nature genetics

Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development

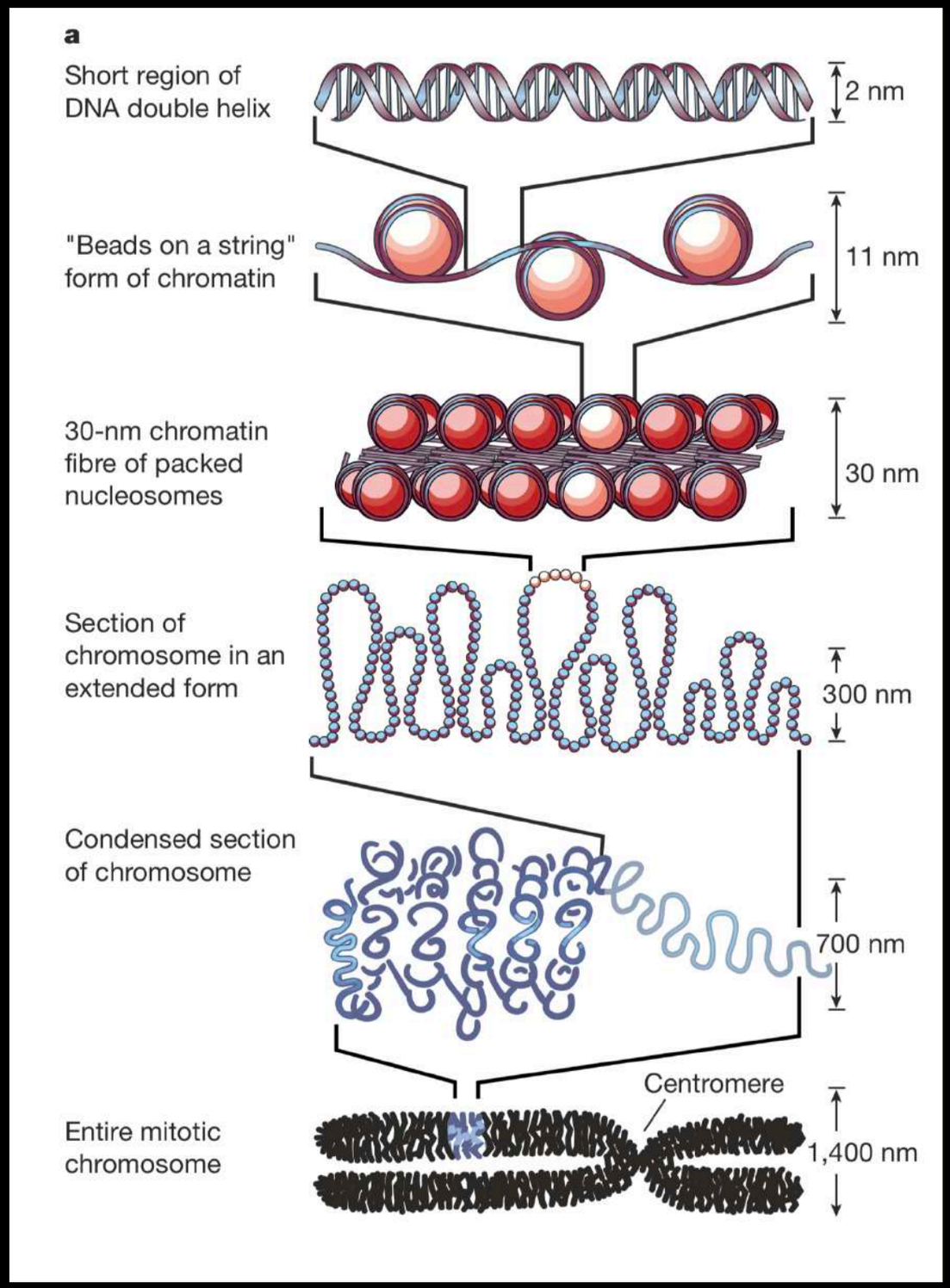
Stefan Bonn[1,2], Robert P Zinzen[1,2], Charles Girardot[1,2], E Hilary Gustafson[1], Alexis Perez-Gonzalez[1], Nicolas Delhomme[1], Yad Ghavi-Helm[1], Bartek Wilczyński[1], Andrew Riddell[1] & Eileen E M Furlong[1]

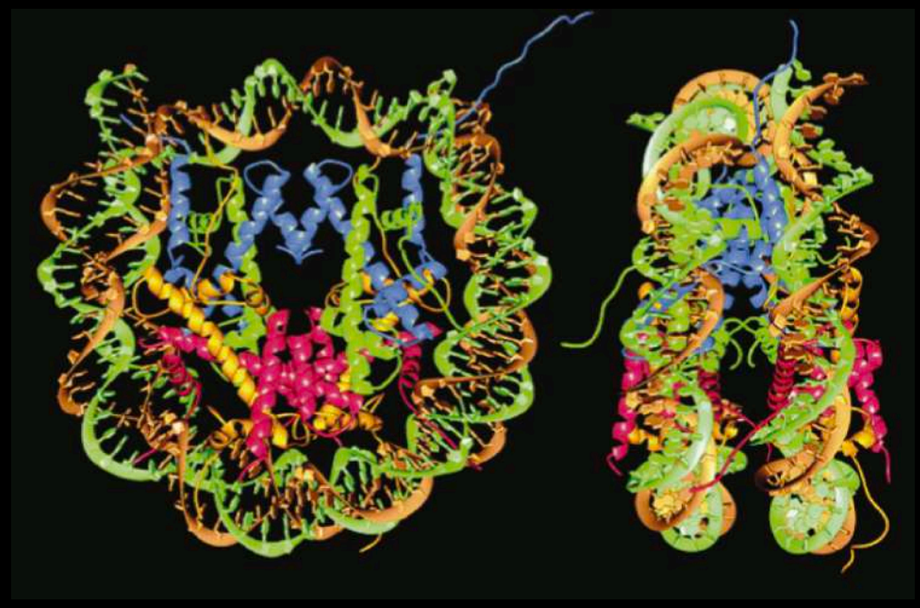
Student Discussion Group
IMSc, Chennai

Chandrani Kumari

28 June 2019

Packaging DNA into Chromatin





The Structure of the nucleosome core particle was uncovered by X-ray diffraction to a resolution 2.8A. It shows the DNA double helix wound around the central histone octamer

Reference: Felsenfeld, Gary, and Mark Groudine. "Controlling the double helix." *Nature* 421.6921 (2003): 448

Histones

- Histones are highly alkaline proteins found in eukaryotic cell nuclei helps in packaging the DNA into structural units called nucleosomes

- They are the chief protein components of chromatin acting as spools around which DNA winds, and play an important role in gene regulation

- Four families of Histones: H2A, H2B, H3 and H4

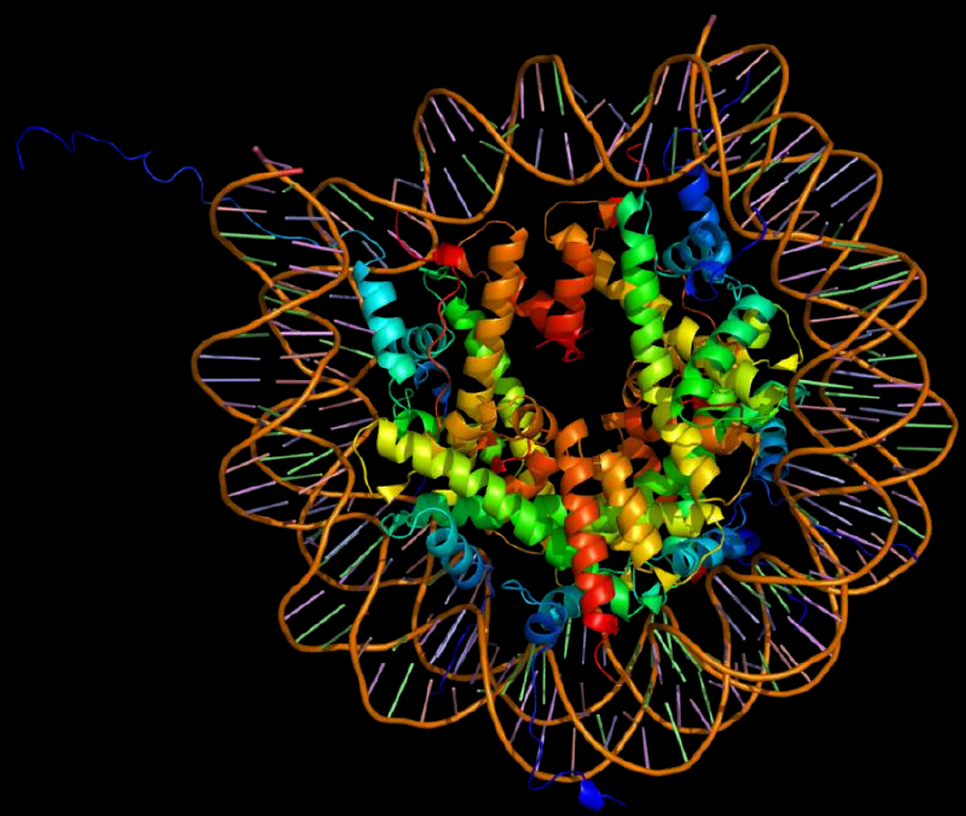


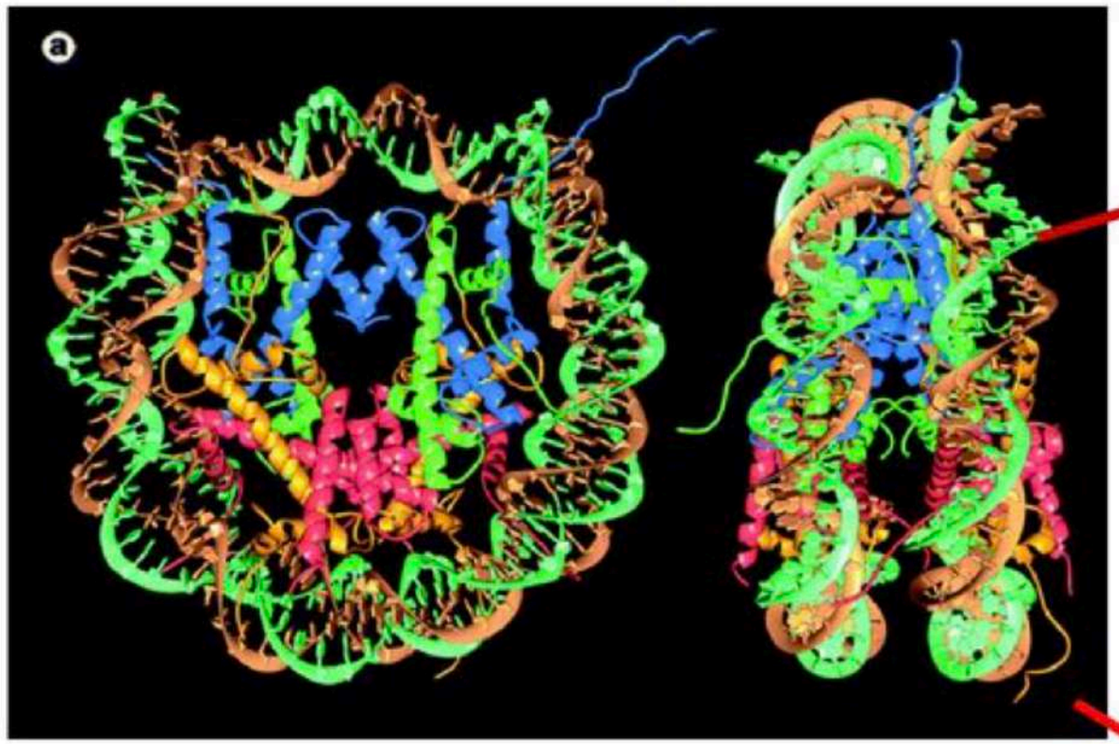
Figure Source: By Emw - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=8814716

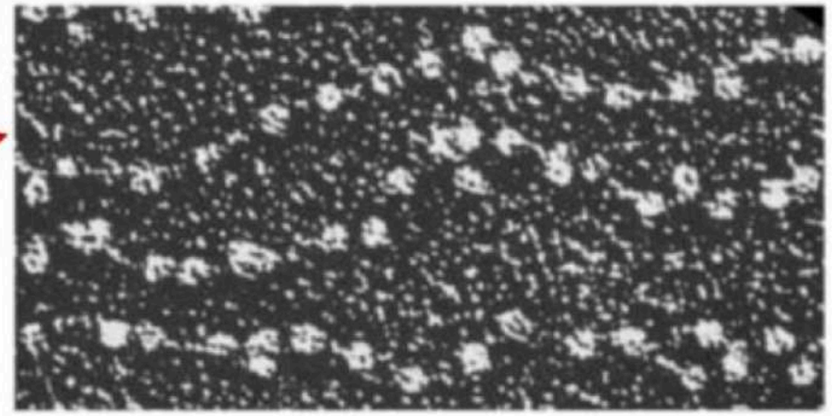
Source : https://en.wikipedia.org/wiki/Histone

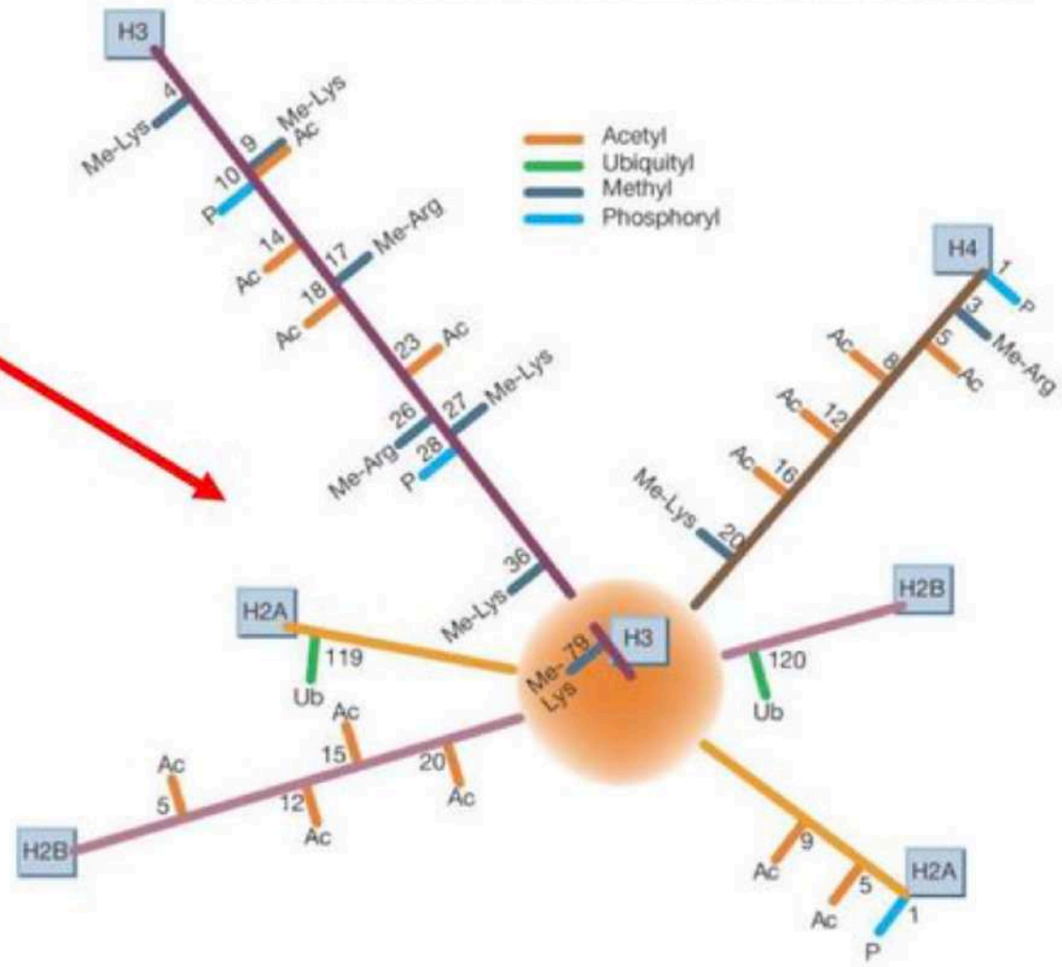
Chromatin regulates gene expression

- Three general ways in which chromatin structure can be altered:
  - Nucleosome remodelling induced by some complex, requires energy
  - Covalent modification of histones within nucleosome
  - Histone variants may replace one or more of the core histones
- Packaging of particular gene in chromatin is required for their expression
- Thus chromatin can be involved in both activation and repression of gene expression

Nucleosome and histone modification







First layer chromatin structure looks like “beads-on-a-string”.

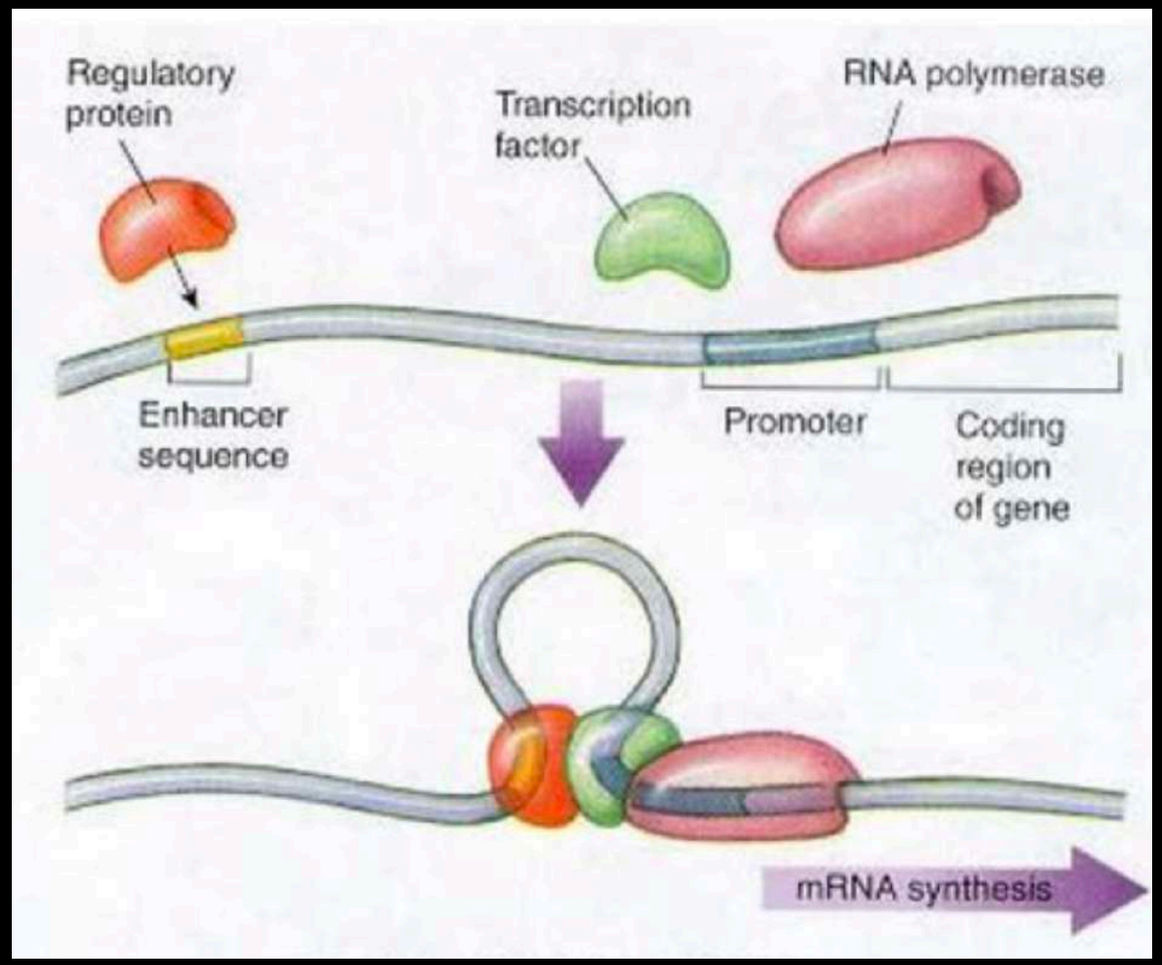
A nucleosome is made of core histone proteins.

The amino acids on the N-terminus of histones can be covalently modified.

Felsenfeld and Groudine 2003

Source : Slide clipped from https://slideplayer.com/slide/13187873/

How enhancers work ?

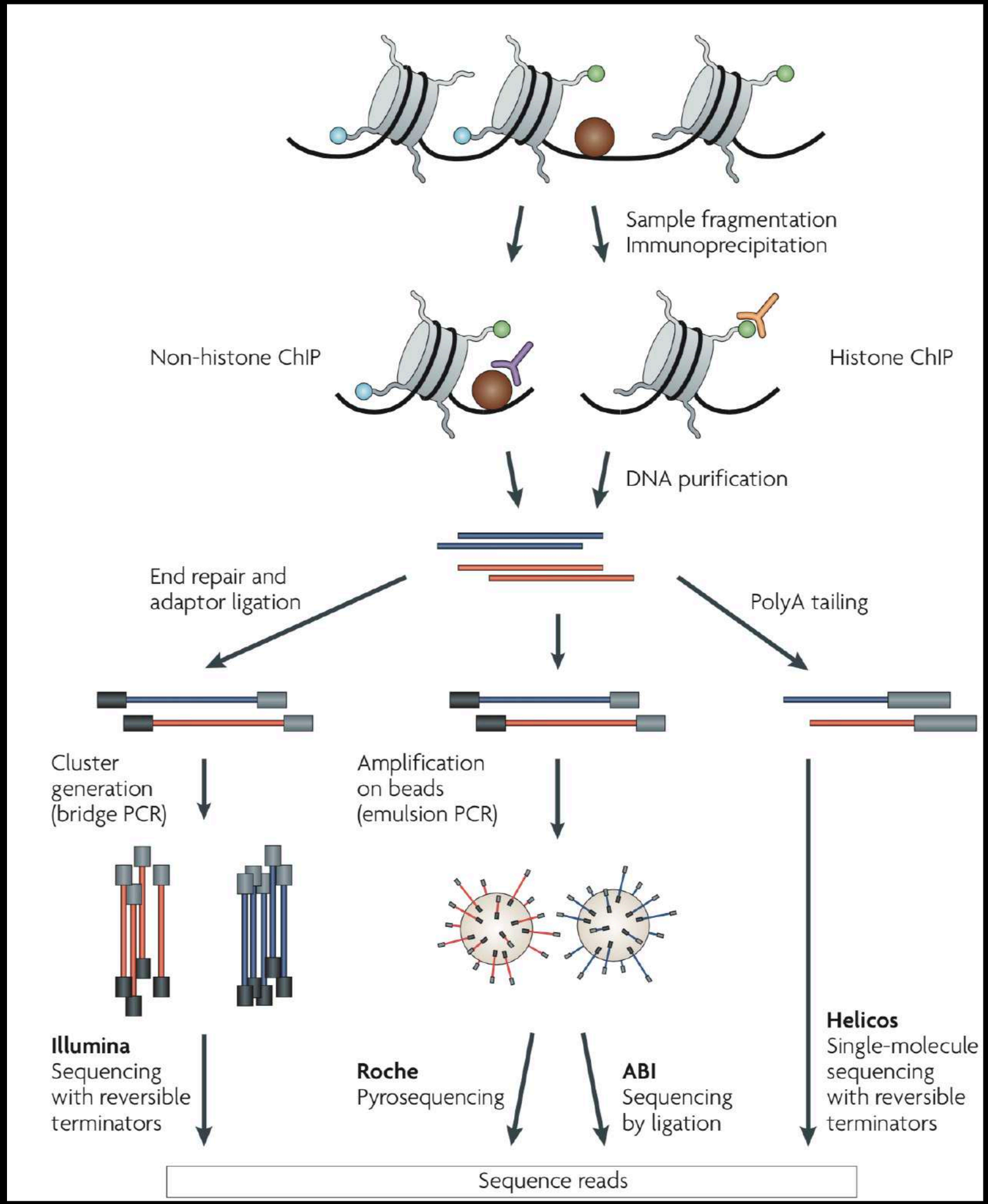


- Enhancer is a DNA sequence that promotes or enhance transcription

- The enhancer site is located far away from the gene being regulated. Binding of a regulatory protein (red) to the enhancer allows the protein to interact with the transcription factors (green) associated with RNA polymerase, activating transcription

https://slideplayer.com/slide/13187873/

ChIP-seq analysis



Reference: Park, Peter J. "ChIP–seq: advantages and challenges of a maturing technology." *Nature reviews genetics* 10.10 (2009): 669.

Introduction

- Chromatin modifications are associated with many aspects of gene expression:
    - H3K4me3, H3K79me3, H3K36me3 reflect promoter activity, gene body transcription, exon intron usage and are highly correlated with gene expression levels
    - H3K4me1 and H3K27ac helps to determine the cis regulatory elements
- Linking chromatin modification to the activity of enhancer remains a key challenge
- Studies in embryonic stem (ES) cells found a positive correlation between the presence of H3K27ac on putative enhancers and the activity of the closest proximal gene, but opposing results were reported for the presence of H3K27me3 on regulatory elements

- Cell fate transitions during embryonic development occur on the order of hours, yet it is not known how this relates to dynamic changes in chromatin state

- It is currently not clear how accurately changes in chromatin modification reflect the precise timing of enhancer, promoter or gene activity

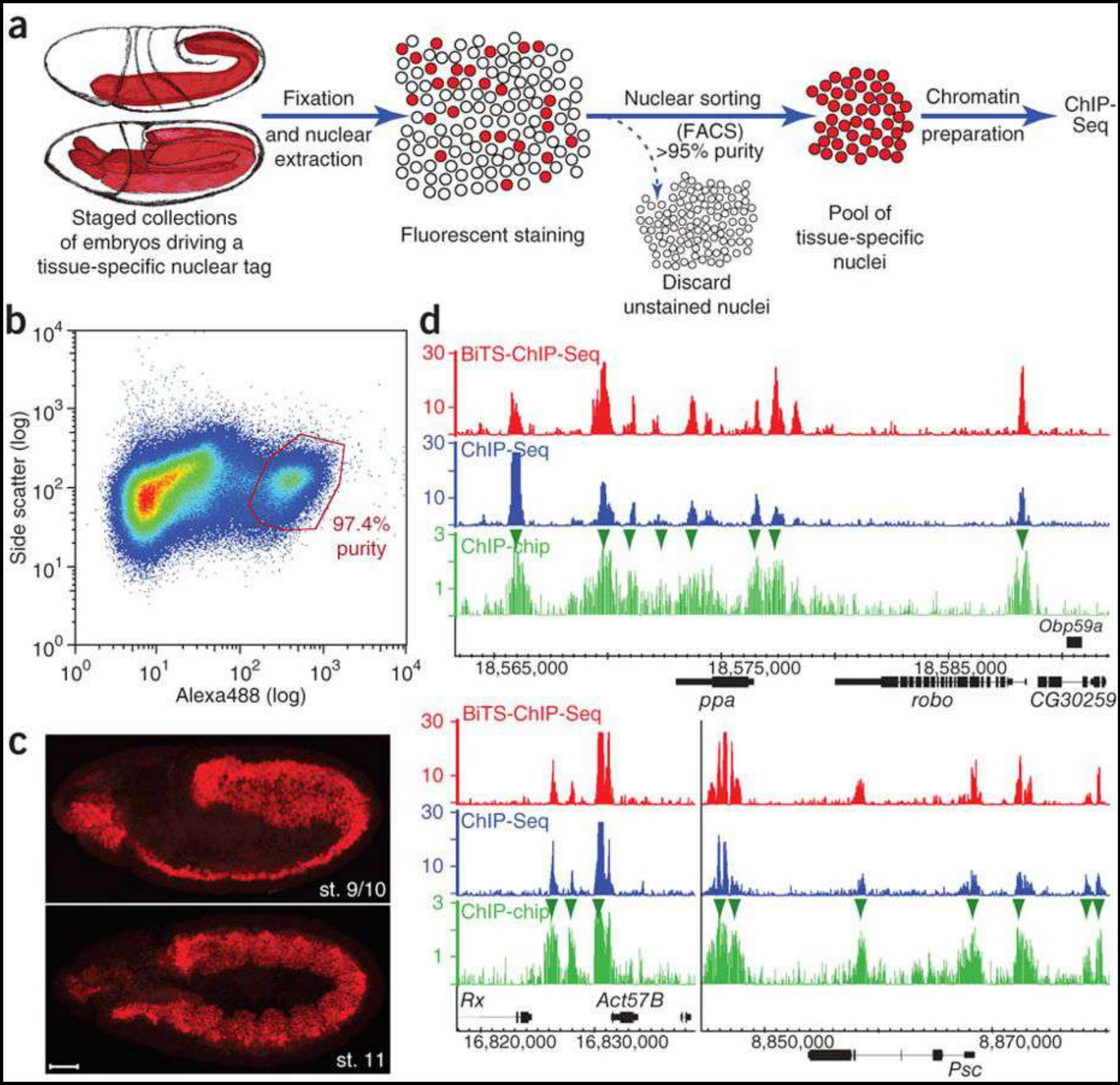
- Available chromatin data comes from dissected tissues or whole embryos, yielding mixed signals from heterogeneous cell types

Cell type-specific ChIP in developing embryos

- Batch isolate tissue-specific chromatin for immunoprecipitation (BiTS-ChIP), which uses a transgene to express a tagged nuclear protein specifically in the cell type of interest

- BiTS-ChIP was applied to six chromatin marks and RNA polymerase II (Pol II) to examine occupancy in mesodermal cells during *Drosophila* development, for which extensive transcription factor occupancy data are available

- Six chromatin marks: promoters (H3K4me3 and H3K27ac), gene bodies (H3K79me3 and H3K36me3), *cis*-regulatory elements (H3K4me1 and H3K27ac) and repressed regions (H3K27me3), Pol II occupancy and histone H3 density



Cell type-specific ChIP in developing embryos

Stages of *Drosophila* development: https://www.sdbonline.org/sites/fly/atlas/00contents.htm

BiTS high sensitivity and specificity

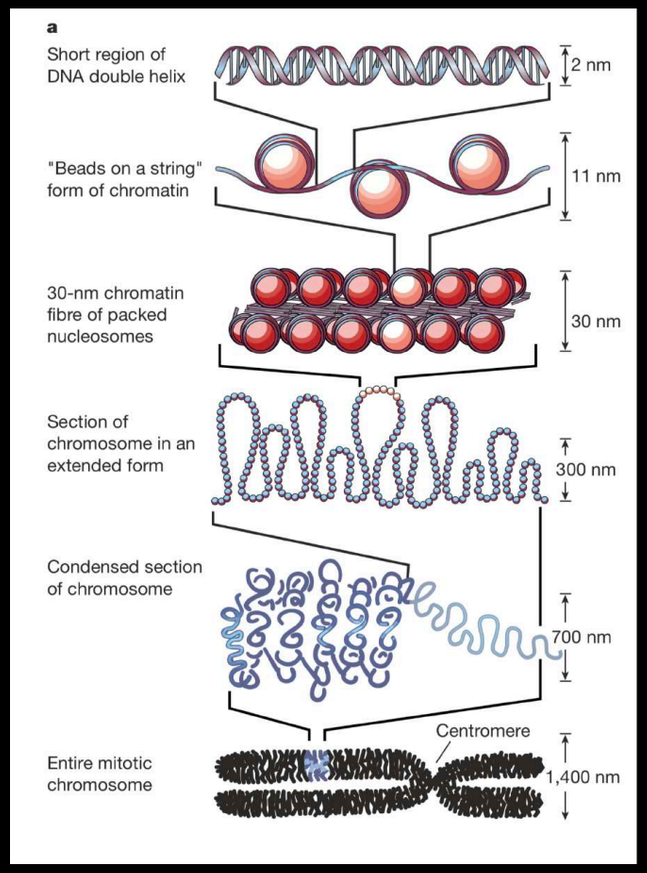
- The dissociation of cells from tissues and embryos leads to a transcriptional stress response, which is typically observed with FACS sorting of live cells

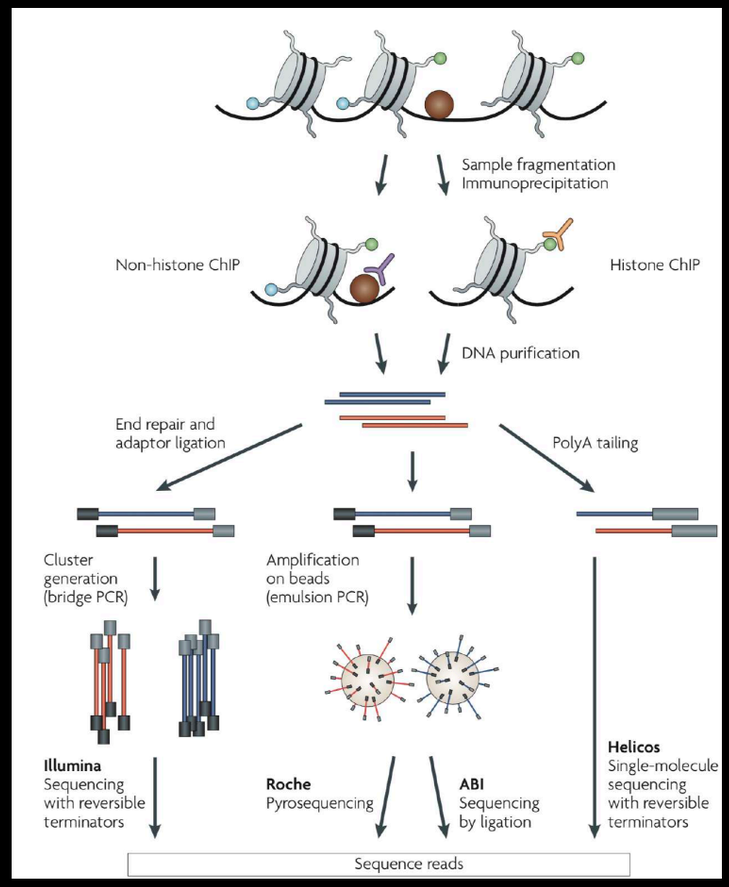
- Covalent cross-linking before embryo dissociation avoids this problem by blocking all transcriptional activity

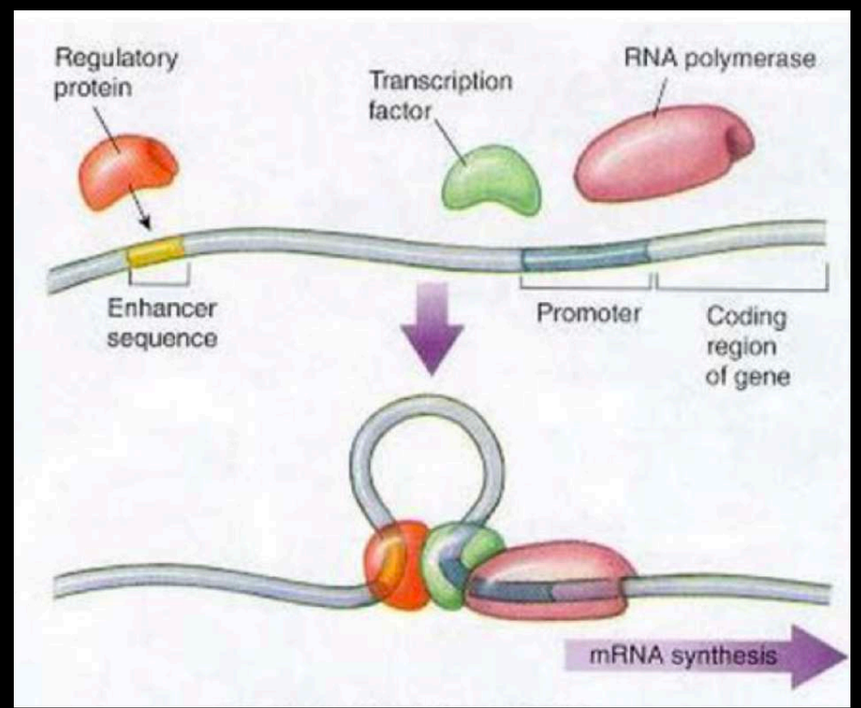
- This key feature of the BiTS ChIP protocol preserves the transcriptional context during nuclear sorting and facilitates cell type–specific analysis of transcription factor binding, which is not possible with native ChIP

- A second important feature of BiTS-ChIP is the high specificity of the data it generates
- Genes that are known to be expressed exclusively in mesoderm at 6–8 h of development showed high enrichment for H3K4me3 and H3K27ac at their promoters and H3K79me3 on their gene bodies

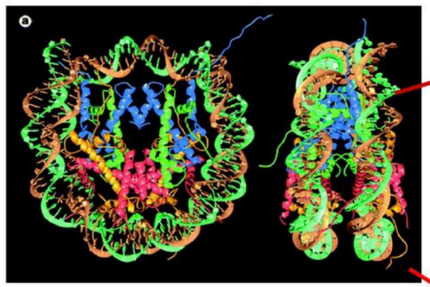
Summary from last Talk







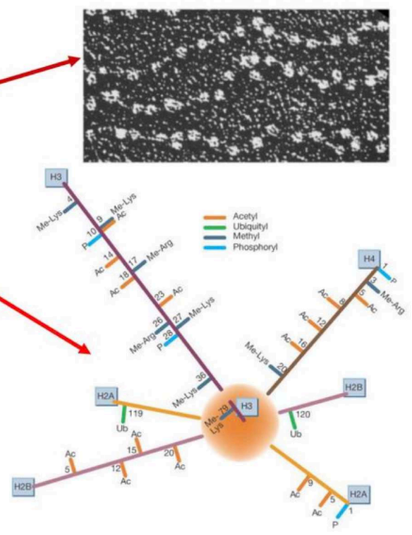
Nucleosome and histone modification



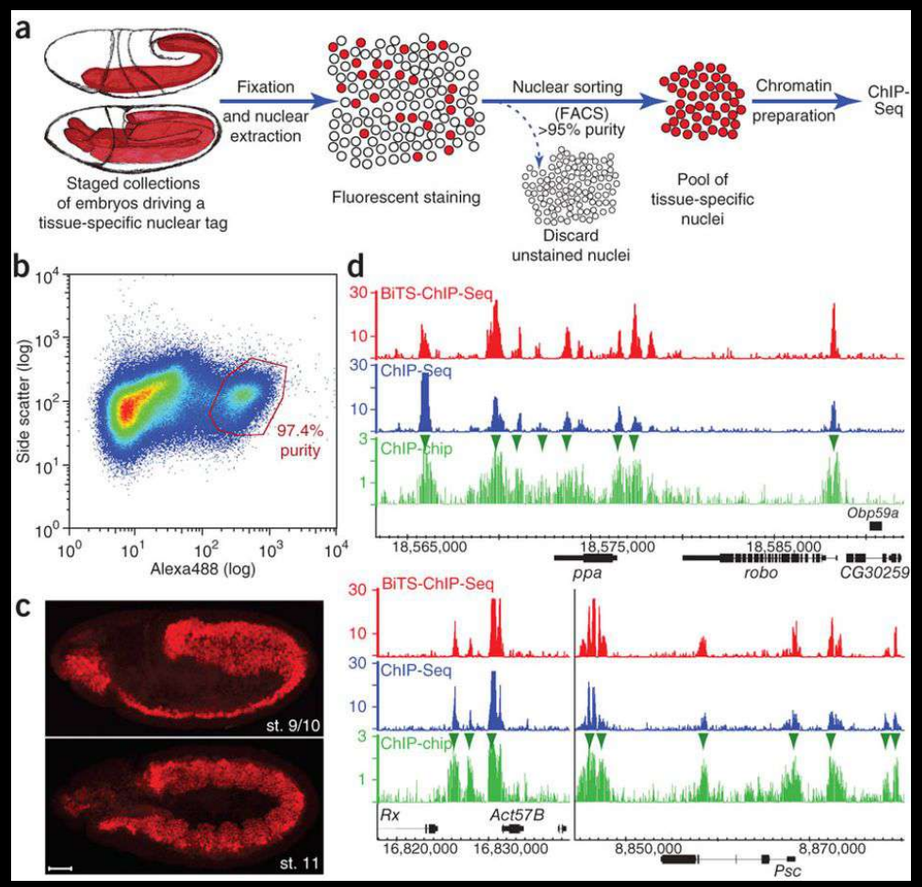
First layer chromatin structure looks like “beads-on-a-string”.

A nucleosome is made of core histone proteins.

The amino acids on the N-terminus of histones can be covalently modified.



Felsenfeld and Groudine 2003



Next Presentation

- A new role for H3K79me3 on developmental enhancer
- Diverse chromatin marks and Pol II indicate active enhancers
- Chromatin modifications and the timing of enhancer activity
- Chromatin Signature can predict enhancer regulatory state
- Predicted regions function as mesodermal enhancer *in vivo*

New role for H3K79me3 on developmental enhancer

- Six Chromatin marks, Pol II occupancy and histone H3 density

- Promoters (H3K4me3 and H3K27ac),
- Gene bodies (H3K79me3 and H3K36me3),
- *cis*-regulatory elements (H3K4me1 and H3K27ac)
- Repressed regions (H3K27me3)

New role for H3K79me3 on developmental enhancer

- To determine the relationship between chromatin modifications and enhancer activity, 465 characterised *Drosophila* enhancer were assembled from publicly available data (CRM activity Database 2 (CAD2))

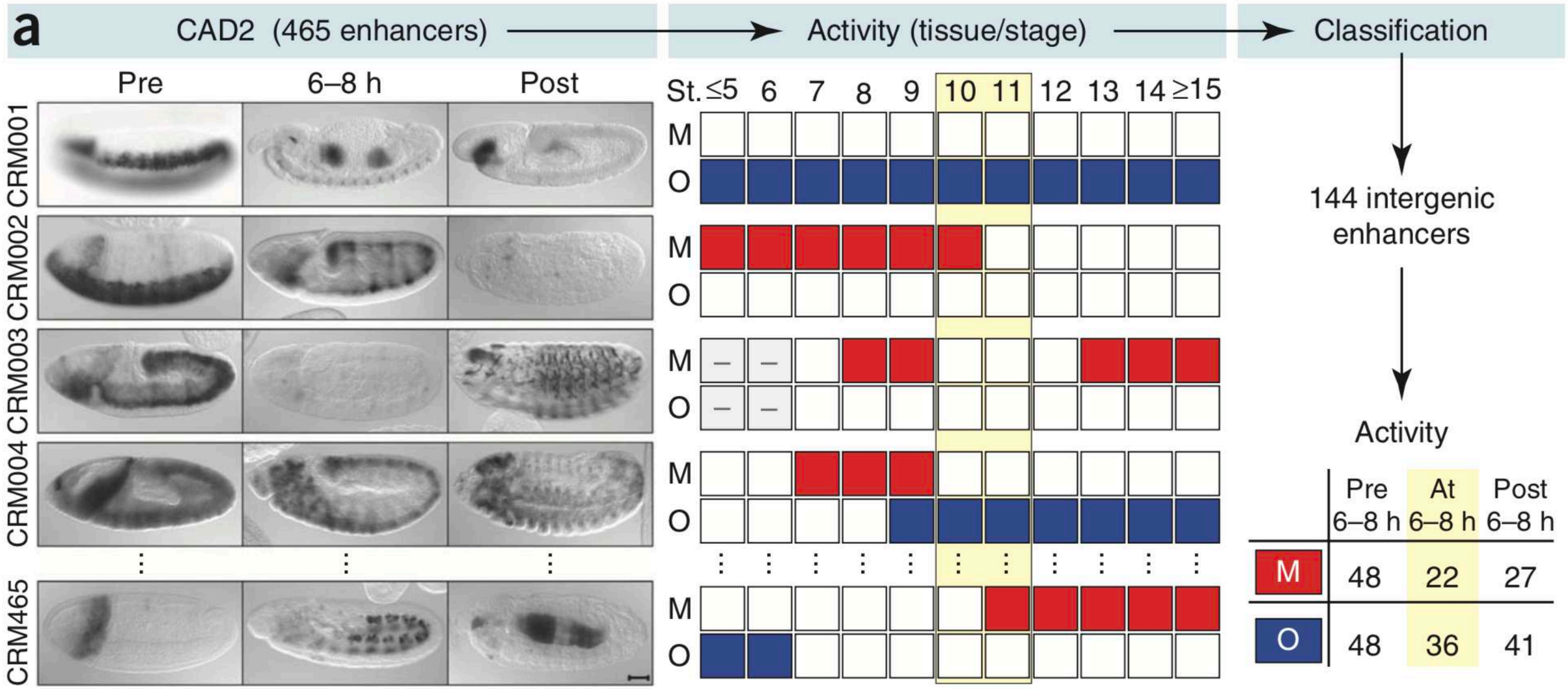
- Literature annotated enhancer activity mapped to the *Drosophila tissue*

- Enhancers were broadly grouped into two categories at each developmental stage:

    - Mesodermal activity

    - Non-mesodermal activity

- Only 144 intergenic enhancers were used for the subsequent analysis



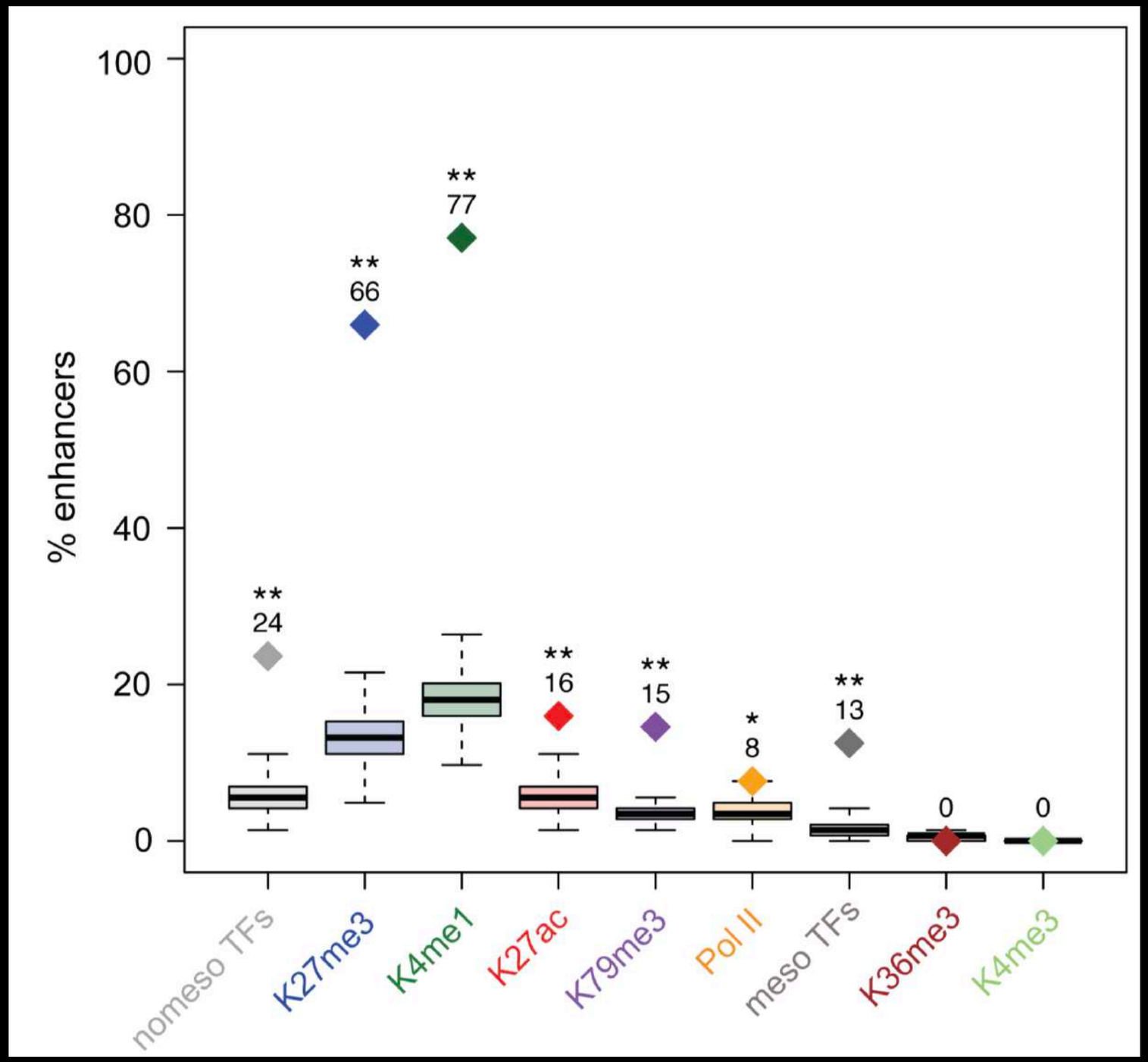
CAD2 enhancer activity annotation and filtering

Left, reporter gene expression directed by CRMs in transgenic embryos.

Middle, reported activity was evaluated by stage for activity in mesoderm (M, red), activity elsewhere (other: O, blue) or no activity (white). −, no information (grey).

Right, 144 of 465 enhancers are located >1 kb away from genes and do not overlap with H3K4me3 peaks. Activity data (bottom right) tabulated non exclusively for tissue and stage. Yellow shading indicates the investigated developmental stages

- General distribution of chromatin marks on developmental enhancers without considering their activity status:



- Solid diamonds show the observed percentage of developmental enhancers in the CAD2 database containing indicated chromatin modifications, Pol II or TF occupancy, regardless of their activity state (the number is indicated above the diamond)

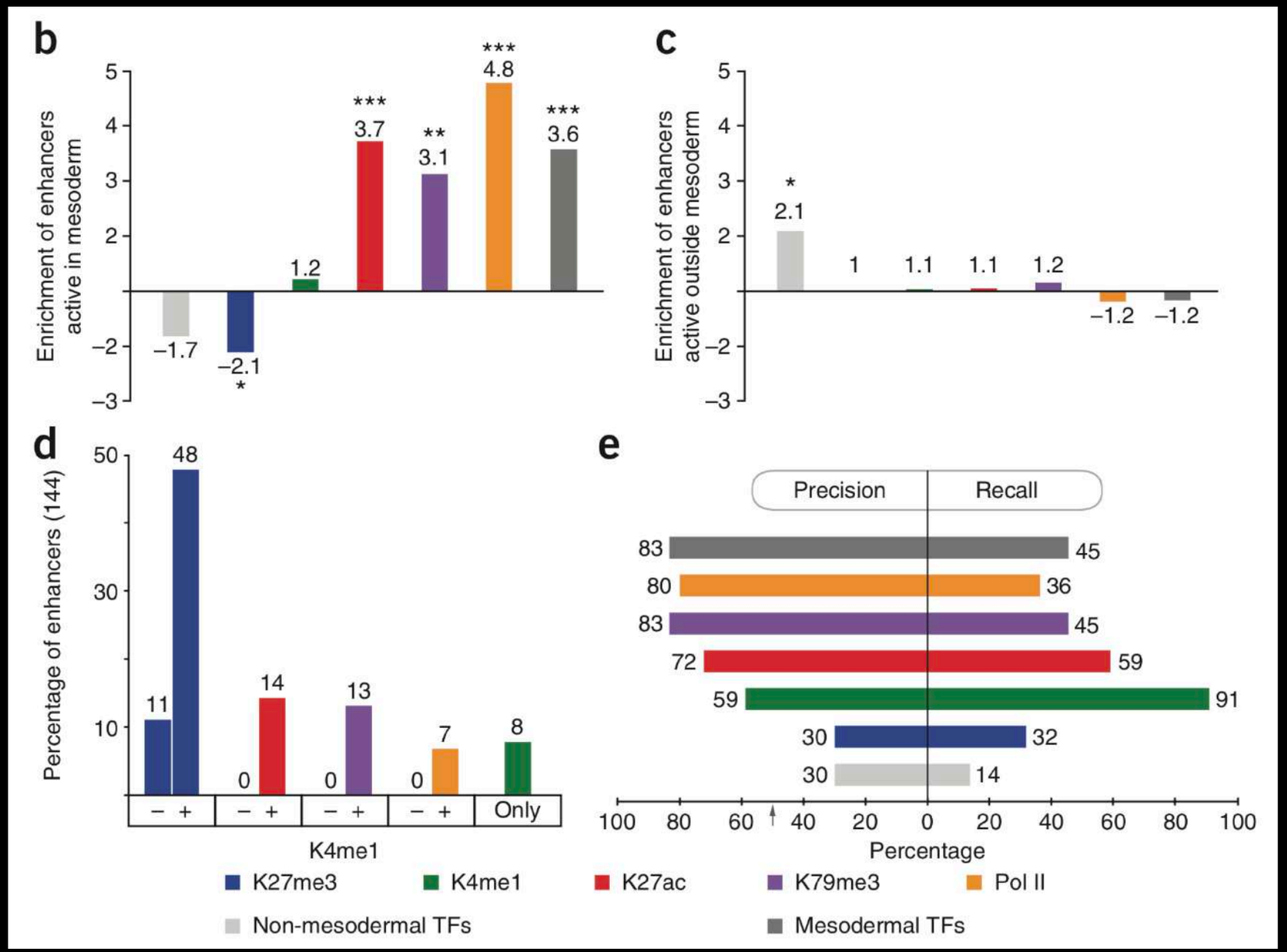
- The box plots show the distribution of percentages obtained using 999 random sets of 144 size matched intergenic regions

- Significant differences of the observed values (diamonds) from the random distribution (box plots) were estimated by bootstrapping

- Two chromatin marks examined are associated with PolI II elongation and active gene transcription (H3K36me3 and H3K79me3), only K79me3 is found on enhancers

- Which indicate potentially new role for K79me3 chromatin mark

Diverse chromatin marks and Pol II indicate active enhancers



Correlating chromatin marks and Pol II occupancy with enhancer activity

- H3K4me1 provides no information on their activity status although its presence has been seen on regulatory elements, as previously reported

- H3K27me3 was significantly depleted on active mesodermal enhancers

- Enhancers with H3K27ac and H3K79me3 marks and Pol II occupancy were significantly enriched for mesodermal activity

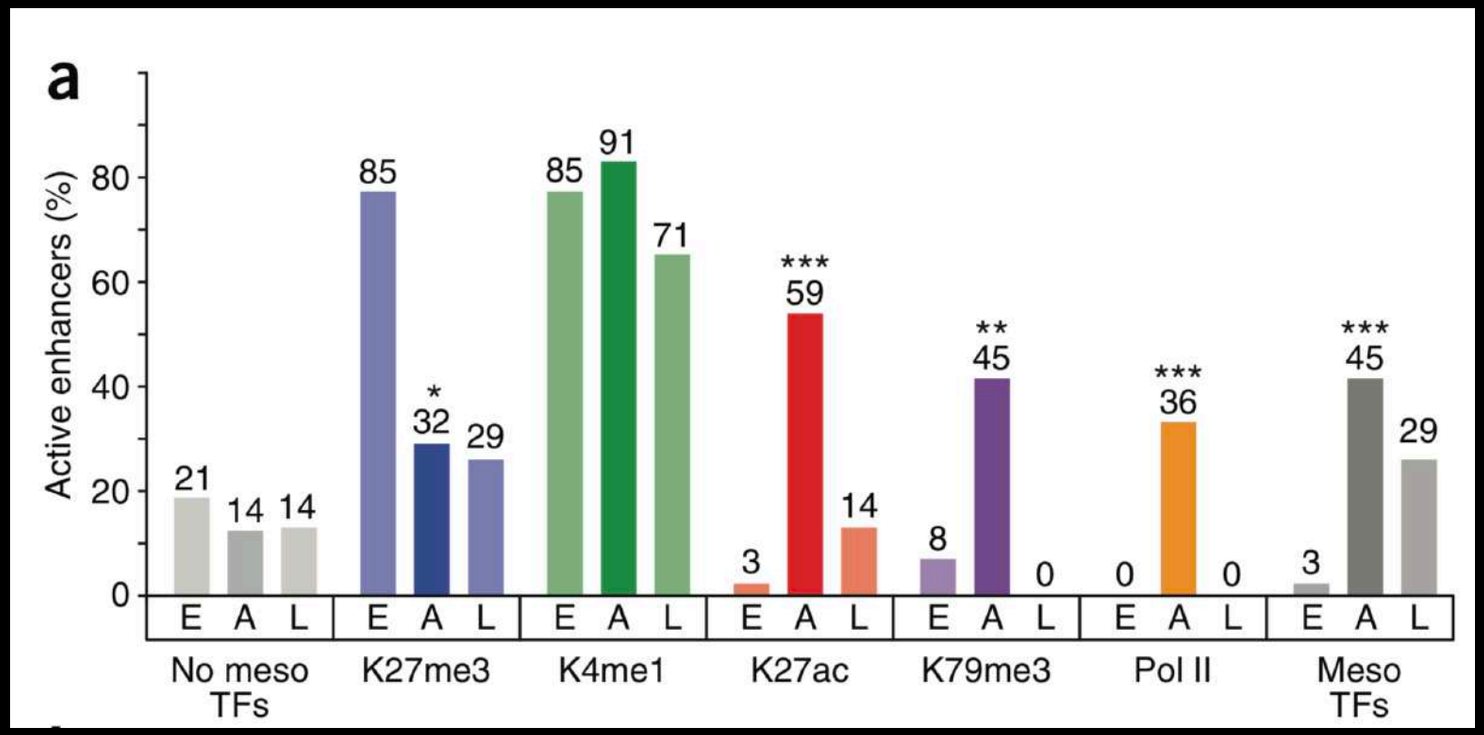
- H3K27ac and H3K79me3 marks and Pol II binding distinguish active and inactive enhancers with high precision

- Enhancers with H3K27ac, H3K79me3 or Pol II are co- marked by H3K4me1

- Results show that not only one specific chromatin mark associated with active enhancers, such as H3K27ac, but instead active regulatory regions are enriched for multiple chromatin modifications and Pol II occupancy

Chromatin modifications and the timing of enhancer activity

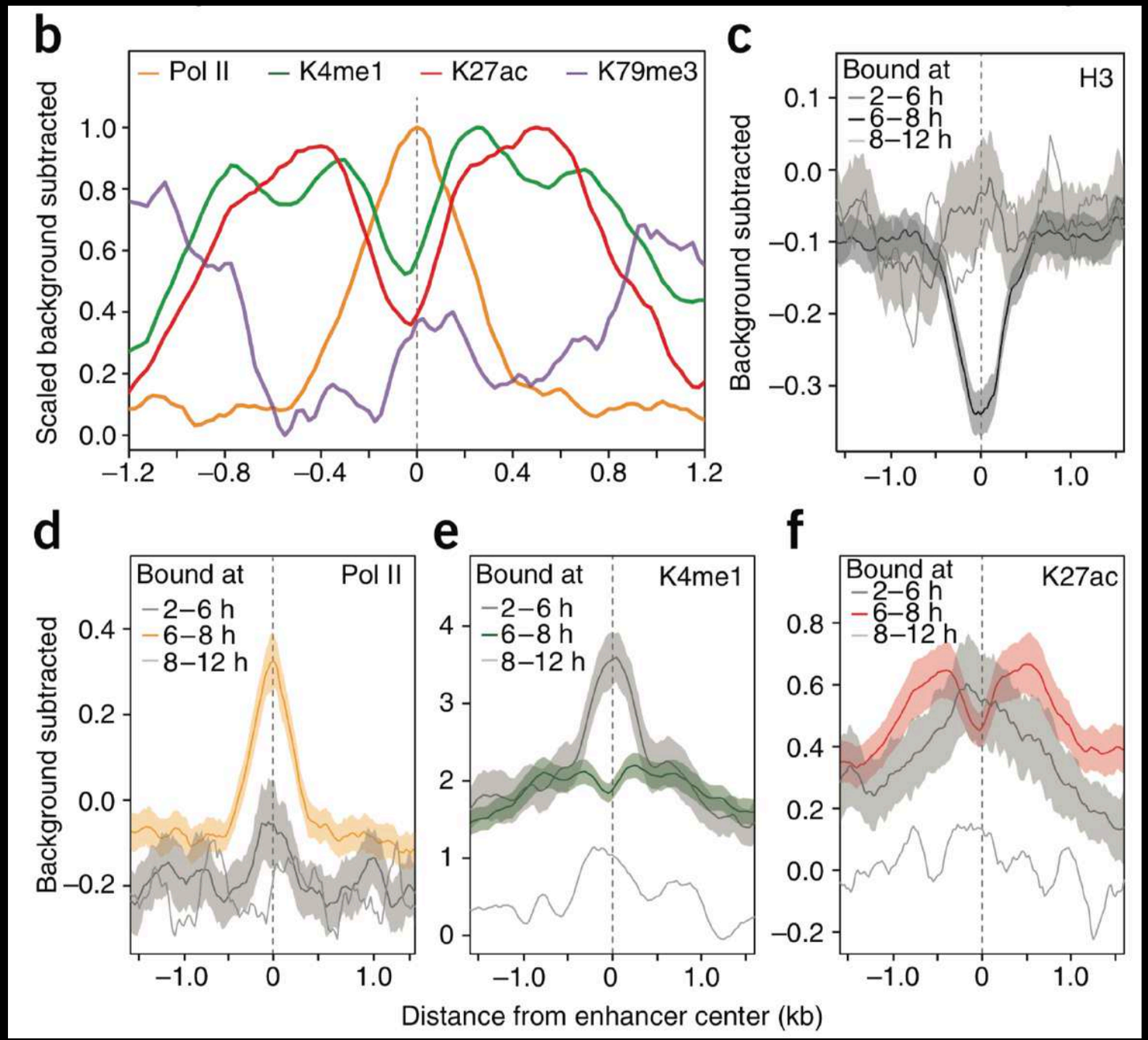
- To assessed the relationship between temporal changes in enhancer activity and chromatin modification within a 2-h window (6–8 h), enhancers were divided into three temporal classes:
    - Mesodermally active during the 2-h time period (at 6–8 h)
    - Only active earlier (<6 h)
    - Only at later stages of development (>8 h)



Analysis of three temporal classes of mesodermal enhancers

- Presence of H3K27ac and H3K79me3 marks and Pol II on enhancers at 6–8 h was highly correlated with the precise timing of enhancer activity

- Pol II occupancy was transient, absent from the enhancers in earlier stages of development and from those that had just become inactive in mesodermal cells



- Spatial distribution of Pol II, H3K4me1, H3K27ac and H3K79me3 on enhancers with signal normalised to [0,1], Pol II signal is centered, and chromatin modifications show bimodal distributions around Pol II

- Pol II signal peaks at the time of transcription factor binding but not when transcription factors are no longer bound

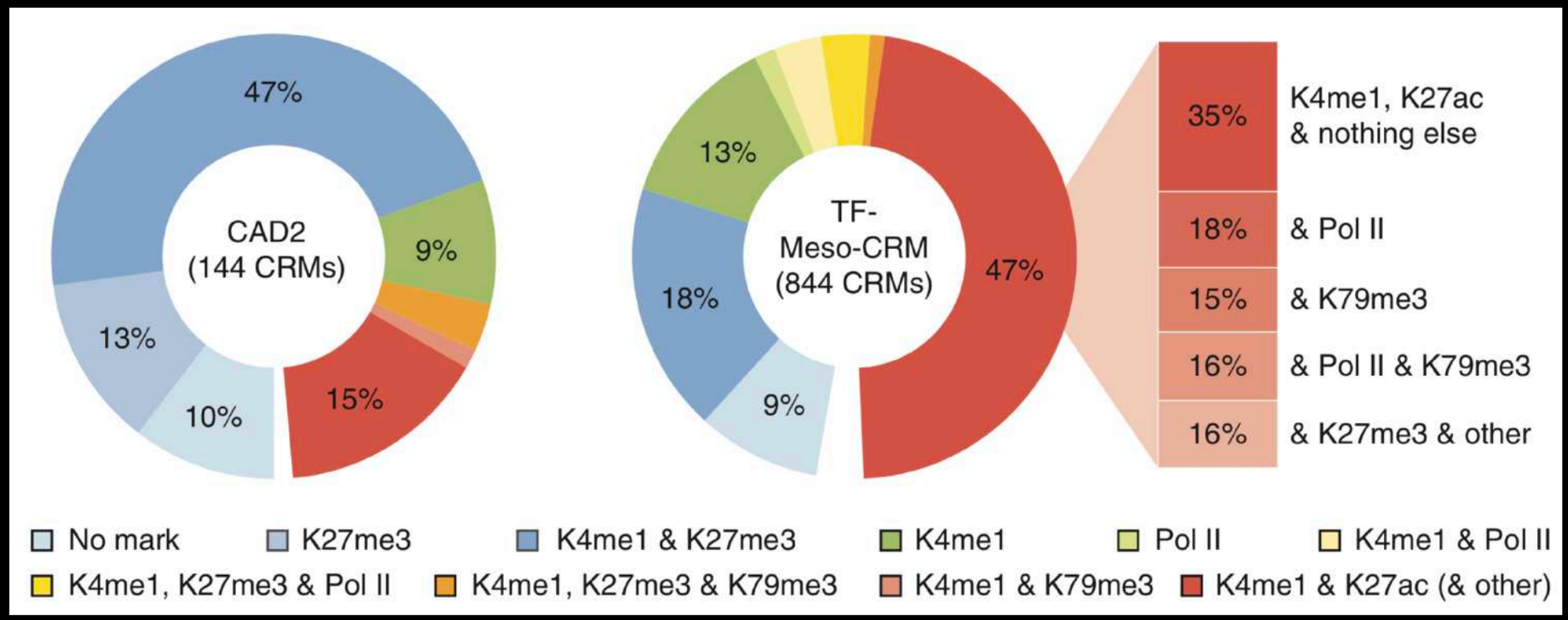
- H3K4me1 and H3K27ac signals exhibit bimodal distributions at the time of transcription factor binding but peak centrally thereafter

Distribution of Pol II and chromatin mark quantitative signals across TF-Meso-CRMs. *x* axes show distance from CRM canter defined by transcription factor binding, *y* axes show background-subtracted signal at 6–8 h

Prediction of enhancer regulatory state using chromatin signature

- Is Combined signature can reflect more accurately enhancer activity and therefore predict activity state *de novo ?*

- Active and inactive enhancers contained heterogeneous combinations of chromatin marks and Pol II binding



- *a priori* we know which combinations of marks and/or Pol II occupancy are important and to what degree, we moved from threshold-based correlations of single features to a probabilistic quantitative model that could directly assess which signatures were informative and in which combinations

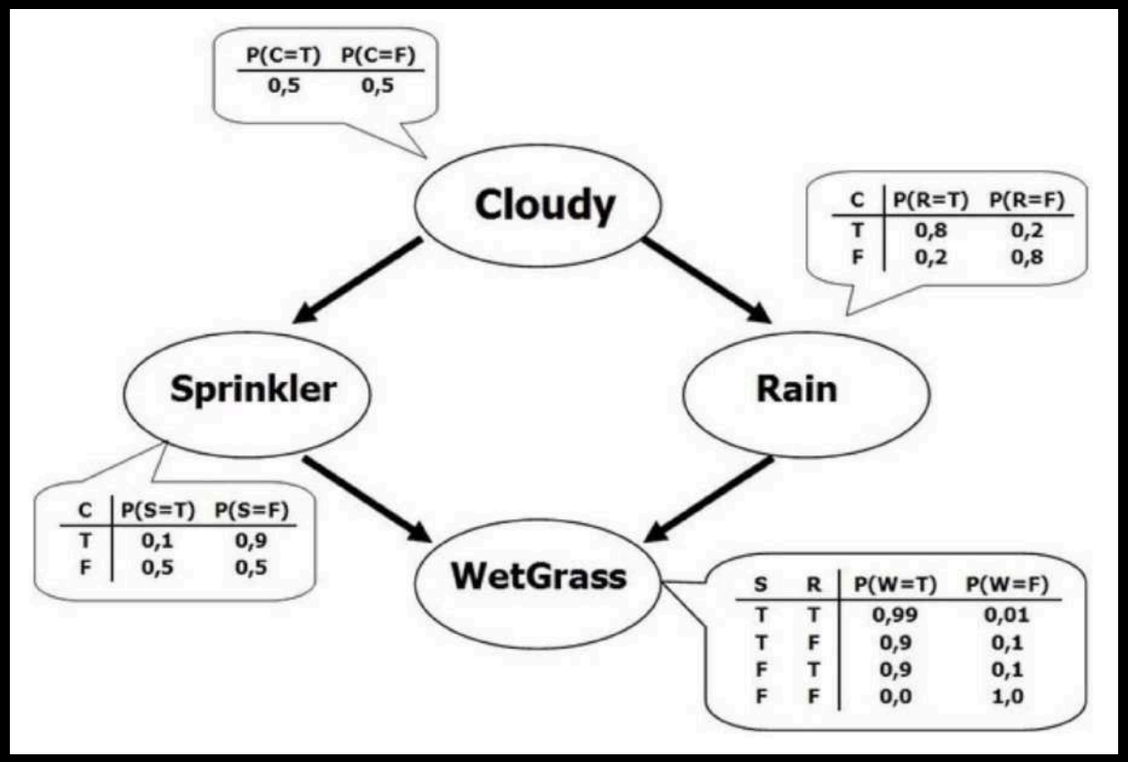
- Here, they are starting with a collection of well-characterised enhancers and used a model to directly learn which features distinguish active and inactive enhancers

- **Bayesian network** inference was used to obtain information on the dependency structure between chromatin marks, Pol II occupancy and enhancer activity

Bayesian Networks

- Bayesian network is a probabilistic graphical model that represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG).

- Bayesian networks are ideal for taking an event that occurred and predicting the likelihood that any one of several possible known causes was the contributing factor



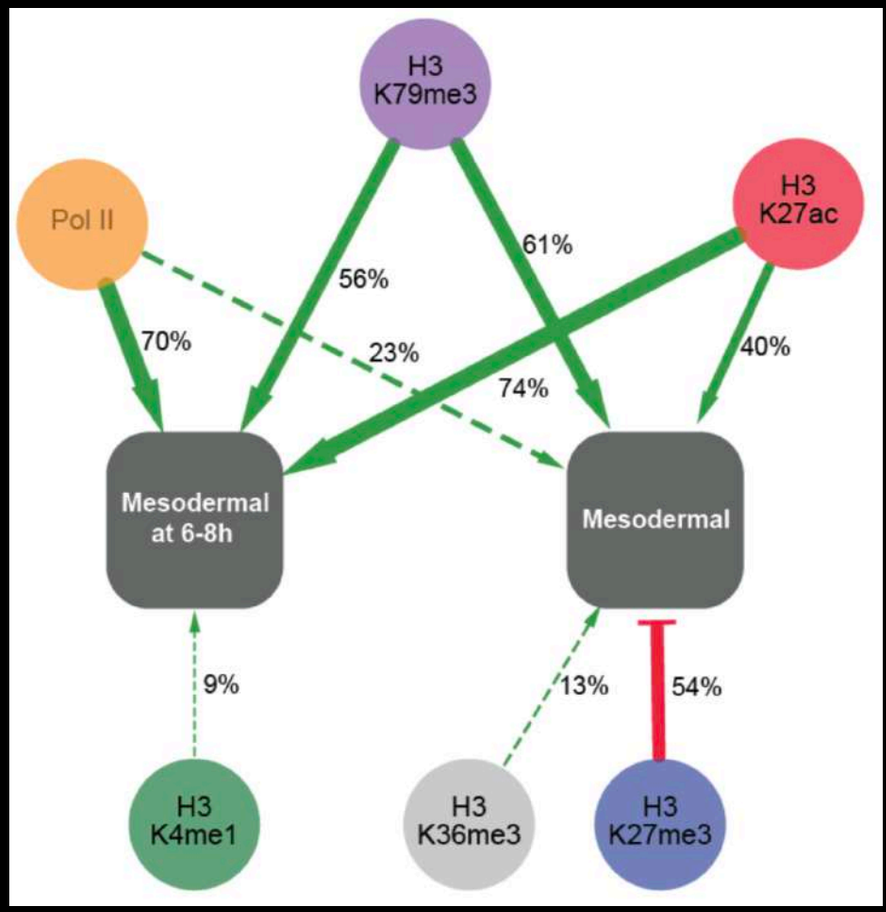
[wikipedia]

Prediction of enhancer regulatory state using chromatin signature

- Bayesian network (BN) inference is a popular tool for describing multivariate probabilistic models with complex structure of dependencies between variables

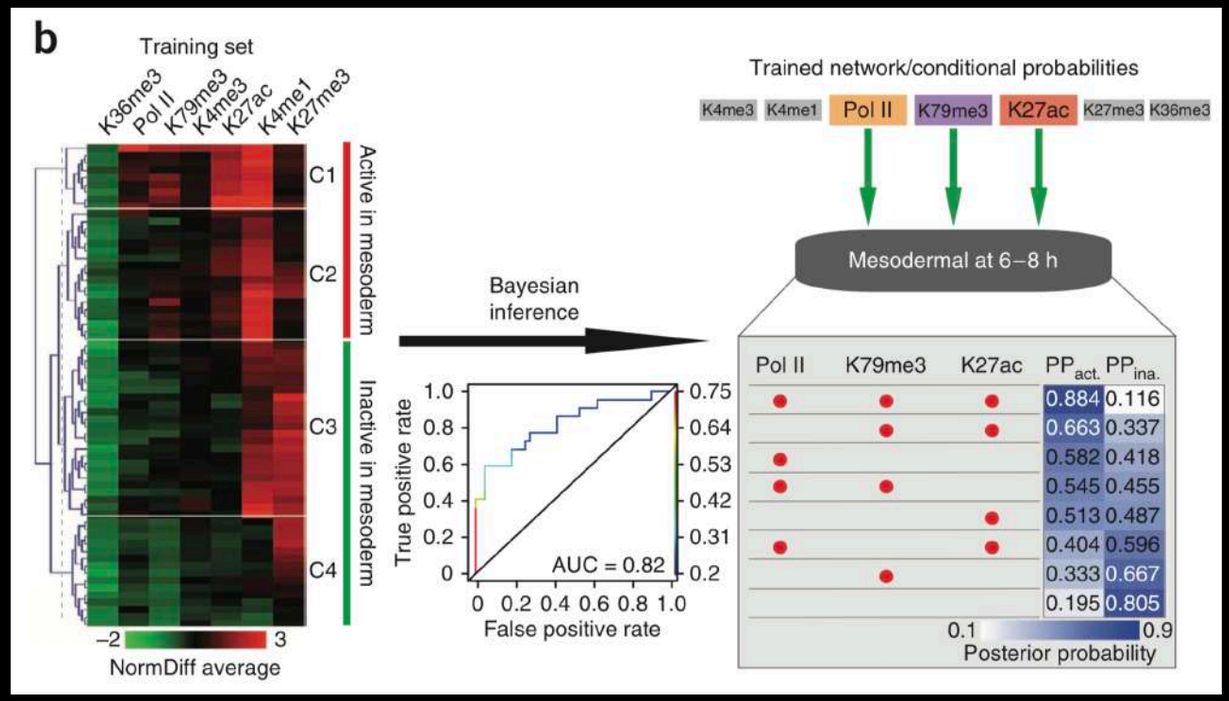
- The model is represented as a graph with nodes representing variables and edges representing conditional dependency of probability distributions between them

- Model consists of two types of variables: observed histone modification and Pol II occupancy quantitative levels are used as “input” variables and the binary enhancer activity classification variables are used as output

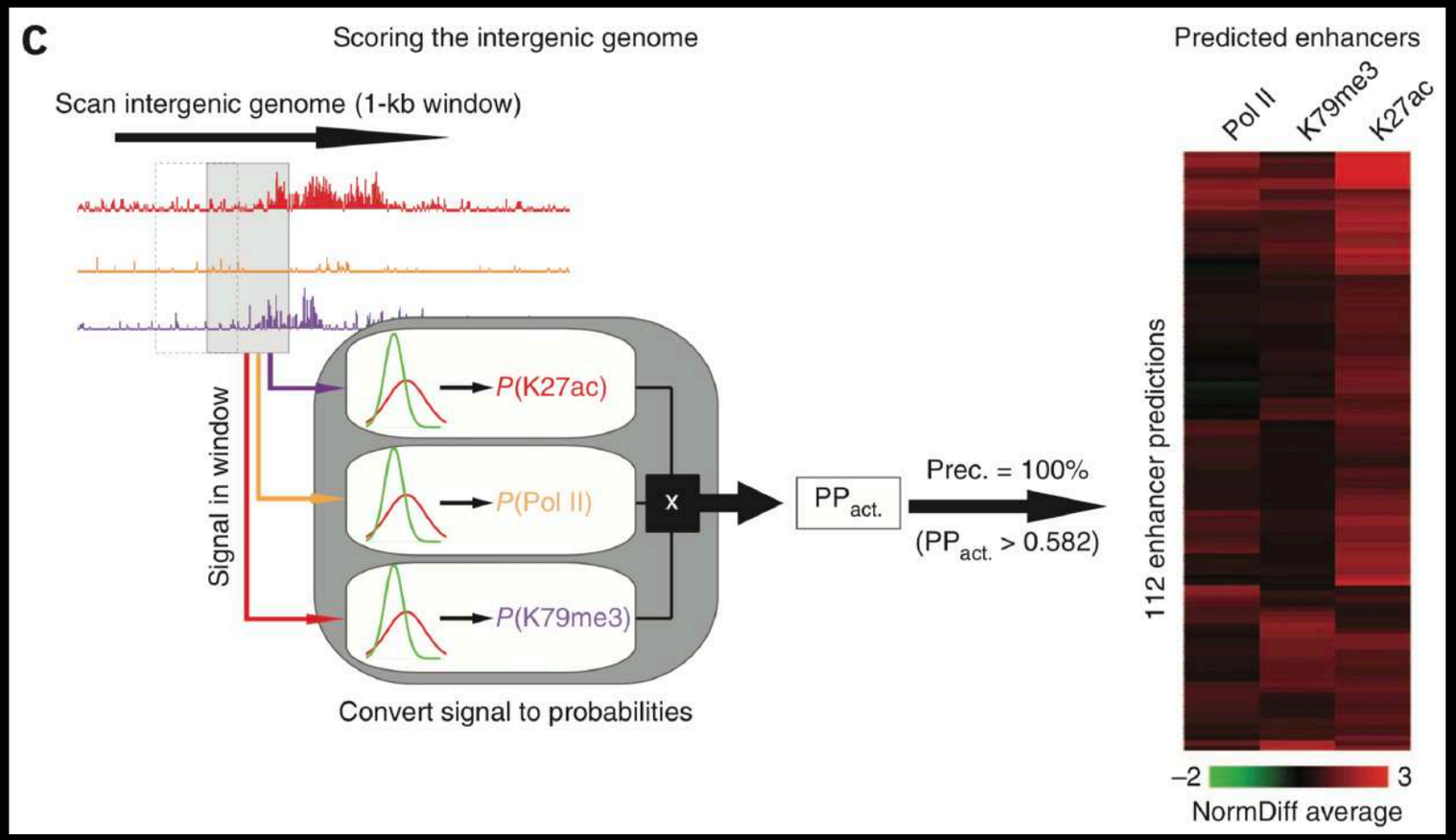


- A Bayesian network topology was reconstructed with two different activity states: a restricted set of enhancers active in mesoderm at 6–8 h of development and a broader set of mesodermally active enhancers

- The trained Bayesian network was validated using a fourfold cross-validation scheme and accurately represents both activity states (area under the receiver operating characteristic (AUC) of 0.82 and 0.76, respectively), independent of enhancer distance from the transcription start site (TSS)



Bayesian modelling of mesodermal enhancers at 6–8 h of development



Predicting mesodermal regulatory regions active at 6–8 h *de novo*

- Trained Bayesian network applied to the *Drosophila* intergenic genome and identified >303 kb of sequence predicted to direct mesodermal activity at 6–8 h of development, using a posterior probability threshold of ≥0.582

- 112 predicted regions had diverse levels of H3K79me3, H3K27ac and Pol II enrichment, with some regions containing all three to varying degrees and other regions lacking one component

- Of the predicted regions, 78% overlapped with TF-Meso- CRMs, indicating that these regions recruit mesodermal transcription factors at exactly these stages of development, suggesting that the majority of predicted regions are likely to function as active regulatory regions *in vivo*


Thank
You