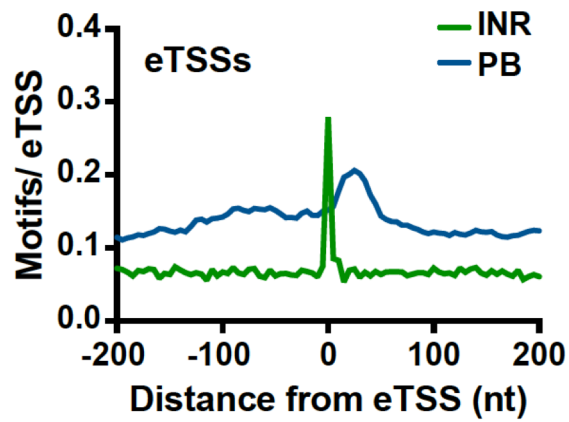# The most active enhancers display H3K4me3, rather than H3K4me1

Core et al. (2014) *Nat. Genet.*

# H3K4me1 / me3 ratios are anti-correlated with enhancer activity and transcription levels

Caution: The use of H3K4me1 to identify enahncers could bias towards those with low activity

Core promoter elements are conserved at enhancer transcription start sites (eTSSs)
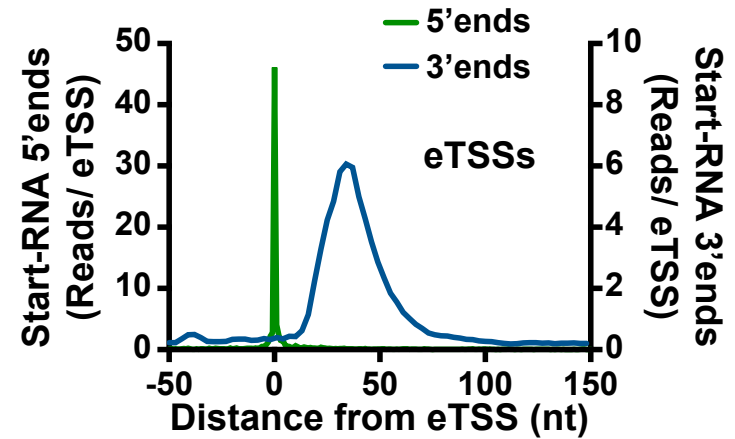

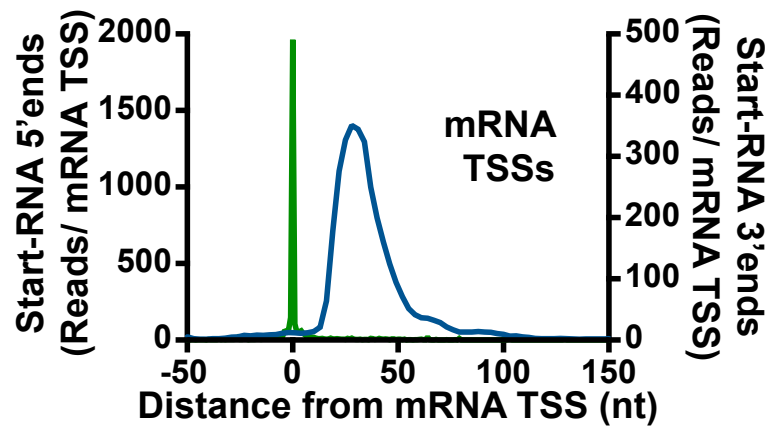INR= Initiator element located at TSS

PB= Pause Button Implicated in pausing during early elongation, located at +20 to +30 nt
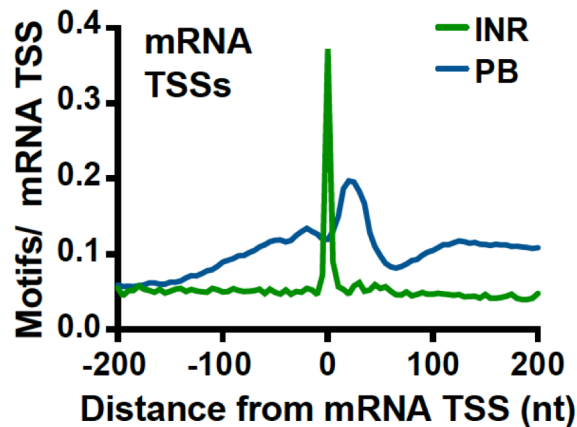
Motifs that impact initiation and early elongation are present at enhancers

TSS-proximal pausing occurs at enhancers

Inhibition of P-TEFb with Flavopiridol significantly increases Pol II occupancy of enhancers
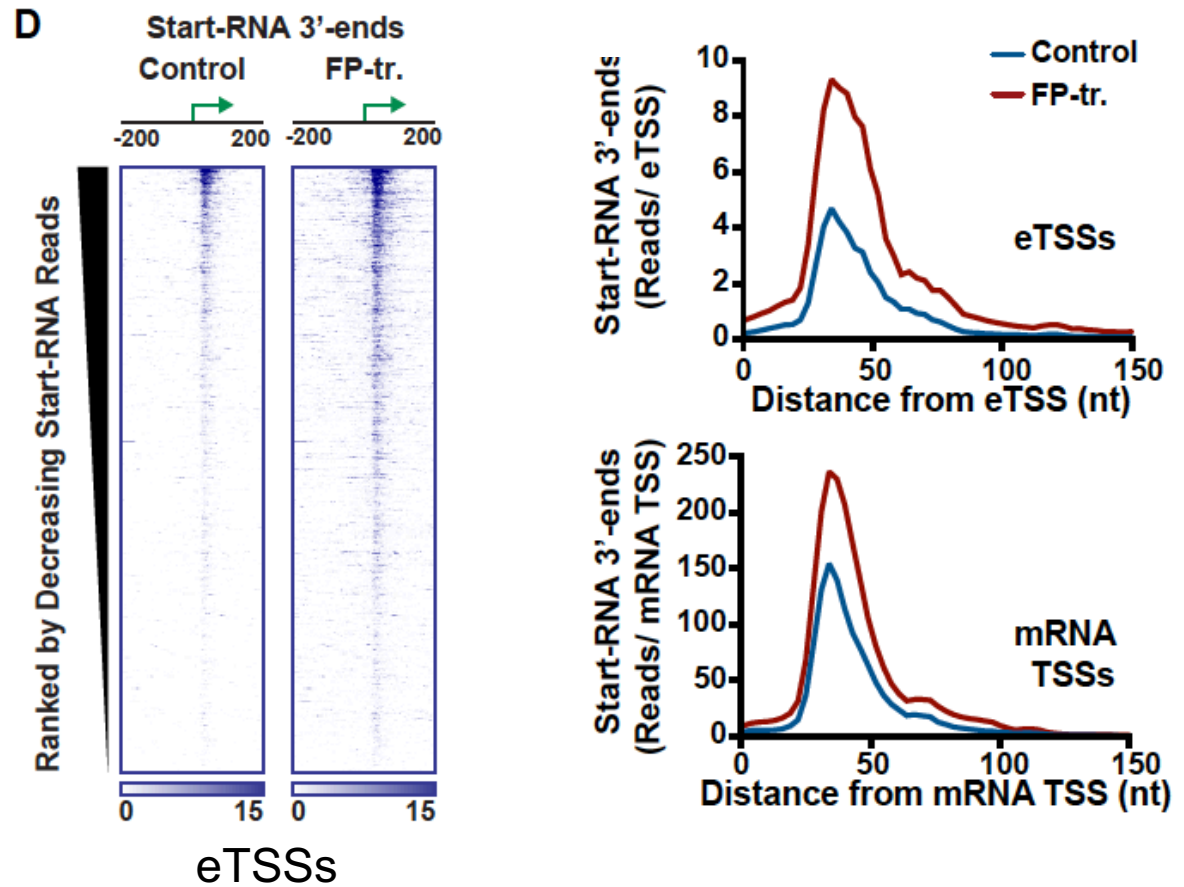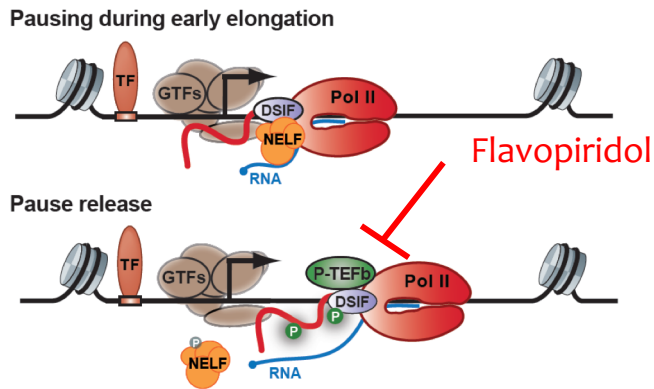
Pol II activity at enhancers requires
P-TEFb-mediated pause release

Does DSIF (Spt4/Spt5) play a role at enhancers?
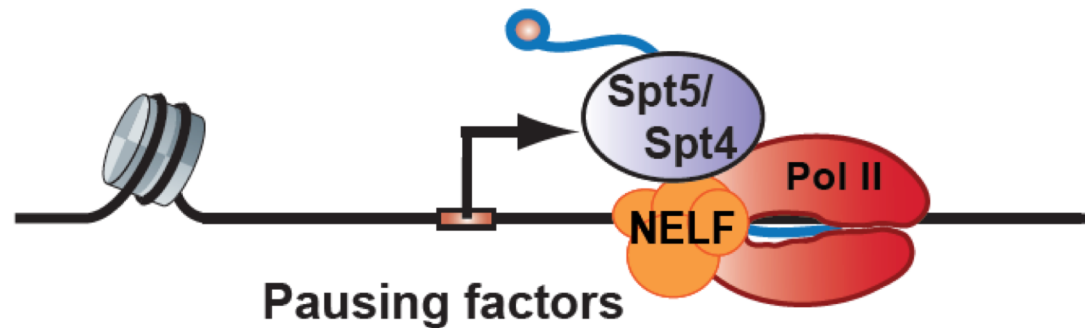
- Spt5 is critical for stabilizing paused Pol II

- After pause release, Spt5 remains associated with Pol II and recruits chromatin and RNA processing factors

Depletion of elongation factor Spt5 strongly reduces Pol II occupancy and transcription at enhancers
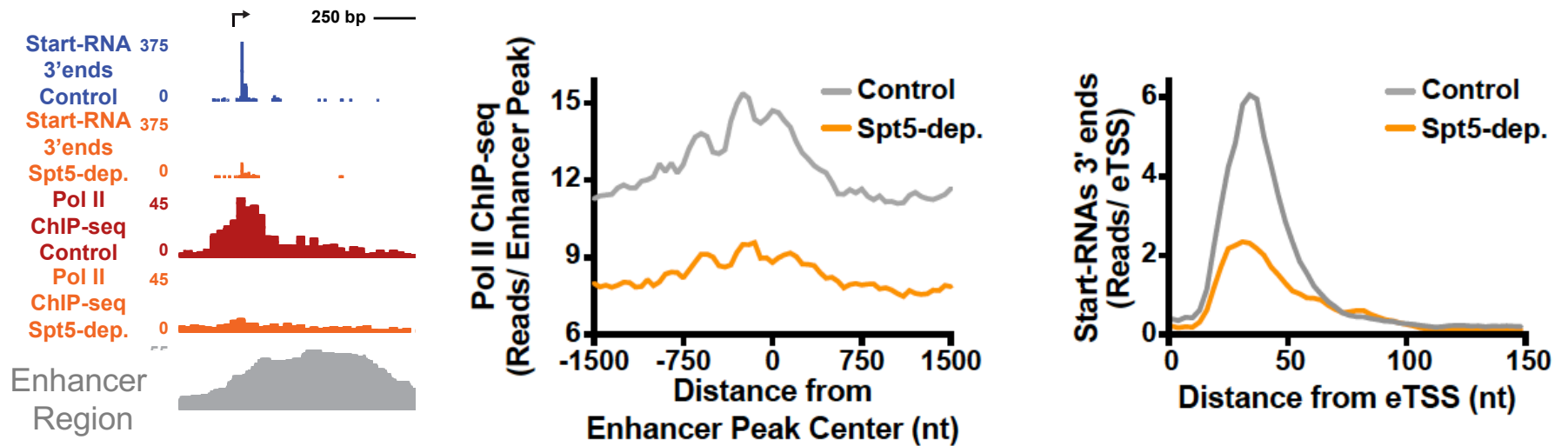
Transcription elongation machinery is critical for eRNA synthesis

How stable is Pol II at enhancers?

Triptolide blocks TFIIH
helicase and new initiation

Productive

elongation

Spt5/
Spt4

Pol II

NELF

Pausing factors

Termination
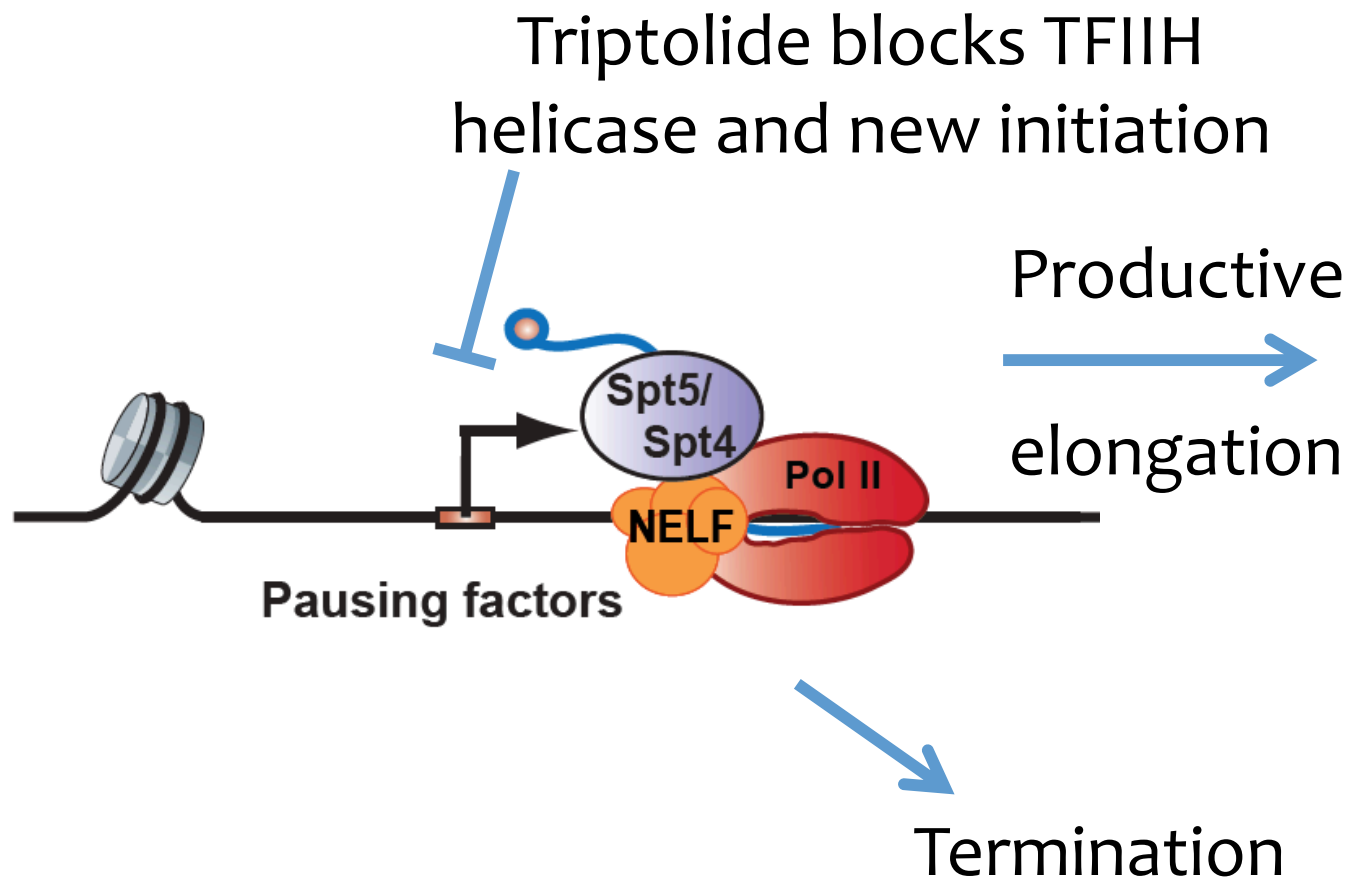
Pol II at enhancers is often less stable than at promoters

How measure stability?

- Start-seq performed over time course of Trp treatment
- Data from promoters and enhancers was clustered together
- Why is enhancer Pol II so unstable?
- Is it rapidly released into productive elongation? Or does it terminate in early elongation?
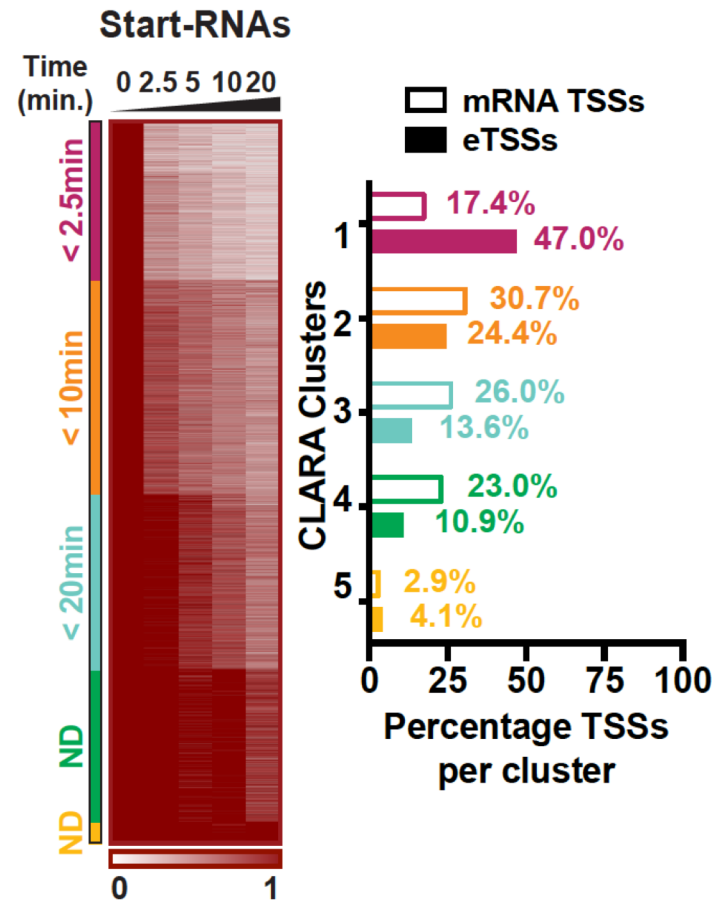
Data from Krebs et al. (2017) *Mol. Cell*

Nascent RNA sequencing techniques
show very little productive elongation at enhancers
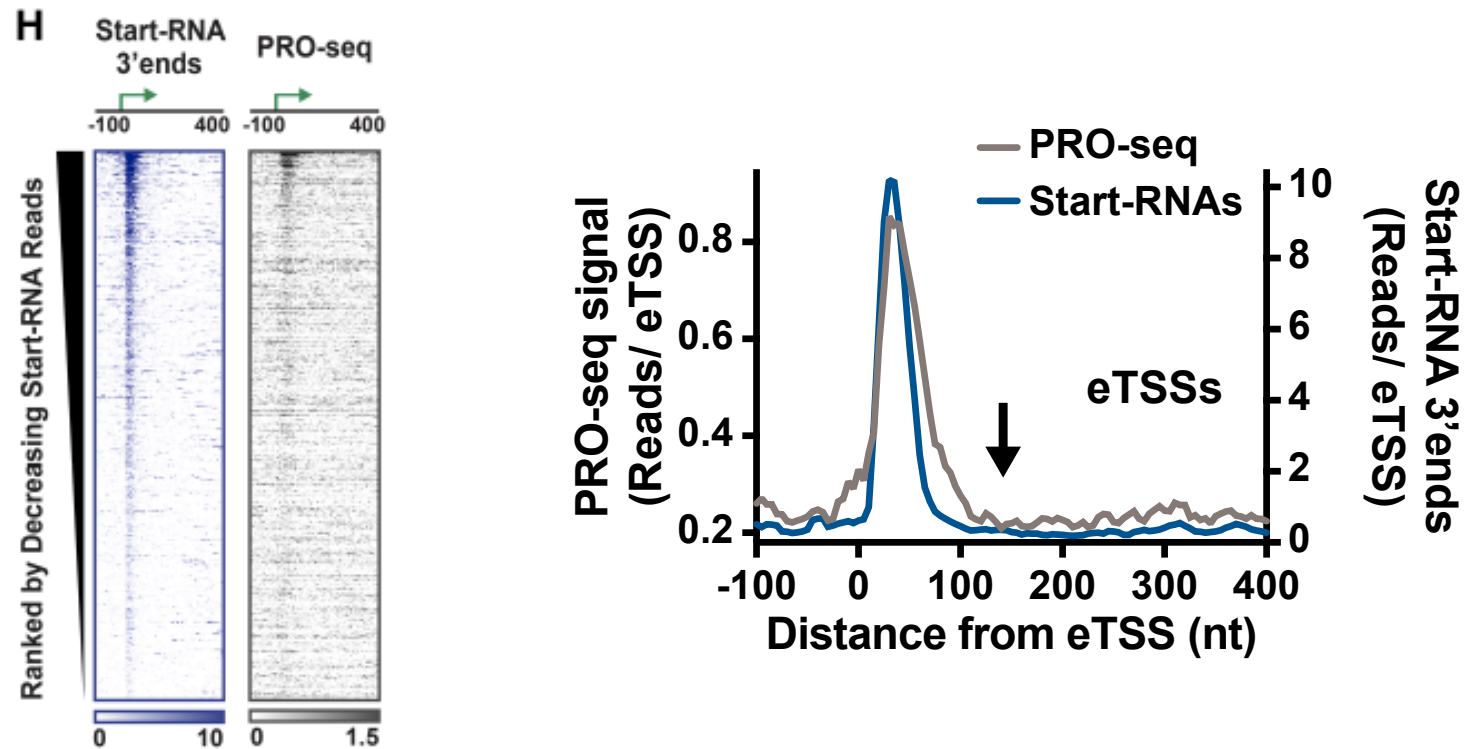
Take home:
eRNAs are very short and very short-lived

1-1000kb

enhancer   mRNA promoter
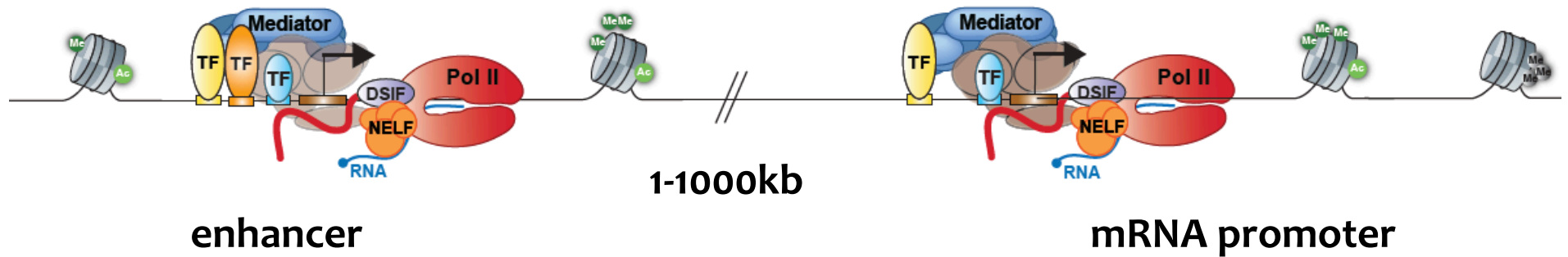
- Nascent RNA production very effectively delineates enhancer location and activity

- Core promoter sequences and elongation factors are present at enhancers

- Regulation of enhancer transcription appears nearly identical to that of protein coding genes, up to the step of pausing

- However, Pol II at enhancers is subject to termination very near the TSS, yielding only short and unstable RNAs

BREAK

AGENDA

- Research Paper
  (including discussion about promoter/enhancer similarities)

- Literature Review on Transcriptional Regulation during Adipogenesis

Adipogenic differentiation
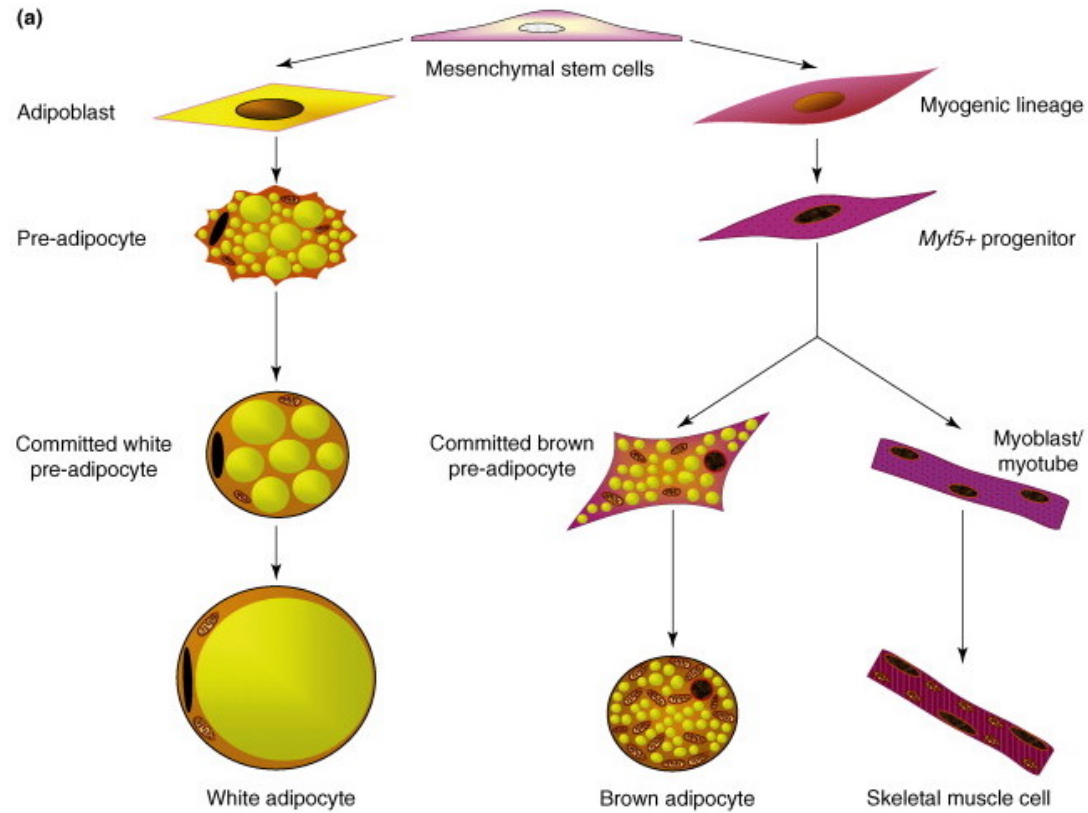
Transcriptional regulation of adipogenesis

Extensive chromatin remodelling and establishment of transcription factor 'hotspots' during early adipogenesis
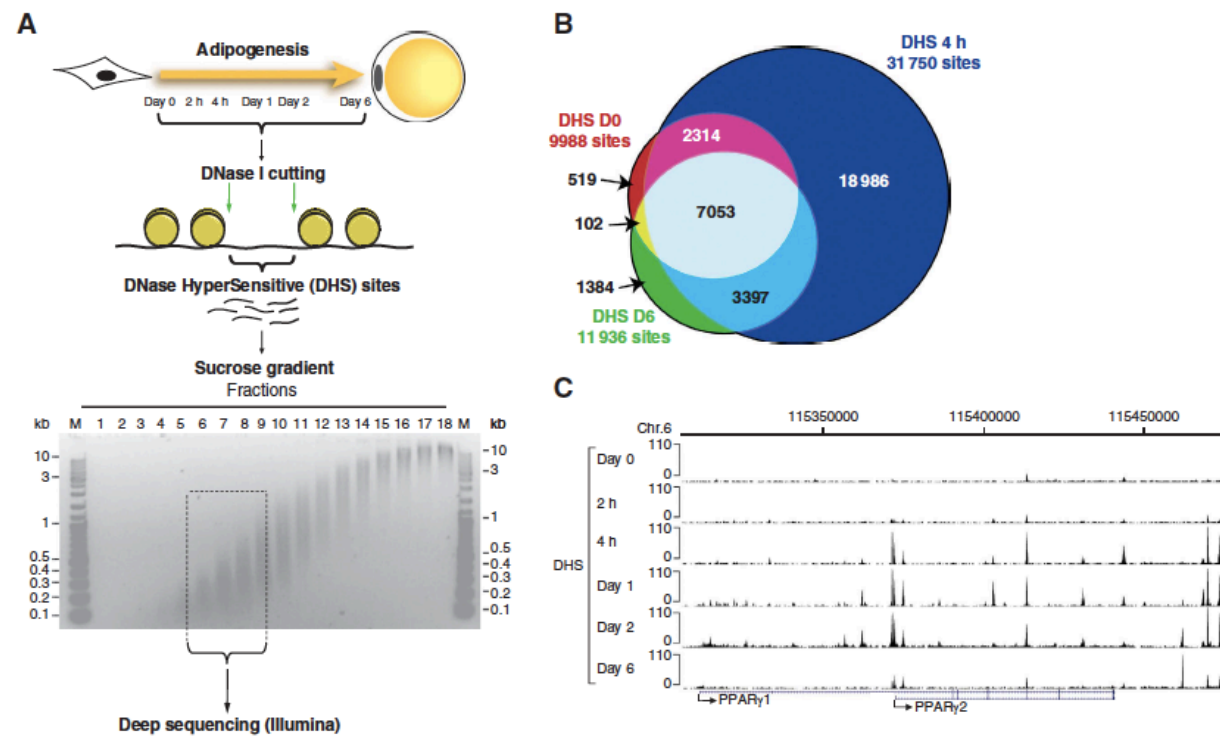
Figure 1 DNase I HyperSensitive (DHS) site analysis during 3T3-L1 adipogenesis. (**A**) Experimental outline. Nuclei from 3T3-L1 cells were isolated at the indicated time points of adipocyte differentiation and subsequently subjected to a limited DNase I digestion. Small DNA fragments where DNase I cut twice (i.e., DHS sites) were purified o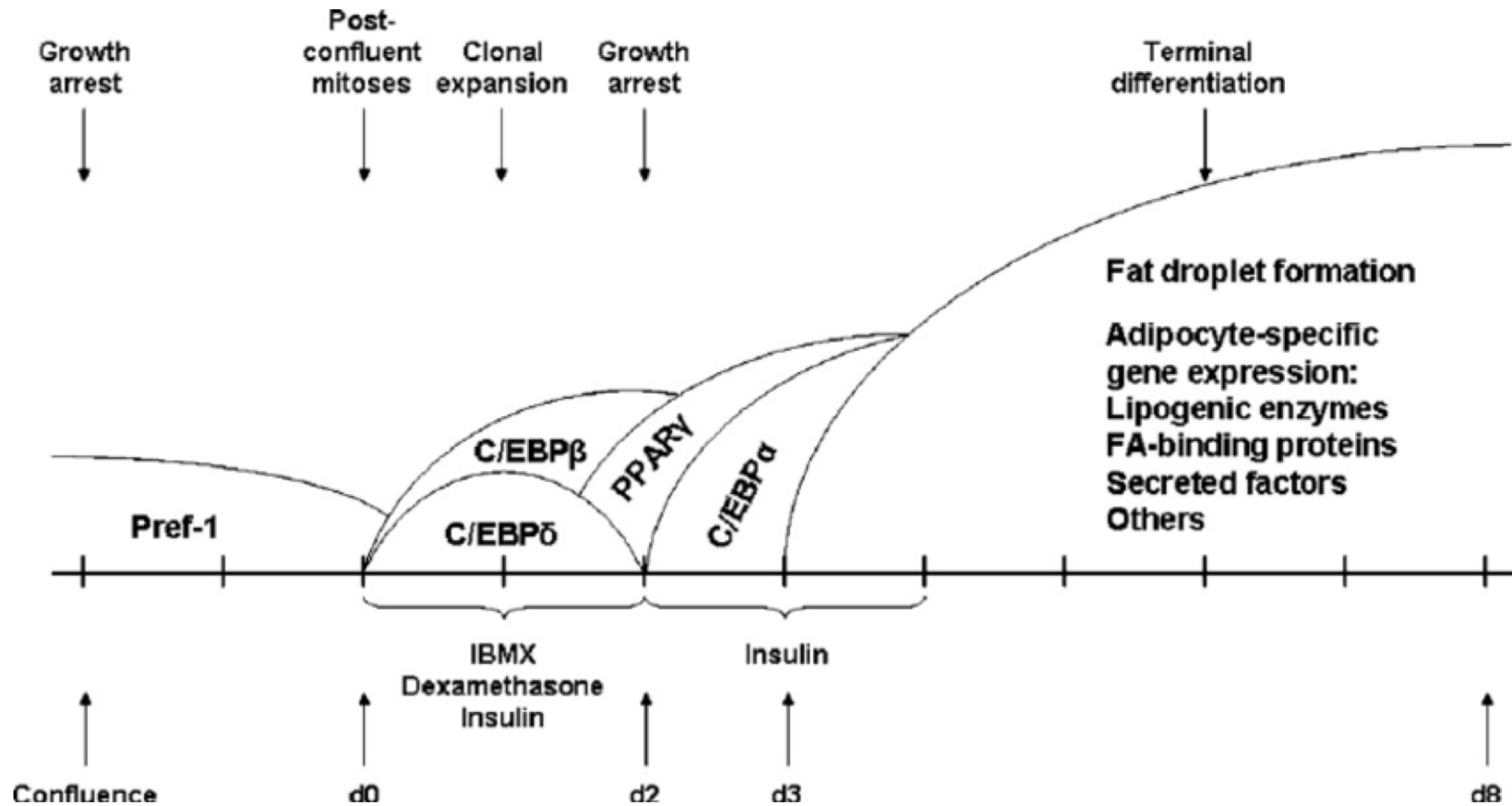ver a sucrose gradient and subsequently subjected to deep sequencing using the Illumina platform. (**B**) Venn diagram representing the overlap between DHS sites in pre-adipocytes immediately before induction of differentiation (day 0; red), in adipocytes (day 6; green), and in cells stimulated for 4 h with the differentiation cocktail (blue). Sizes of the circles are proportional to the number of sites. (**C**) DHS-seq data at the *PPARγ* locus viewed in the UCSC genome browser.

C

Chr.6

C/EBPδ
- Day 0 — 180 to 1
- 4 h — 180 to 1

C/EBPβ
- Day 0 — 220 to 1
- 4 h — 220 to 1
- Day 2 — 220 to 1

PPARγ
- Day 0 — 150 to 1
- Day 1 — 150 to 1
- Day 2 — 150 to 1
- Day 6 — 150 to 1

DHS
- Day 0 — 80 to 0
- 4 h — 80 to 0
- Day 6 — 80 to 0

Cav2   Cav1

PPARγ target sites at day 6

D

No DHS at 4 h
3692 sites (67%)

1324 sites (24%)

477 sites (9%)

DHS at 4 h

■ No C/EBPβ at 4 h   ■ C/EBPβ at 4 h
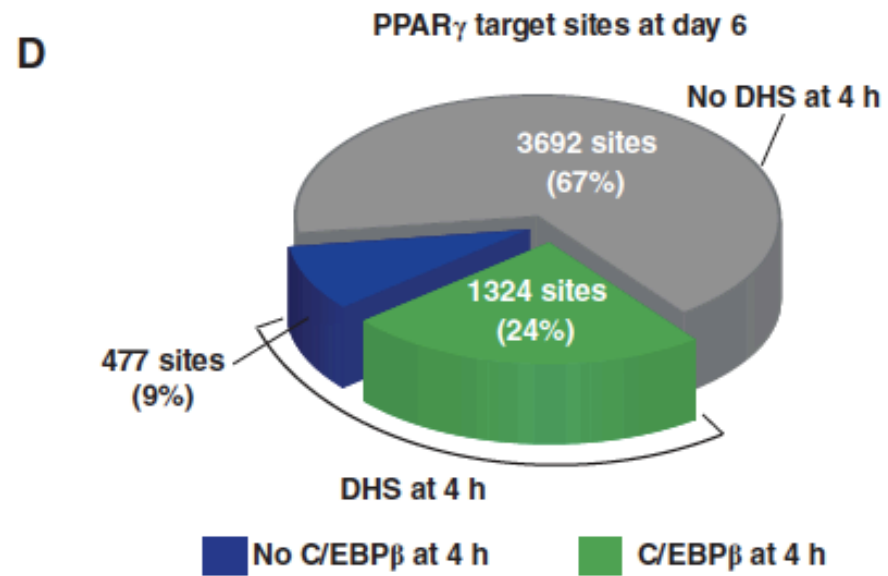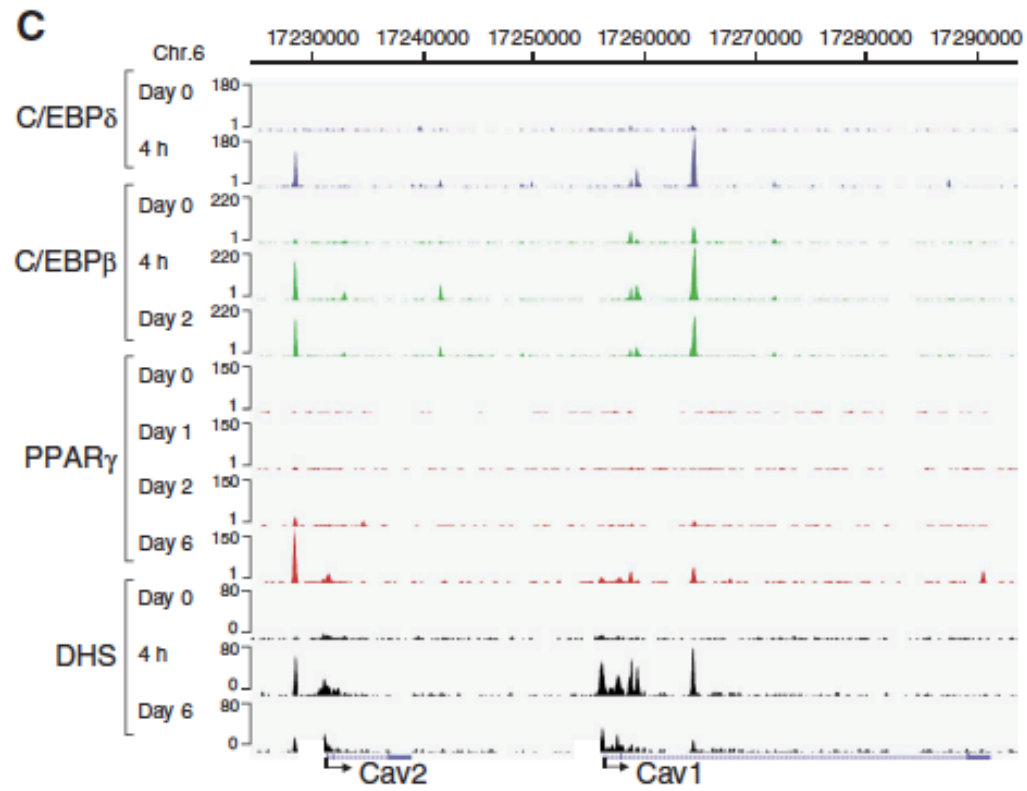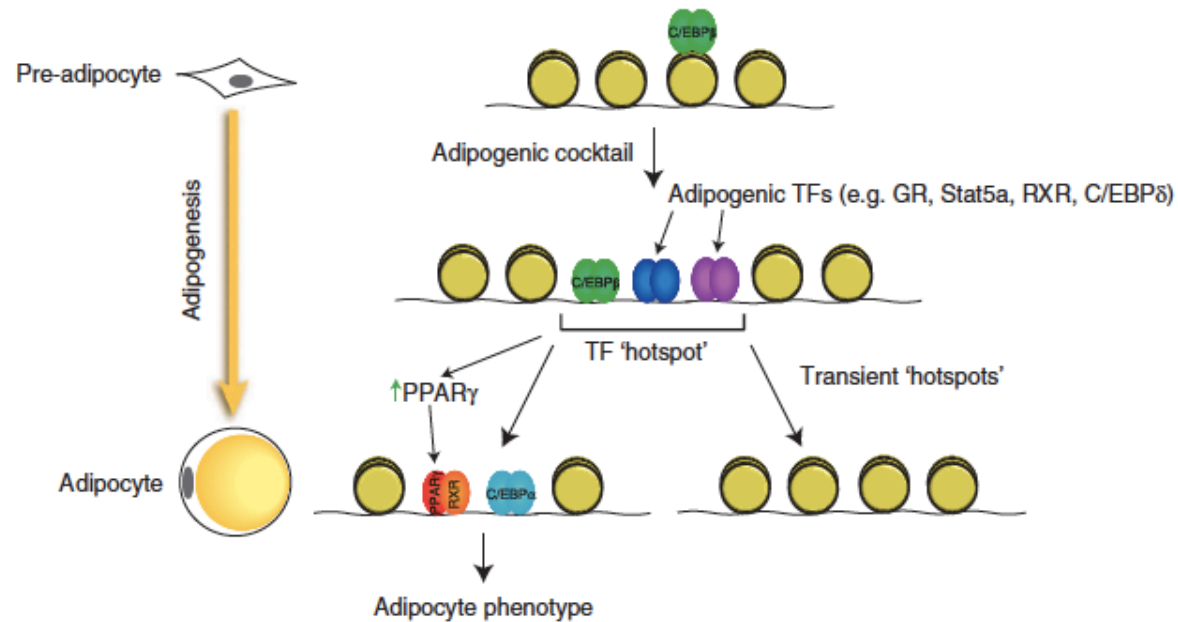
Figure 9 Model illustrating C/EBPβ as a pioneering factor for adipogenic transcription factors and chromatin remodelling. C/EBPβ binds to closed chromatin in pre-adipocytes. Upon induction of differentiation, several other transcription factors are activated and recruited to C/EBPβ sites, resulting in remodelling of the chromatin structure and formation of transcription factor 'hotspots'. Some of these are transient in nature, whereas others persist throughout differentiation and are later occupied by PPARγ and C/EBPα, which induce the mature adipocyte phenotype.

Article

Transcription Factor Cooperativity in Early Adipogenic Hotspots and Super-Enhancers

Rasmus Siersbæk,[1,2] Atefeh Rabiee,[1,2] Ronni Nielsen,[1] Simone Sidoli,[1] Sofie Traynor,[1] Anne Loft,[1] Lars La Cour Poulsen,[1] Adelina Rogowska-Wrzesinska,[1] Ole N. Jensen,[1] and Susanne Mandrup[1,*]
[1]Department of Biochemistry and Molecular Biology, University of Southern Denmark, 5230 Odense M, Denmark
[2]Co-first author
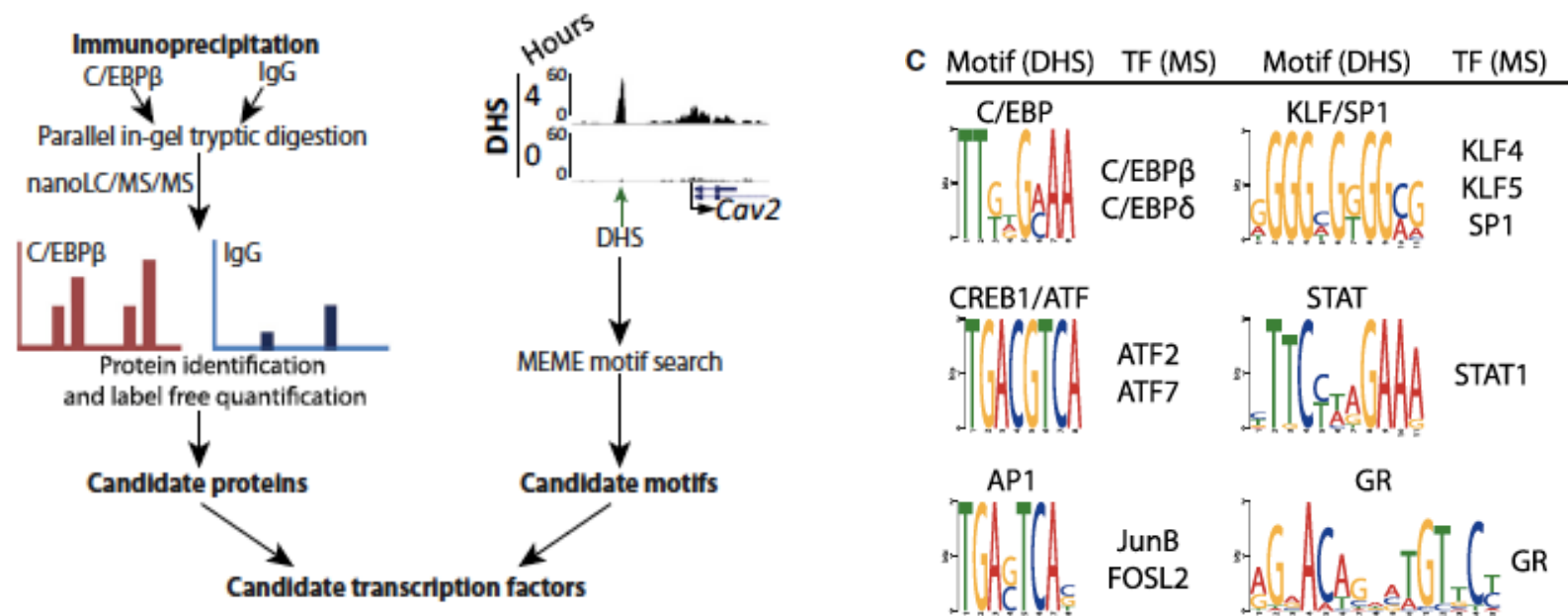
Figure 1. Schematic Overview of the Combined Genomics and Proteomics Approach Used to Identify Key Early Regulators of Adipocyte Differentiation

Motif analyses of DNA sequences at DHS sites 4 hr after induction of differentiation of 3T3-L1 cells obtained from previous analyses (Siersbæk et al., 2011) were combined with proteomics analyses of C/EBPβ-associated proteins to confidently identify candidate transcription factors involved in early adipogenic reprogramming.
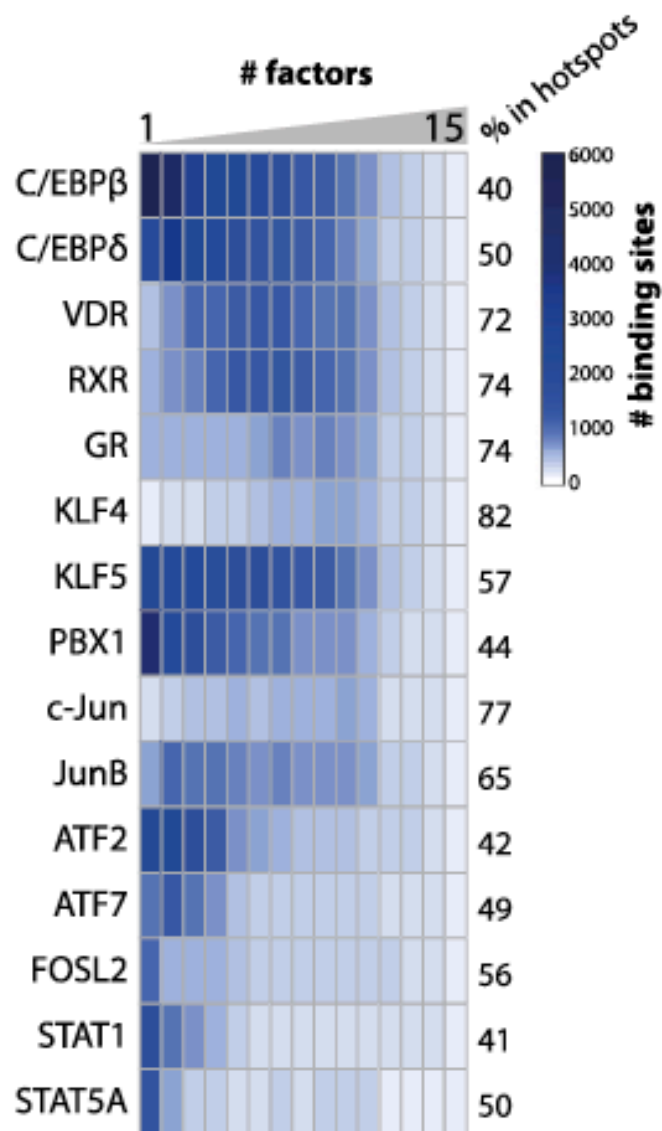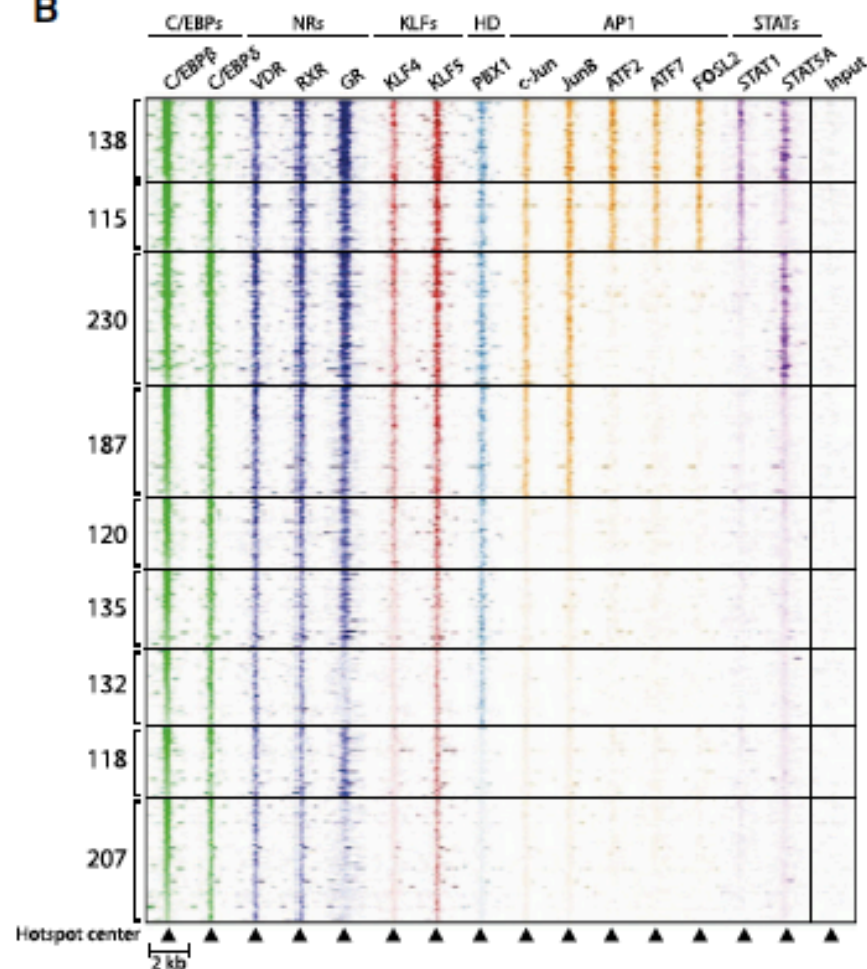
Figure 3. Transcription Factors in the Early Adipogenic Network Colocalize at Transcription Factor Hotspots

(A) For each factor, the number of binding sites that are occupied by one (only the factor itself) to 15 factors is shown along with the percentage of binding sites that are located in hotspot regions. The numbers of all transcription factor binding sites based on the ChIP-seq data sets and all binding sites redistributed randomly in the genome that are occupied by one to 15 factors are shown at the bottom.

(B) Heatmap of transcription factor binding in a 2 kb region around the center of the nine largest groups of hotspots. Input signal (Siersbæk et al., 2011) is shown as a control.
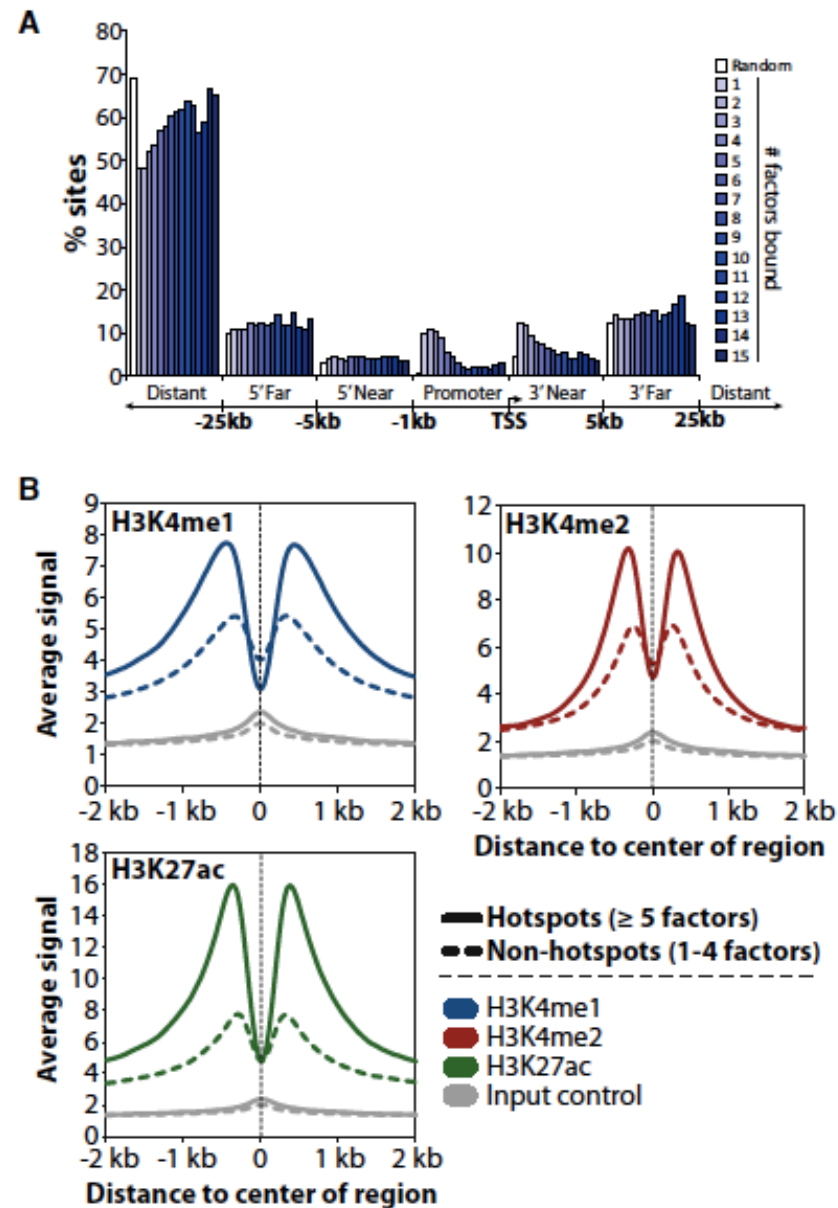
Figure 4. Extensive Transcription Factor Cooperativity at the Level of Hotspots

(A) Location of transcription factor binding sites occupied by one to 15 factors relative to the transcription start site (TSS) of RefSeq genes. The location of randomly placed binding sites of the same size is shown as a reference.

(B) The level of three histone marks characteristic of enhancers regions (i.e., H3K4me1, H3K4me2, and H3K27ac) in the vicinity of distal (>2 kb away from the TSS) non-hotspots (occupied by one to four factors) and hotspots (occupied by ≥5 factors). Input (Siersbæk et al., 2011) is shown as a control.
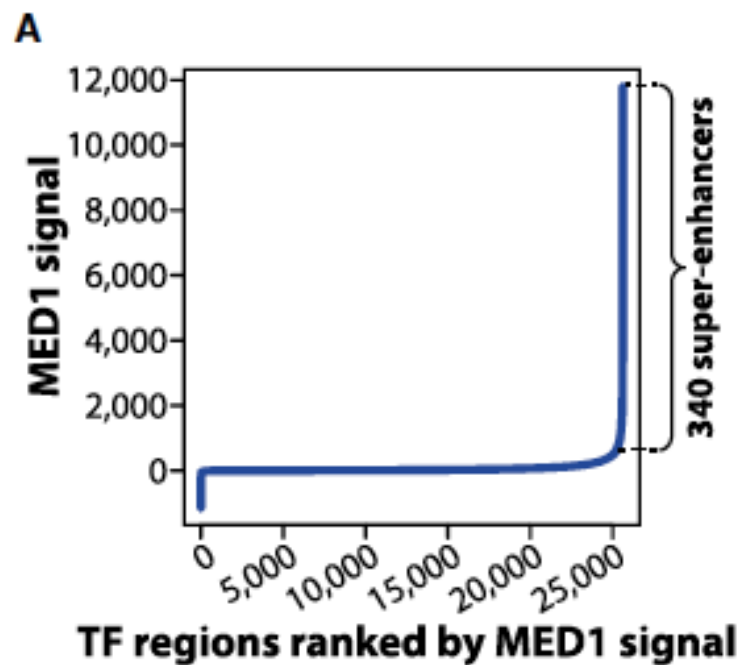
Figure 5. Hotspots Are Enriched in Early Adipogenic Super-Enhancers

(A) All the identified transcription factor binding sites (54,724) that were within 12.5 kb of each other were merged, resulting in 25,632 regions. These regions were ranked by their MED1 signal, where the input background (Siersbæk et al., 2011) had been subtracted. Regions with a MED1 signal (minus background) above 700 reads per 10 M total reads were defined as super-enhancers. All other regions were denoted as regular transcription factor binding regions.
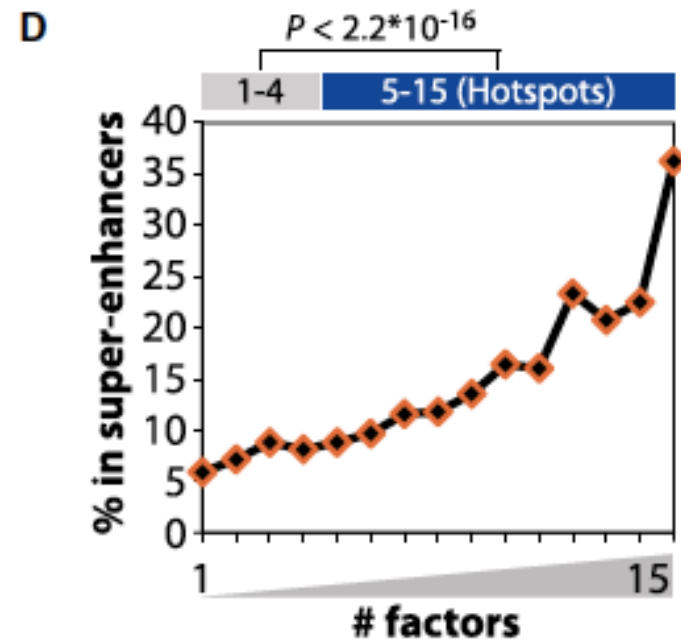
(D) Fraction of transcription factor binding sites occupied by one to 15 factors that are found in super-enhancer regions. The significance of the higher occurrence of hotspots (i.e., binding sites occupied by at least five factors) relative to non-hotspots (i.e., binding sites occupied by one to four factors) within super-enhancer regions as determined by Fisher's exact test is shown at the top.
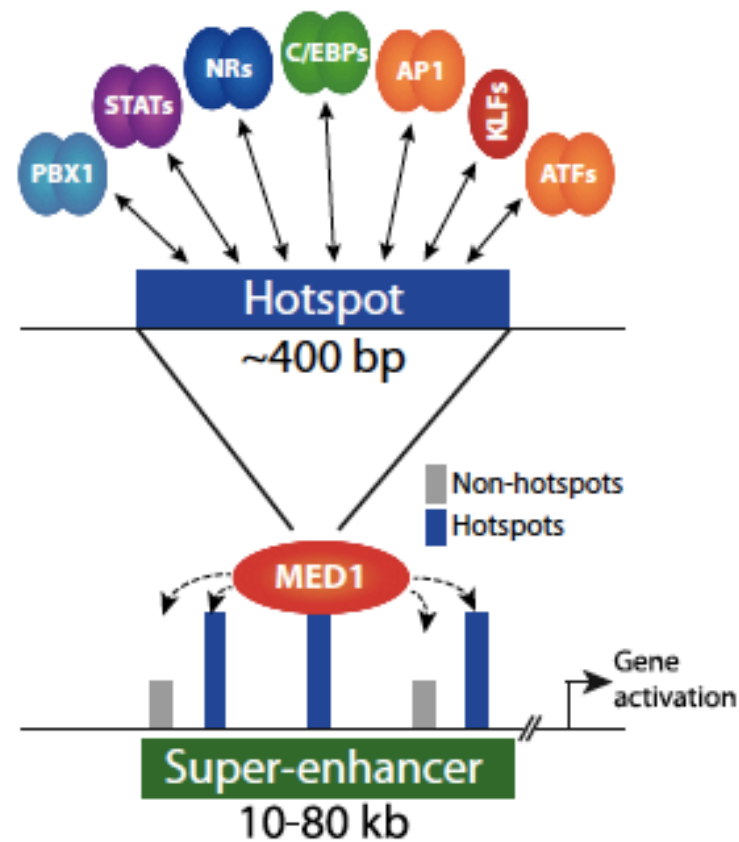
Figure 7. Model of Transcription Factor Cooperativity in Adipogenic Hotspots and Super-Enhancers

Multiple diverse transcription factors colocalize at small genomic regions termed transcription factor hotspots (~400 bp), which are central constituents in large super-enhancers (10–80 kb). Super-enhancers are characterized by very high levels of MED1 recruitment, and several lines of evidence suggest that constituents within super-enhancers cooperate to recruit MED1. Ultimately, establishment of super-enhancer regions results in activation of nearby genes characteristic of the early phase of adipogenesis.

Molecular Architecture of Transcription Factor Hotspots in Early Adipogenesis

Rasmus Siersbæk,[1] Songjoon Baek,[2] Atefeh Rabiee,[1] Ronni Nielsen,[1] Sofie Traynor,[1] Nicholas Clark,[3] Albin Sandelin,[3] Ole N. Jensen,[1] Myong-Hee Sung,[2] Gordon L. Hager,[2] and Susanne Mandrup[1,*]
[1]Department of Biochemistry and Molecular Biology, University of Southern Denmark, 5230 Odense M, Denmark
[2]Laboratory of Receptor Biology and Gene Expression, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, MD 20892, USA
[3]The Bioinformatics Centre, Department of Biology & Biotech Research and Innovation Centre, University of Copenhagen, 2200 Copenhagen N, Denmark

Digital Genomic Footprinting Reveals Precise Protein Footprints at a Genome-wide Level

To begin to understand how transcription factors communicate in hotspots, it is essential to know how the factors are organized at these regions, including which factors are engaged in direct DNA interactions. We therefore employed high-resolution digital genomic footprinting (Figure 1A), a recently developed method to identify protein footprints (i.e., areas of restricted nuclease access) within DNase I hypersensitive (DHS) regions based on ultradeep sequencing (>100 M sequence tags) of DHS-seq
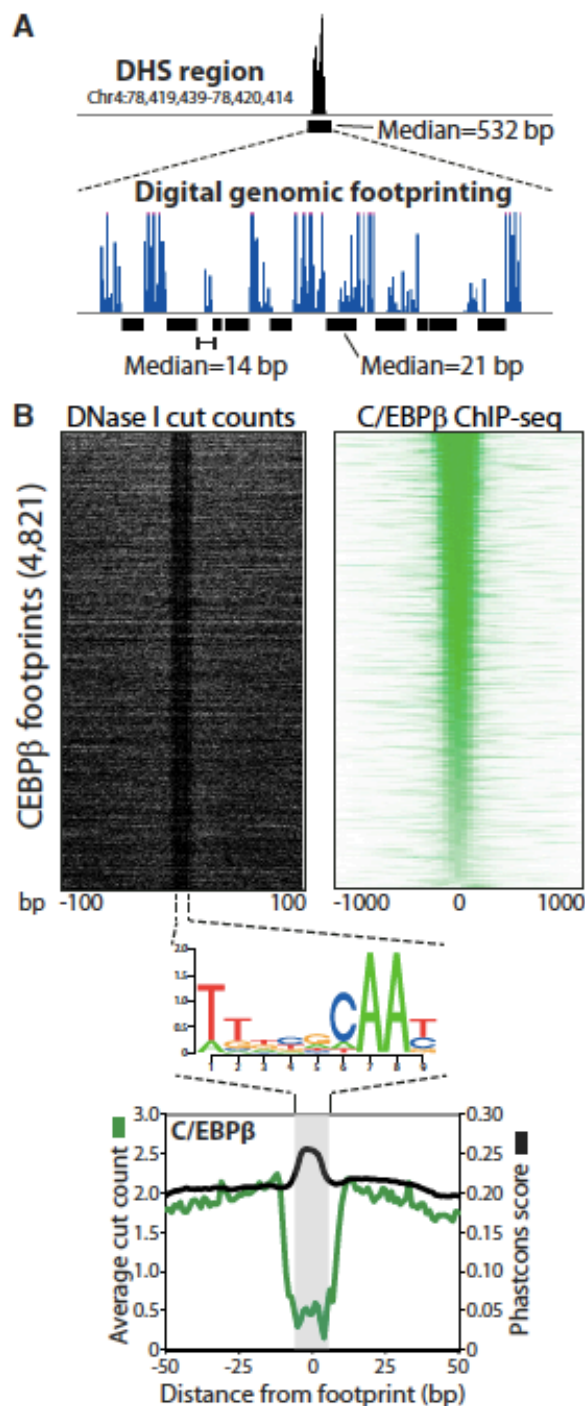
Figure 1. Digital Genomic Footprinting Reveals Transcription Factor Footprints at a Genome-wide Level

(A) Schematic overview of digital genomic footprinting. Extensive sequencing reveals small protected areas of 8–30 bp within overall DNase I hypersensitive regions corresponding to protein footprints. The median size of DHS regions and footprints as well as the median distance from one footprint to the nearest neighbor obtained from Figure S1 are indicated.

(B) DNase I cut counts (left) and C/EBPβ ChIP-seq signal (right) in the vicinity of C/EBPβ footprints (top). These regions were defined as footprints containing a C/EBP predicted site that overlap a ChIP-seq peak for C/EBPβ. Note the different scales used for visualization of DNase I cut counts and ChIP-seq data. Average DNase I cut counts and phastCons score in the vicinity of C/EBPβ footprints are shown at the bottom.
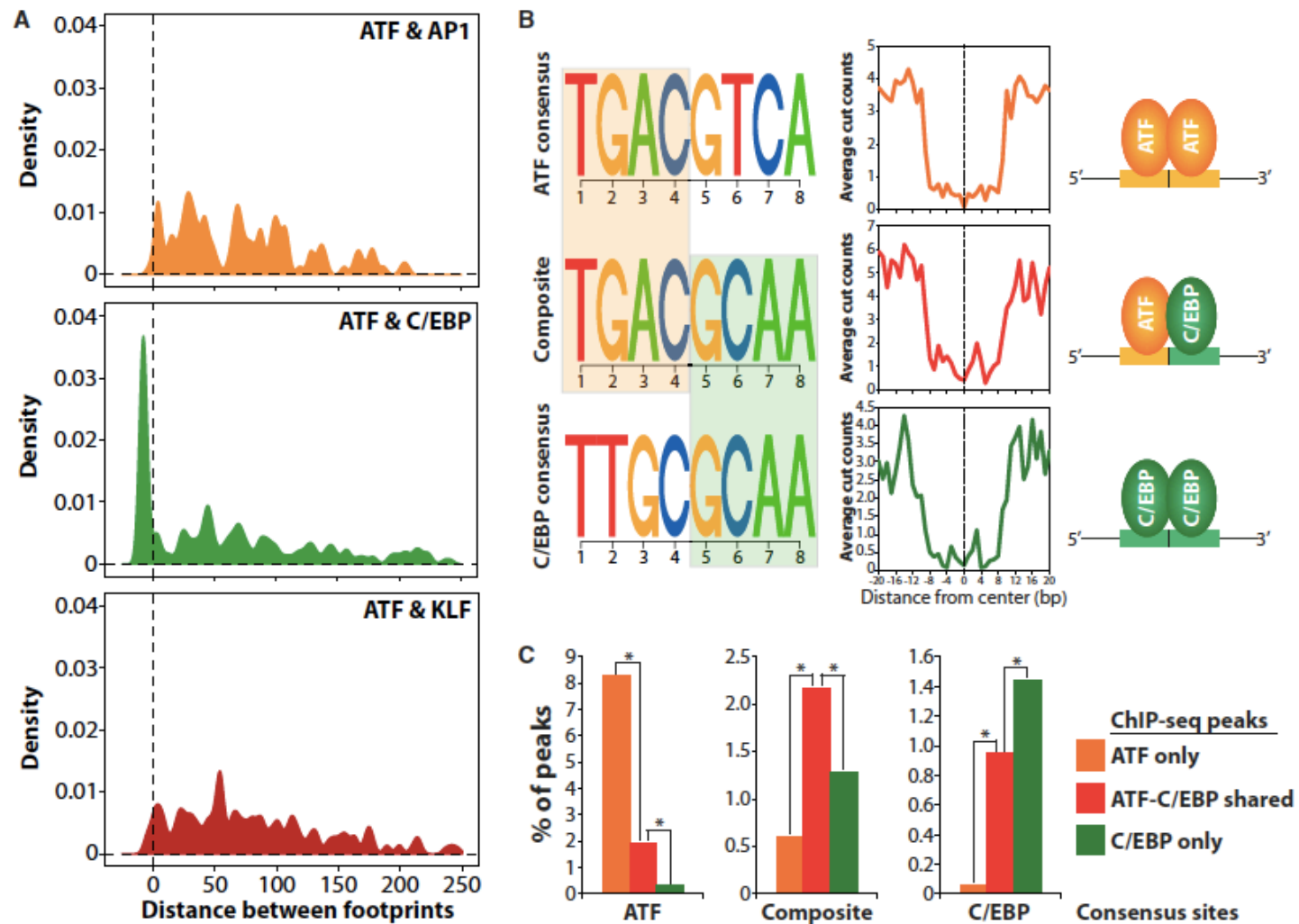
Figure 2. Footprint at a Composite Motif Demonstrates ATF-C/EBP Heterodimer Formation on Chromatin

(A) Distance between footprints for different transcription factor pairs (i.e., ATFs and AP1, ATFs and C/EBPs, and ATFs and KLFs) found within the same DHS sites. Negative distances mean that the footprints for the two types of factors overlap. Here, footprints refer to predicted binding sites found in footprint regions overlapping a ChIP-seq peak.

(B) Consensus binding sites for ATFs and C/EBPs (i.e., the core predicted binding sites that best fit the position weight matrices for these factors and which have previously been shown to be strong binding sequences for ATF and C/EBP homodimers; Mann et al., 2013) as well as a composite DNA element found at the overlapping footprints identified in (A) containing an ATF and a C/EBP half site (left). Average DNase I cut counts in the vicinity of the predicted binding sites for ATF and C/EBP as well as the composite ATF-C/EBP DNA element found in footprint regions (middle). Schematic view of transcription factor binding to the different types of predicted sites (right).
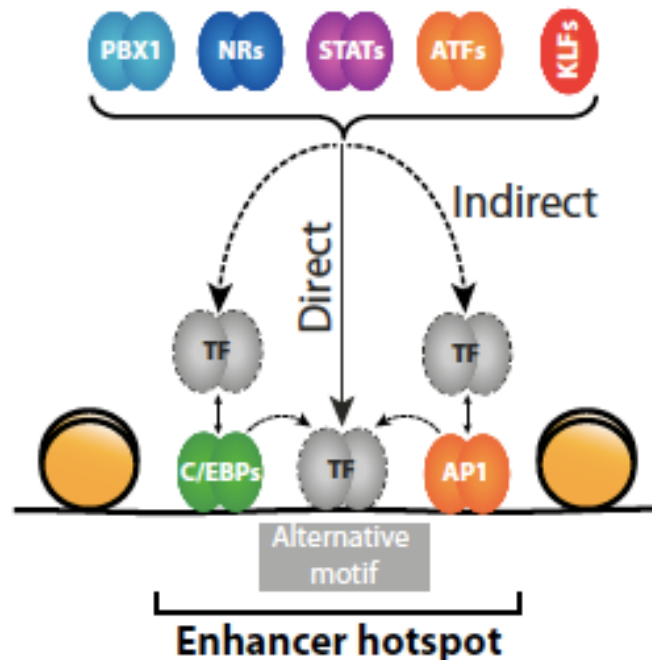
(C) Percentage of ATF only, C/EBP only, and shared ATF-C/EBP ChIP-seq peaks that contain the different types of predicted consensus sites shown in (B). *p < 0.01 as determined by Fisher's exact test.

See also Figure S2.

F

Model for alternative recruitment to early adipogenic hotspots


Enhancer hotspot

Interestingly, our genomics analyses also revealed that alternative binding of transcription factors not involving their regular binding sites is a key mechanism through which transcription factors are recruited to hotspots. This is consistent with a recent study demonstrating that a large fraction of ChIP-seq peaks for several different factors is associated with a binding site for another factor, but not a site for the factor itself (Wang et al., 2012). Whether these alternative binding events occur via indirect binding or assisted loading to nonconsensus motifs remains unclear.

Our footprinting analyses demonstrate that C/EBPs, KLFs, and AP1 factors make many strong footprints that are highly associated with alternative mechanisms of binding of other factors, indicating that these factors may be involved in facilitating recruitment of additional factors to hotspots through mechanisms not involving their known motifs. Interestingly, however, we have observed that different factors have different abilities

Dynamic Rewiring of Promoter-Anchored Chromatin Loops during Adipocyte Differentiation

Graphical Abstract
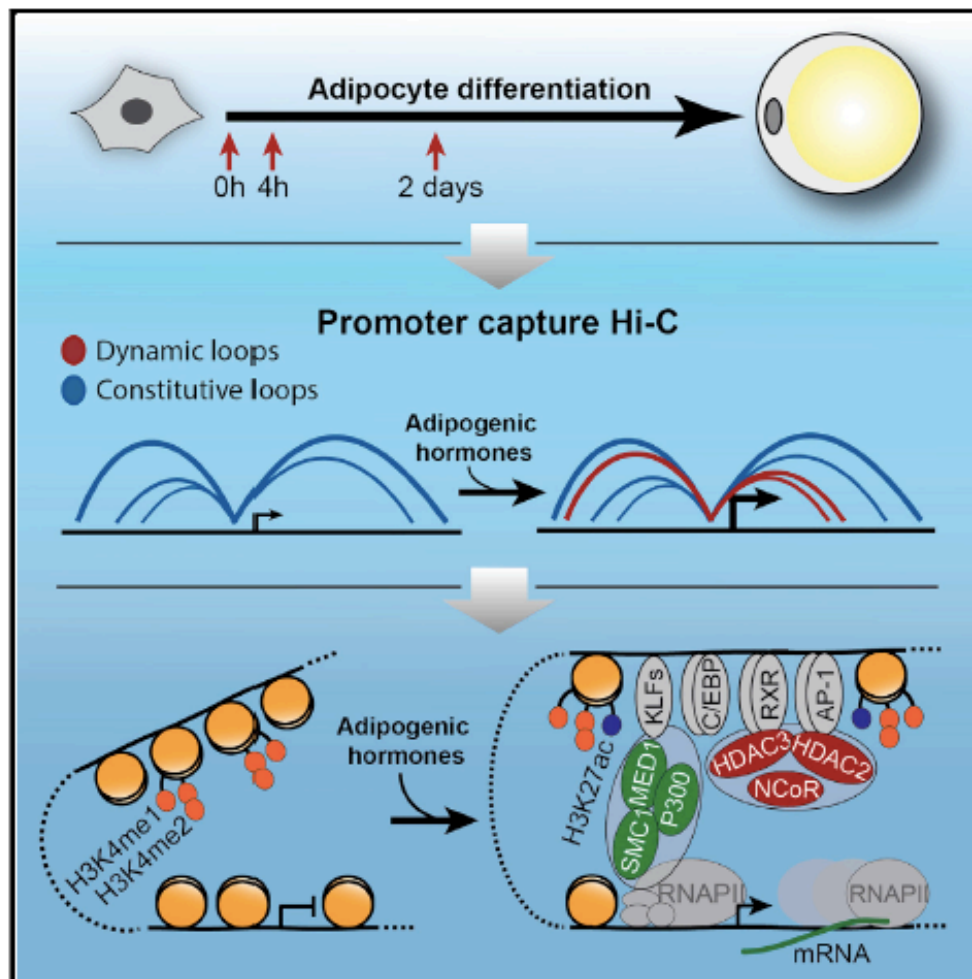
Authors

Rasmus Siersbæk,
Jesper Grud Skat Madsen,
Biola Maria Javierre, ...,
Mikhail Spivakov, Peter Fraser,
Susanne Mandrup

Correspondence

peter.fraser@babraham.ac.uk (P.F.),
s.mandrup@bmb.sdu.dk (S.M.)

In Brief

Siersbæk et al. demonstrate a rapid and dynamic rewiring of promoter-anchored chromatin loops during adipocyte differentiation, which is linked to changes in the activity of promoters and enhancers. Loop formation involves activation of poised enhancers and is associated with extensive recruitment of both coactivators and corepressors.
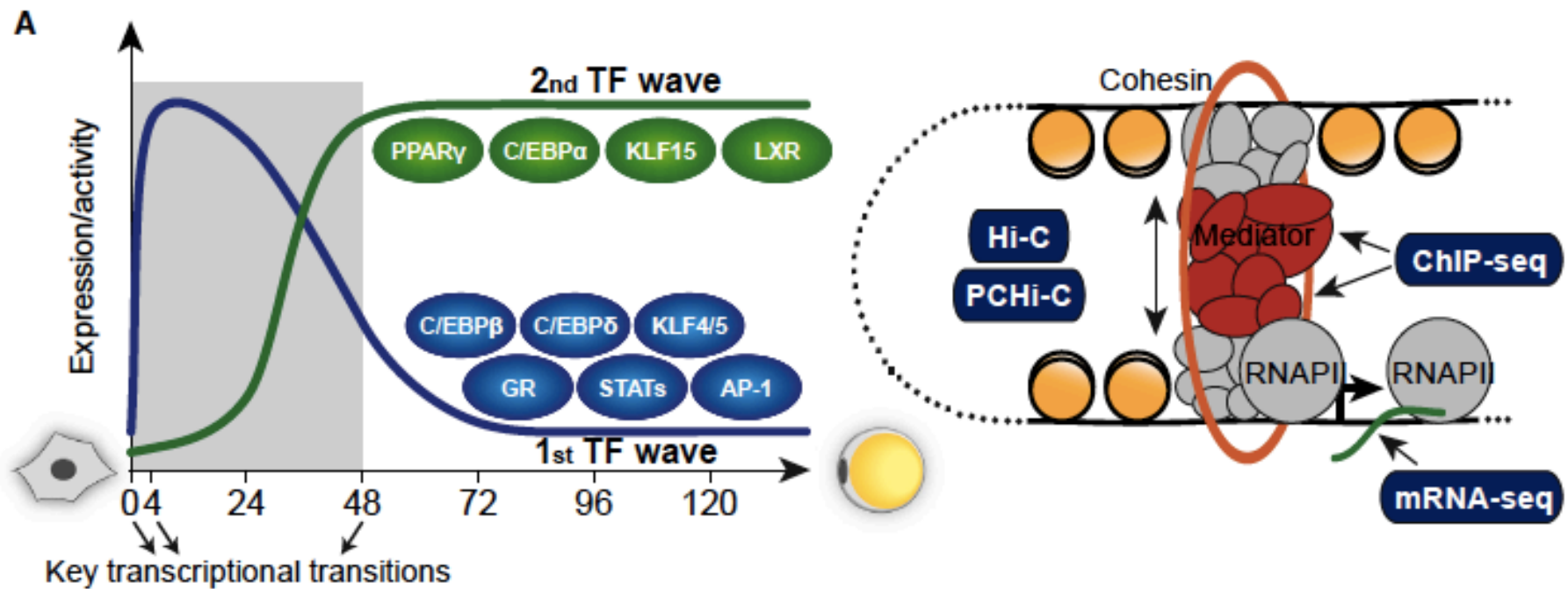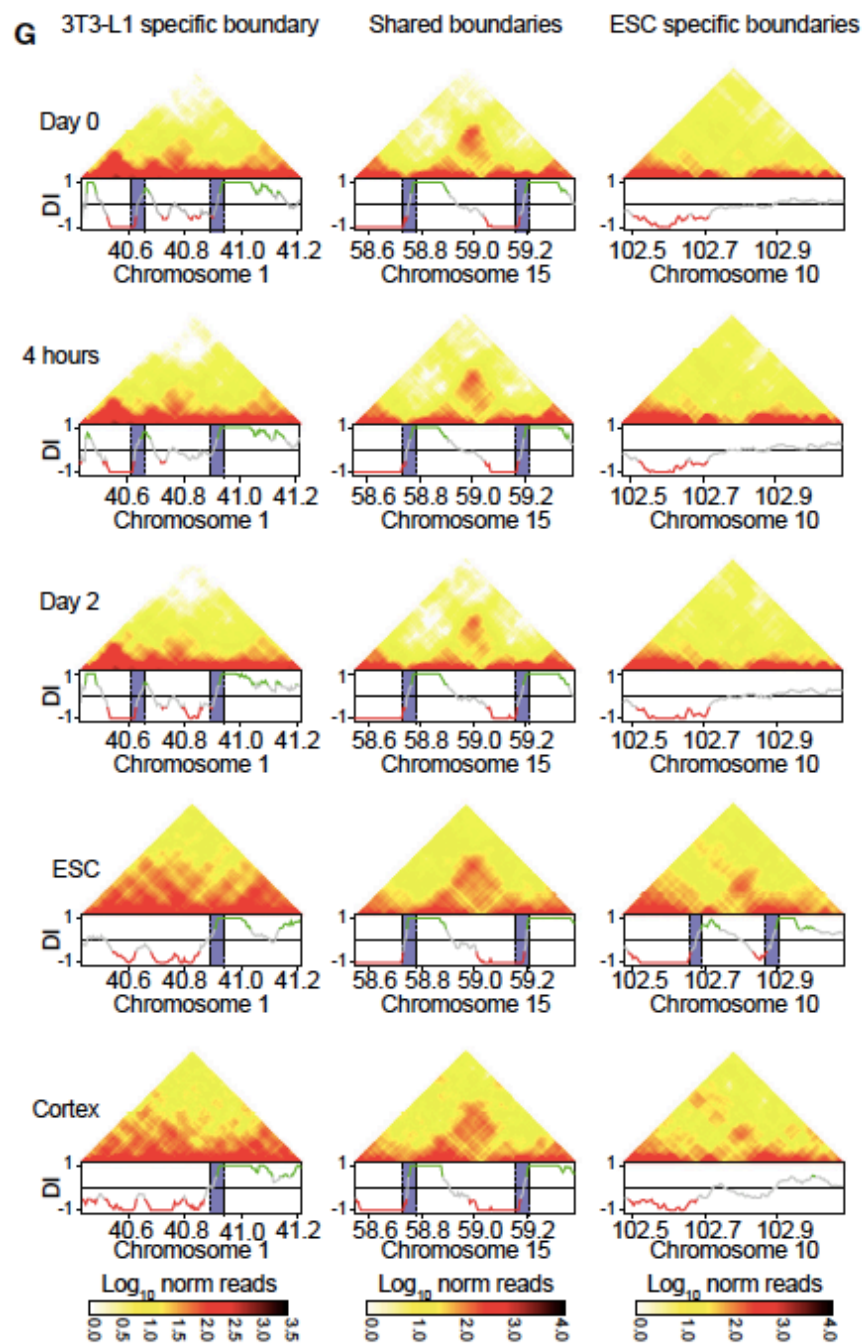
Figure 1. TADs Do Not Change during 3T3-L1 Adipocyte Differentiation

(A) Schematic overview of the experimental approach. Peroxisome proliferator-activated receptor γ, PPARγ; CCAAT/enhancer-binding protein, C/EBP; Krüppel-like factor, KLF; liver X receptor, LXR; glucocorticoid receptor, GR; signal transducer and activator of transcription, STAT; activator protein 1, AP-1.

B

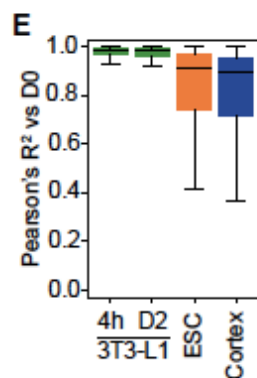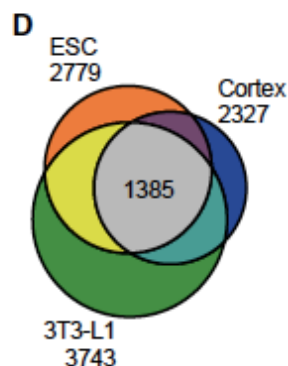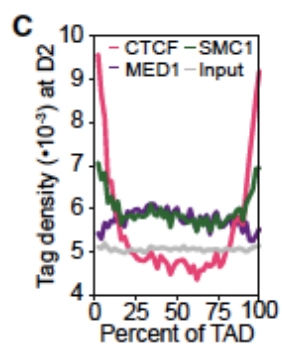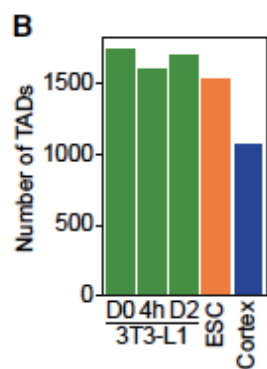Number of TADs

D0 4h D2 — 3T3-L1 | ESC | Cortex

C

Tag density (·10⁻³) at D2

CTCF — SMC1
MED1 — Input

Percent of TAD

D

ESC 2779
Cortex 2327
1385
3T3-L1 3743

E

Pearson's R² vs D0

4h D2 — 3T3-L1 | ESC | Cortex

F

1385

443
708
234
1207
452
265

D0 4h D2 ESC Cortex
3T3-L1

Directionality index

-1.0   0.0   1.0

G

3T3-L1 specific boundary | Shared boundaries | ESC specific boundaries

Day 0

DI

Chromosome 1 | Chromosome 15 | Chromosome 10

4 hours

DI

Chromosome 1 | Chromosome 15 | Chromosome 10

Day 2

DI

Chromosome 1 | Chromosome 15 | Chromosome 10

ESC

DI

Chromosome 1 | Chromosome 15 | Chromosome 10

Cortex

DI

Chromosome 1 | Chromosome 15 | Chromosome 10
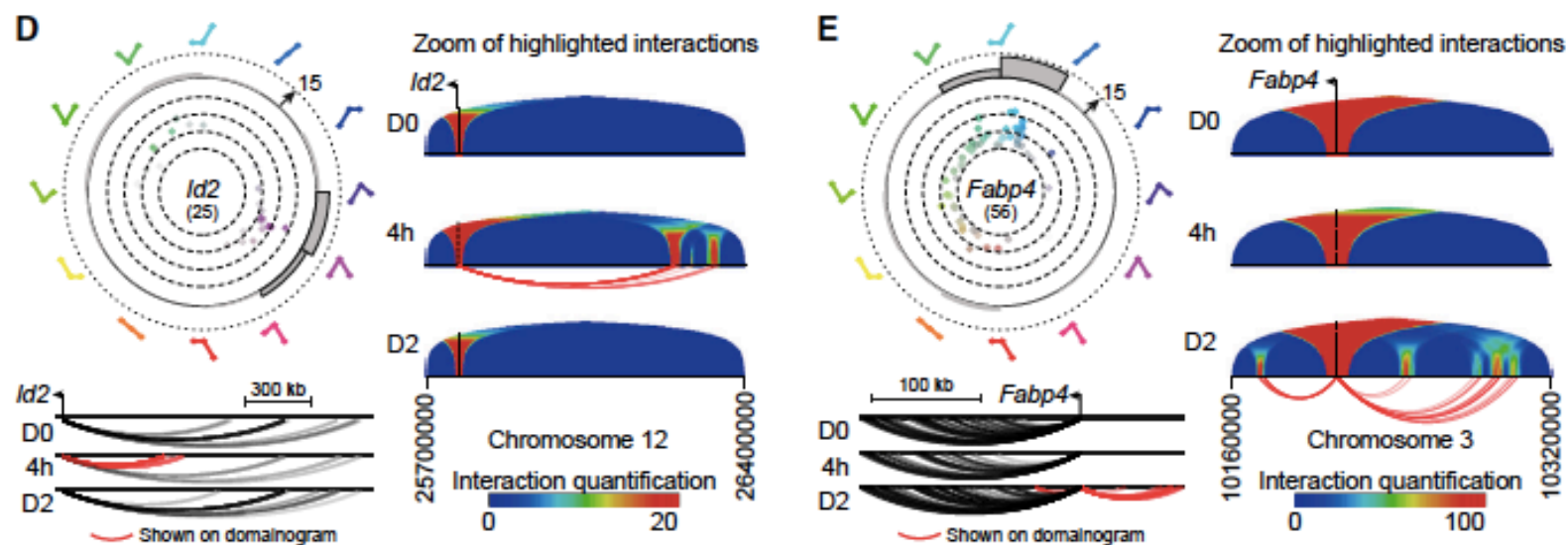
Log₁₀ norm reads

Figure 3. Dynamic Rewiring of Promoter-Anchored Chromatin Loops during 3T3-L1 Adipocyte Differentiation

(A) Fraction of interactions changing interaction strength as determined by tag counts during the course of differentiation.

(D and E) Examples of gene promoters enriched for transient (*Id2*) (D) and late (*Fabp4*) (E) interactions, respectively. Top left, all interactions involving the indicated promoter are visualized in a hue circle. The number of interactions changed at least 1.5-fold in hue bins of 30° is indicated by bar plots. The number by the arrow and the dashed line in each plot indicates the span of the y axis, which describes the number of interactions in each bin. The total number of interactions for each gene is indicated below the gene name. Bottom left, overview of nearby interactions involving the indicated promoter. Right, virtual 4C analyses of nearby interactions involving the indicated promoter. These domainograms visualize the actual strength of the identified interactions based on the tag counts connecting the distal regions and the target promoter. Related to Figure S1.
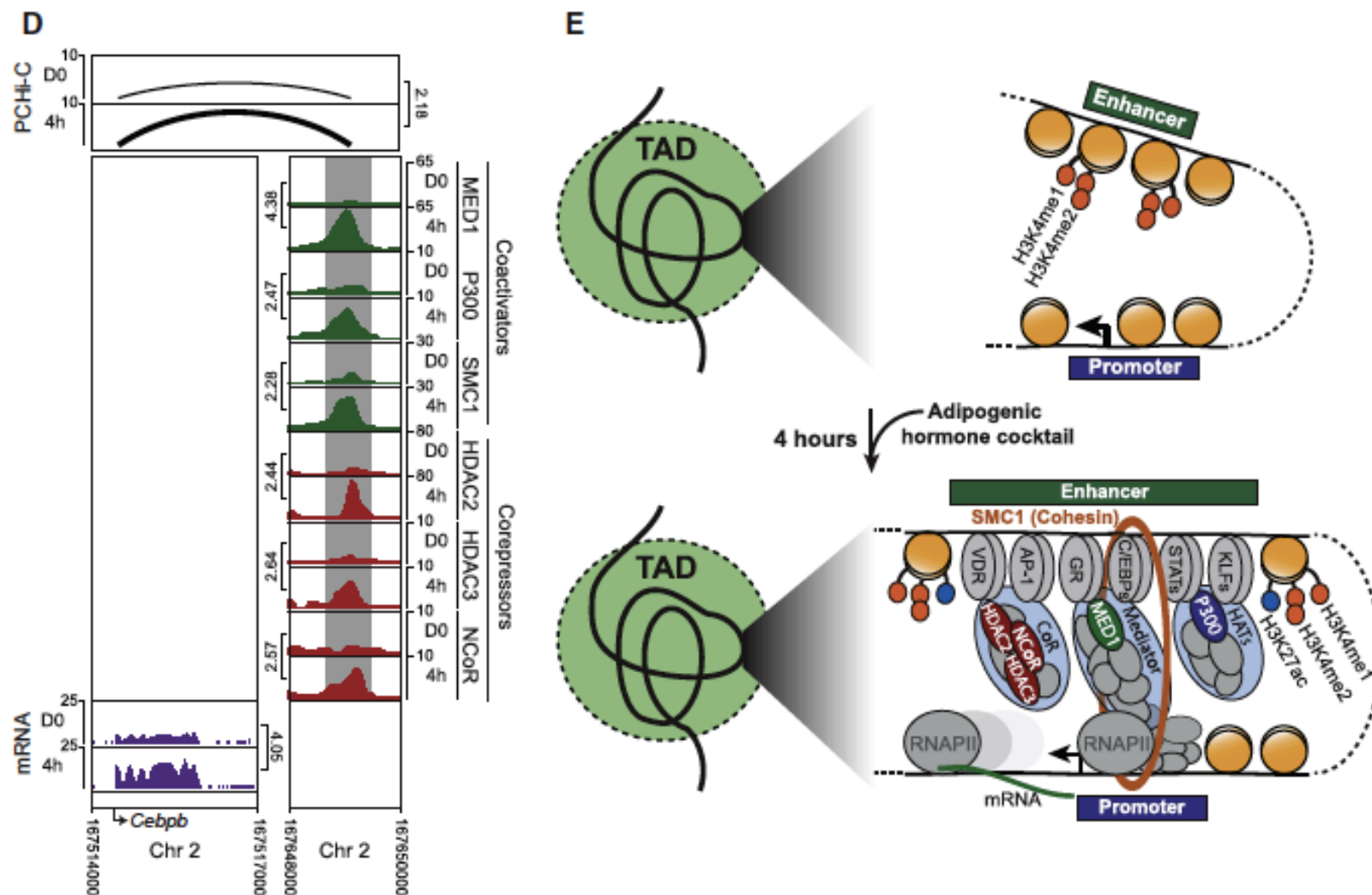
Figure 7. Corepressors Are Recruited to Activated Enhancers that Control Gene Activation through Chromatin Loop Formation

(D) WashU screenshot of *Cebpb* and a connected enhancer. Binding of corepressors and coactivators to the enhancer as determined by ChIP-seq (bottom right), promoter-enhancer chromatin loop formation as determined by PCHi-C (top), and *Cebpb* expression as determined by mRNA-seq (bottom left) in preadipocytes (i.e., prior to induction of differentiation) and 4 hr after stimulation with the adipogenic hormone cocktail are shown. The height of the loop as well as the thickness of the loop line indicates the strength of the promoter-enhancer interaction (top). Fold changes for interaction strength, corepressor/coactivator binding, and expression level are indicated in vertical orientation.

(E) Proposed model of the mechanism controlling enhancer reprogramming and 3D reorganization during adipocyte differentiation. Related to Figures S5–S7.