crosslinking approaches to characterize chromatin composition and organization

CBB752
Monday, March 30, 2009

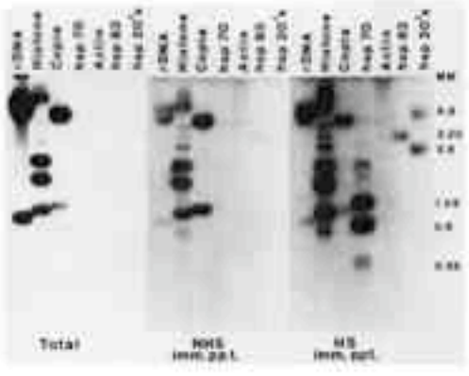
Tae Hoon Kim - taehoon.kim@yale.edu
http://genome.med.yale.edu/index.php/Teaching

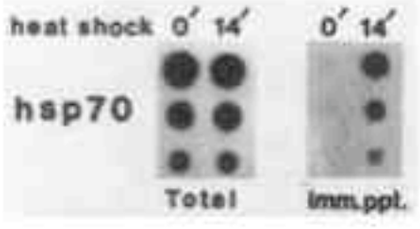
main crosslinking approaches

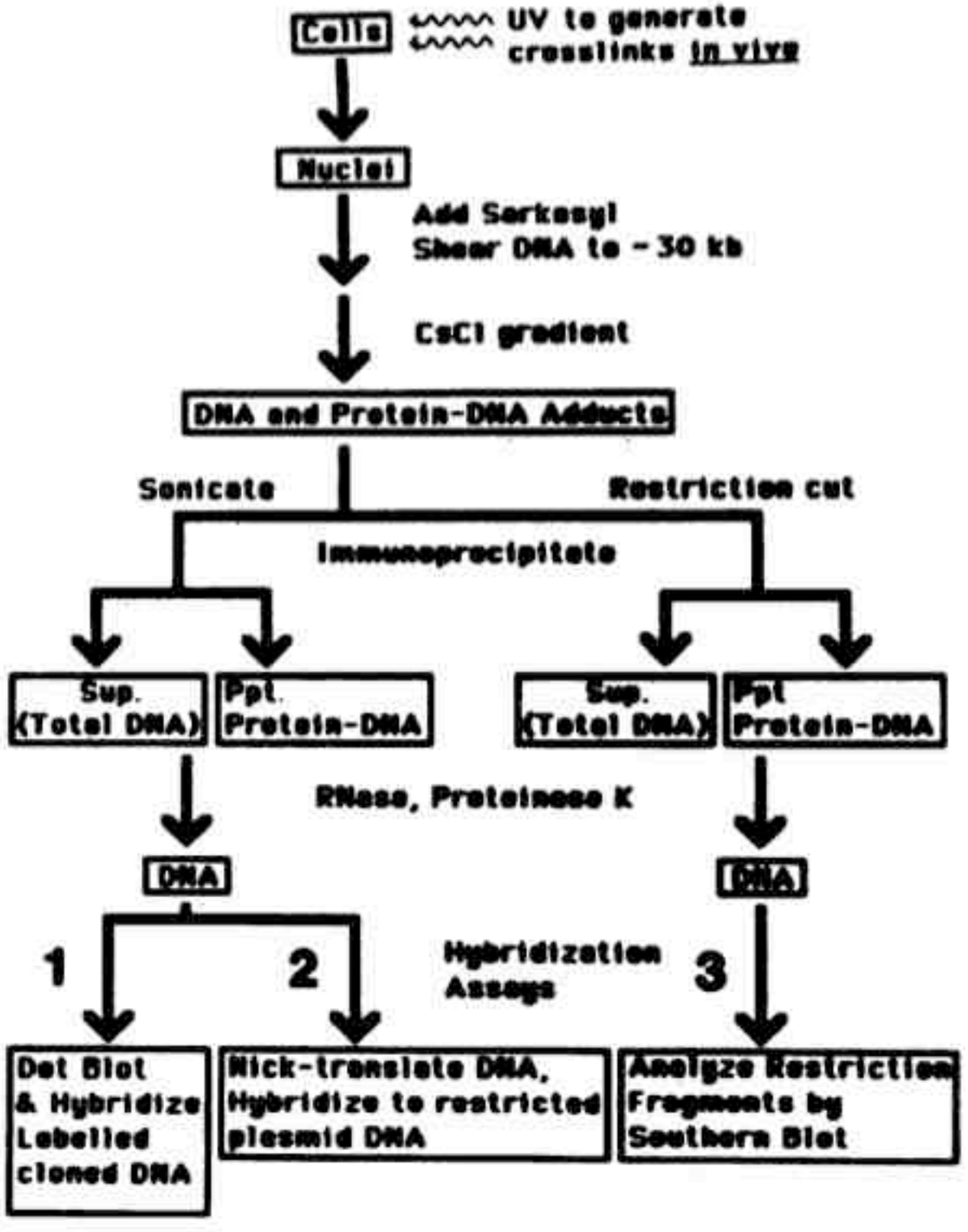
- *ch*romatin *i*mmuno*p*recipitation - *ChIP*
  - *history, rationale, development*
- *c*hromosome *c*onformation *c*apture *(3C)*
  - *history, rationale, development*

chromatin immunoprecipitation

- *first conceptualized by Gilmour and Lis in 1984 & 1985*
- *used UV light to crosslink proteins to DNA in vivo*
- *lyse the cells and immunoprecipitate*
- *Dot or Southern blot to detect DNA fragments associated with the protein*
- *Gilmour and Lis used this method to demonstrate that increased association of RNAP at heat shock genes upon heat shock*

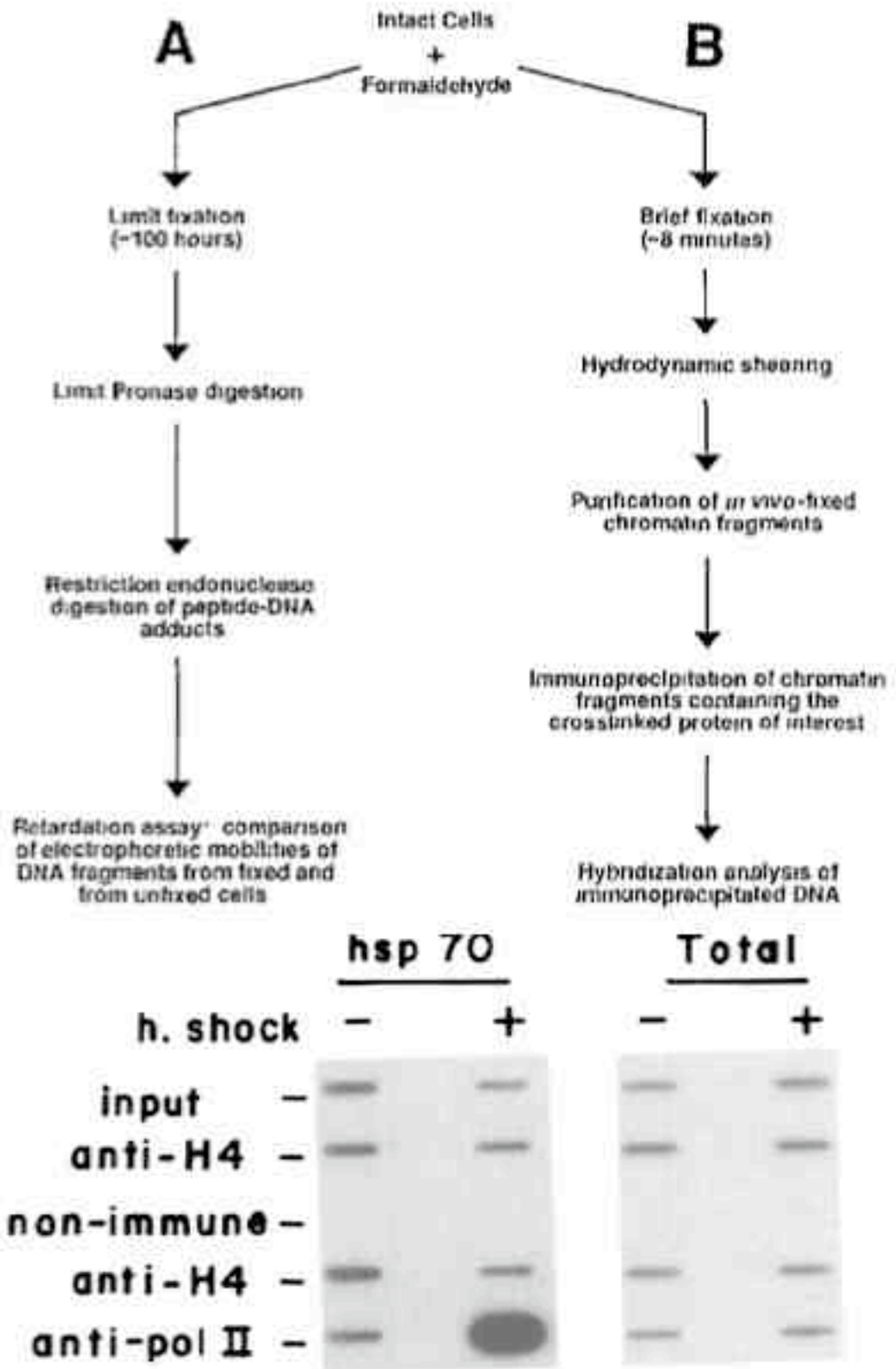






development of formaldehyde for ChIP

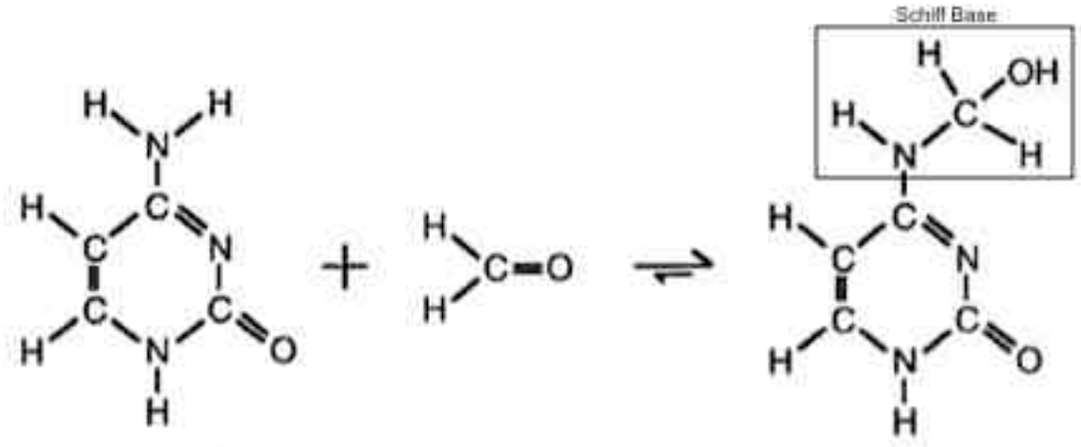
- *UV is a general and efficient crosslinker, but not reversible, thus making the crosslinked DNA largely unsuitable for molecular analysis*
- *Varshavsky pioneered formaldehyde as a crosslinking reagent*
  - *immobilization of protein-DNA interactions (1960s - 1980s)*
  - *restriction enzyme accessibility (Varshavsky, 1979)*
  - *reversal of formaldehyde crosslinks (Solomon and Varshavsky, 1985)*
  - *ChIP using formaldehyde crosslinking (Solomon and Varshavsky, 1988)*



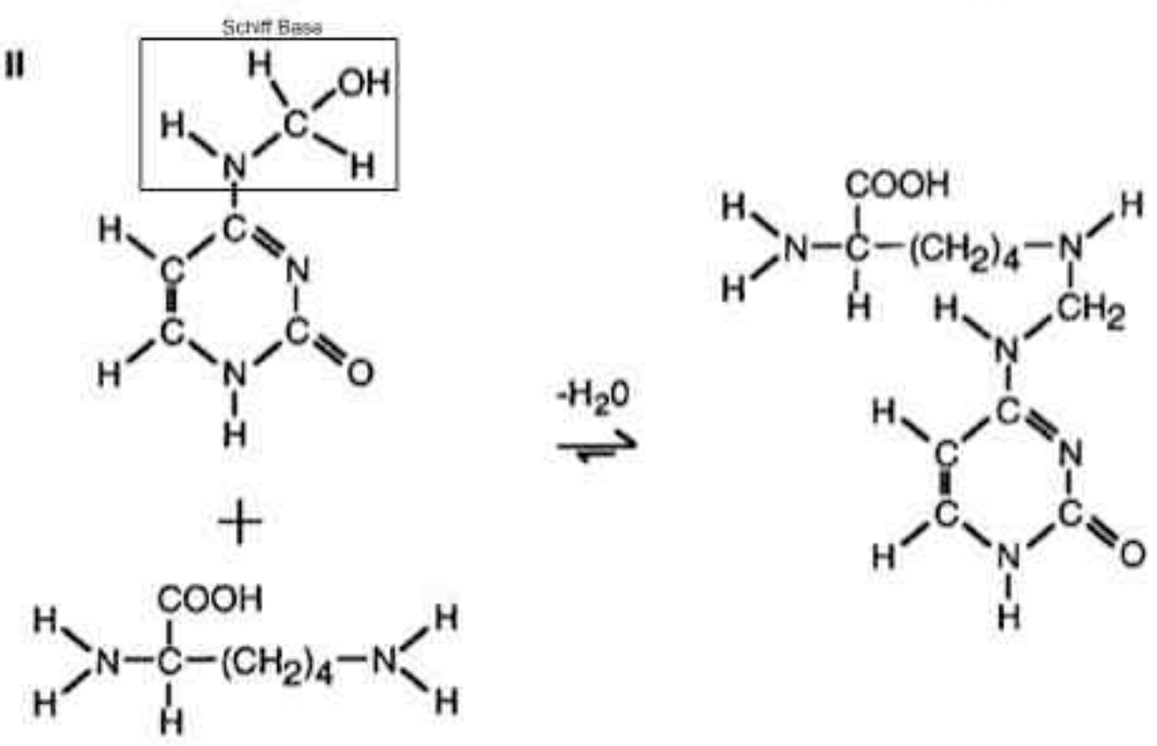
chemistry of formaldehyde crosslinking

- *versatile 2 Å crosslinker*
- *highly cell permeable*
- *limited by availability of primary amines in the vicinity*
- *a selective crosslinker - not general*
- *reversible*

Reaction I



Reaction II

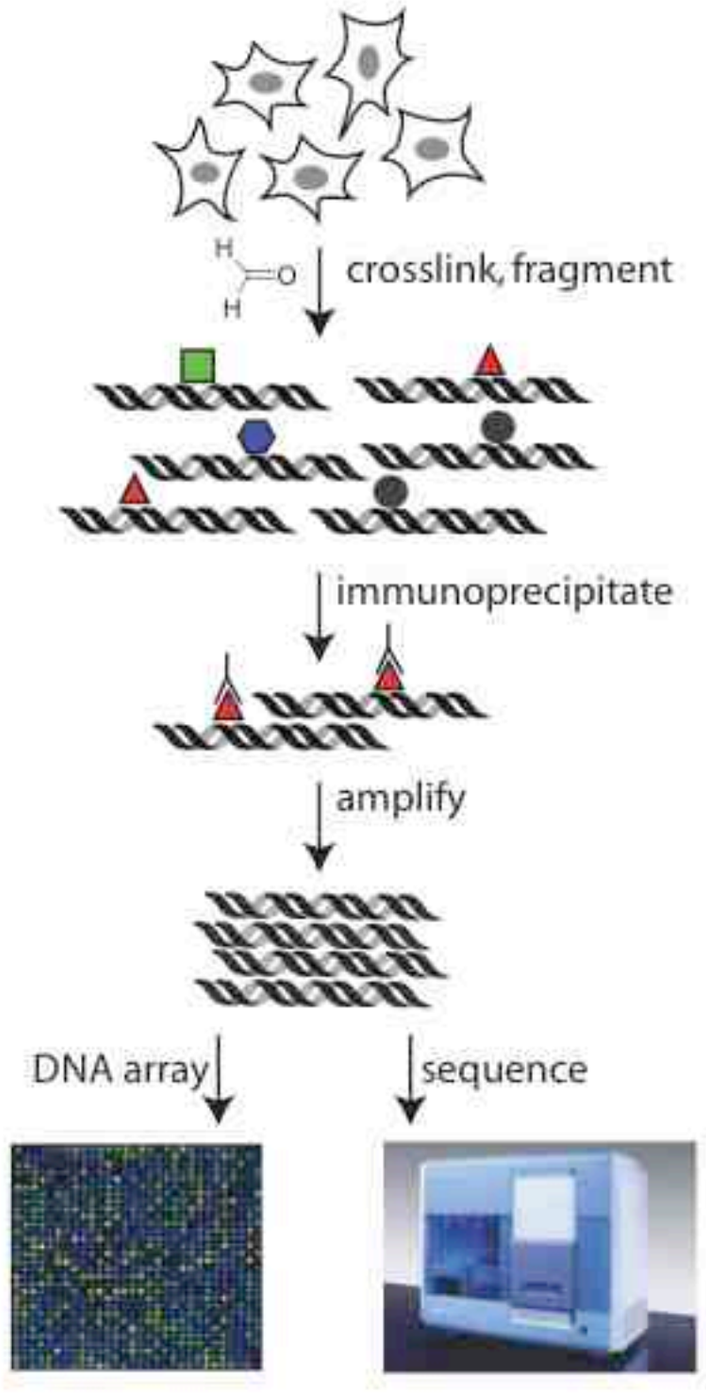


a brief history of ChIP

- *Southern blot to detect immunoprecipitated chromatin (Gilmour & Lis, 1985)*
- *PCR to detect immunoprecipitated chromatin (Hecht & Grunstein, 1996)*
- *high-throughput ChIP*
  - *ChIP-chip (Ren et al, 2000; Iyer et al, 2001)*
  - *ChIP-SAGE, ChIP-SACO, ChIP-PET (Zhao; Goodman; Wen) - Sanger Sequencing*
  - *ChIP-Seq (Johnson et al 2007; Mikkelsen et al 2007) - Illumina Sequencing*
- *3C (Dekker & Kleckner 2002)*

chromatin immunoprecipitation

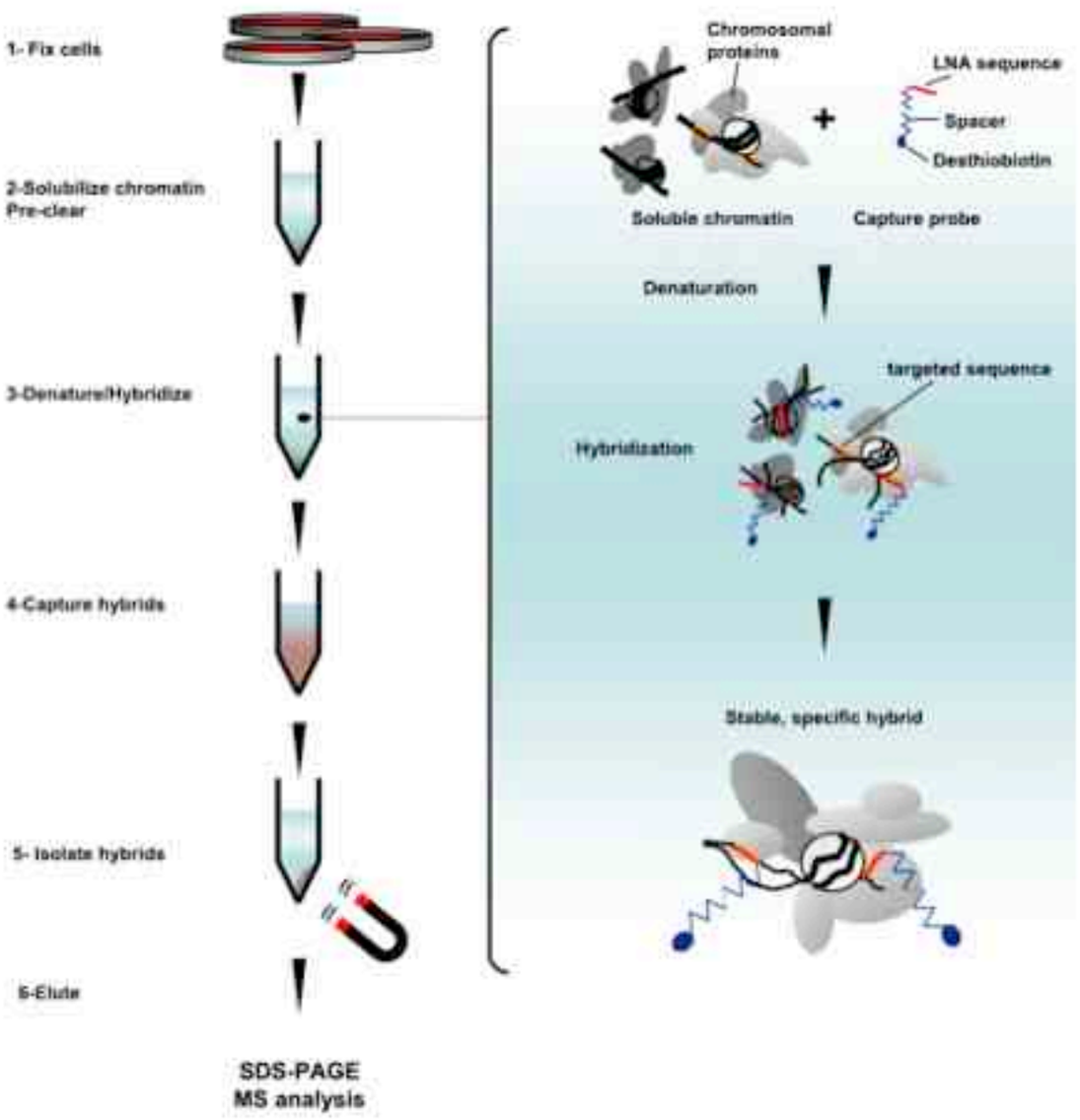
- *allows determination of in vivo transcription factor binding sites*
- *living cells are fixed with crosslinker (formaldehyde)*
- *crosslinked DNA is isolated*
- *covalent protein-DNA complexes are purified using antibodies*
- *isolated DNA is identified using DNA microarrays or high throughput sequencing to determine the binding sites*

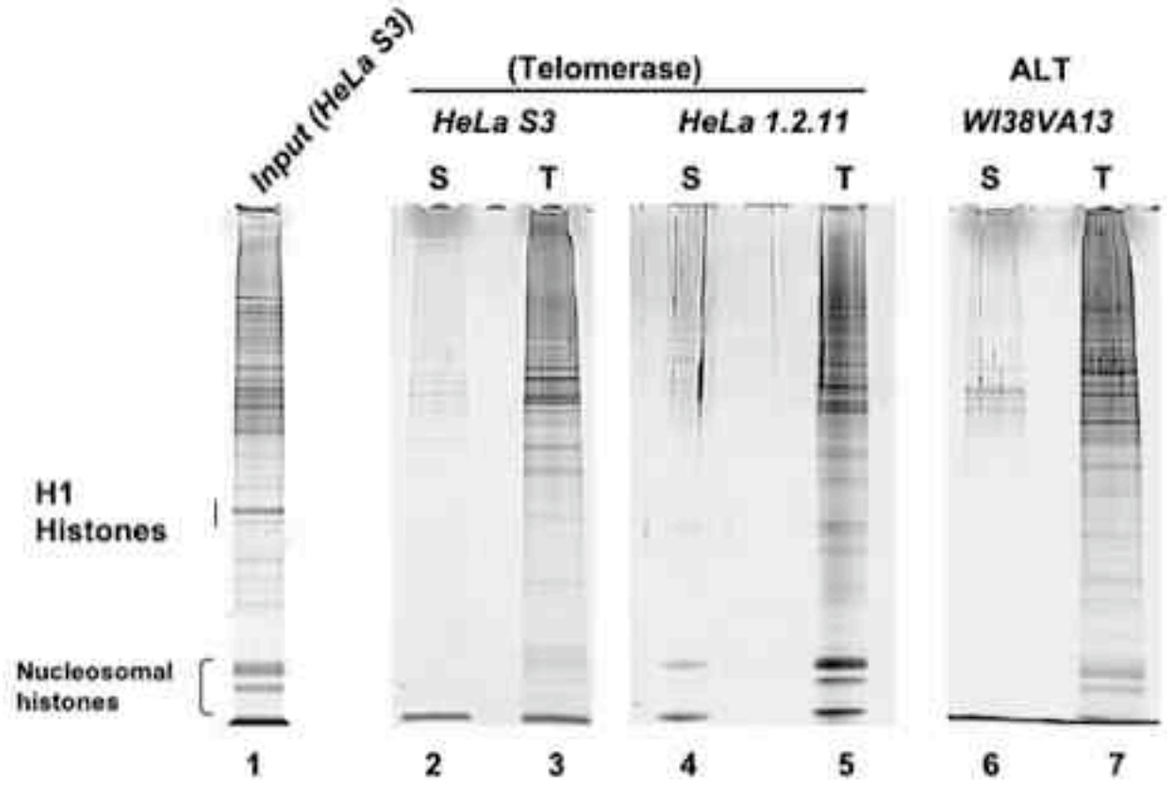


chromatin and regulatory features in the genome

- *histone code for euchromatin and heterochromatin*
- *bivalent chromatin in ES cells*
- *coordination of histone methylation and DNA methylation*

PICh (proteomics of isolated chromatin segments) reverse ChIP

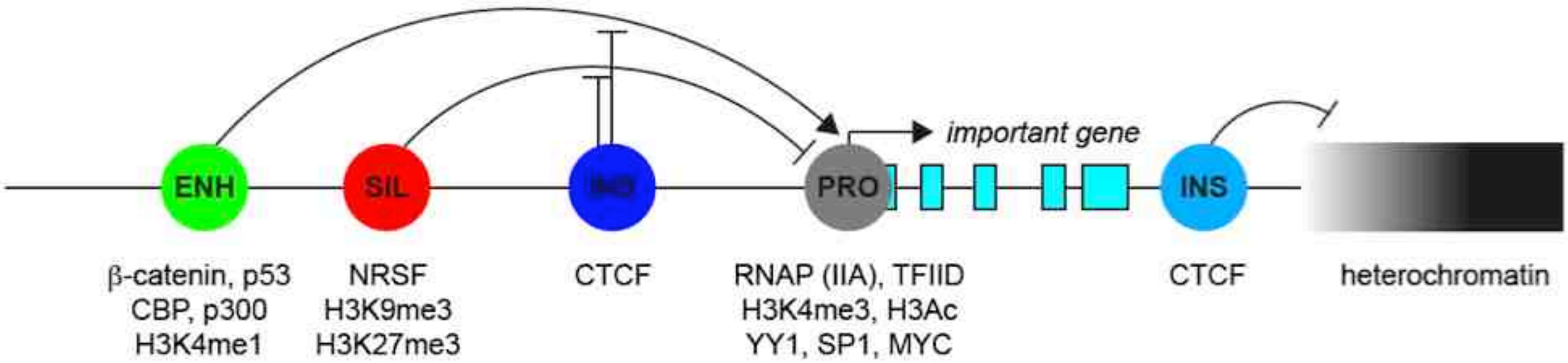




Déjardin and Kingston Cell 136:175 (2009)

mapping all non-coding functional elements

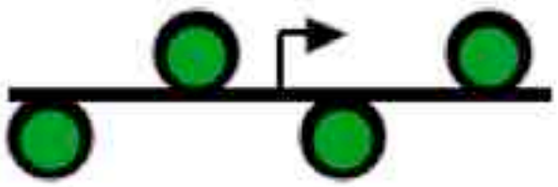
- *map all types of known regulatory elements*
  - *promoters, enhancers, silencers, insulators*
- *bound by specific transcription factors*
- *associated with different histone modification marks or histone variants*



promoter mapping

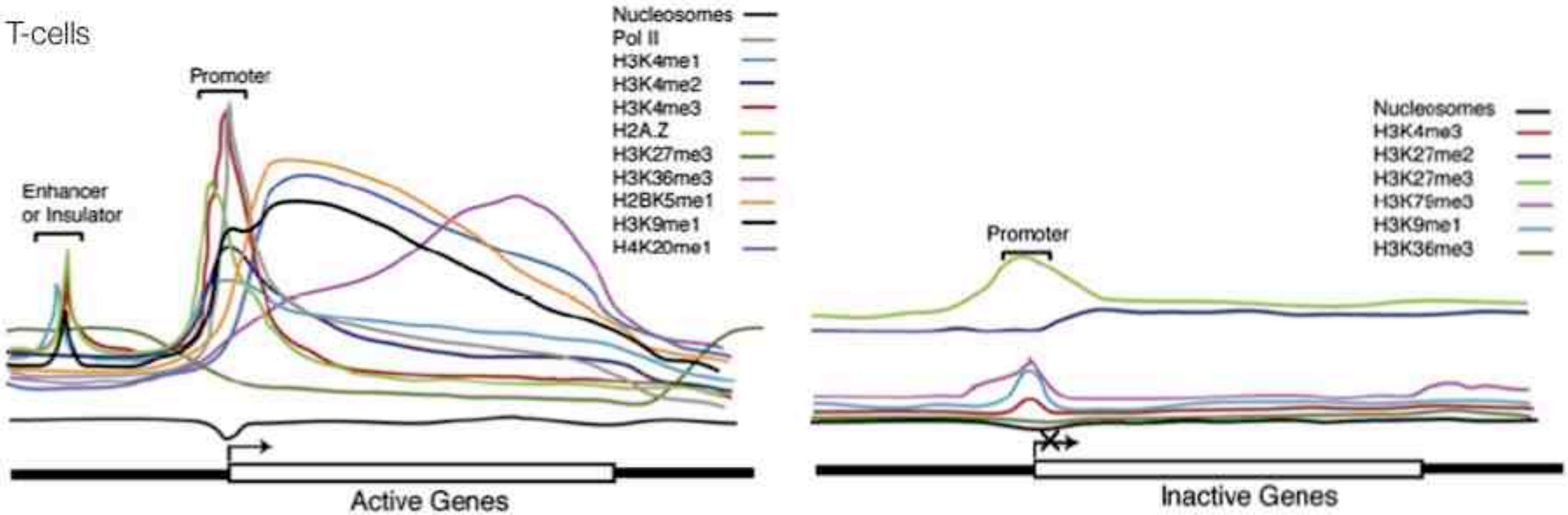
- *ChIP or cDNA library sequencing*
- *use promoter binding information*
  - *to find genes, classify active fraction of the genome and relationship between expression and promoter structure*
- *complexity of promoter architecture*
  - *many genes have multiple promoters*
- *absence of TATA box in most promoters*
- *active promoters are associated with H3K4me3 and other histone marks*
- *transcriptional units (of genes) can be defined by combination of histone marks - without the knowledge of cDNAs*

simple histone code for active and inactive chromatin



Active Promoter Marks		Silent Promoter Marks
H2BK5ac	**H4K5ac**	H3K9me2,3
H2BK12ac	**H4K8ac**	H3K27me2,3
H2BK20ac	**H4K91ac**	H4K20me3
H2BK120ac	**H2A.Z**	
H3K9me1	H3K27me1	
H3K4me1,2,3	H4K20me1	
H3K4ac	H3K36me3	
H3K9ac	H2BK5me1	
H3K18ac		
H3K27ac		
H3K36ac		

histone modifications at promoters and enhancers



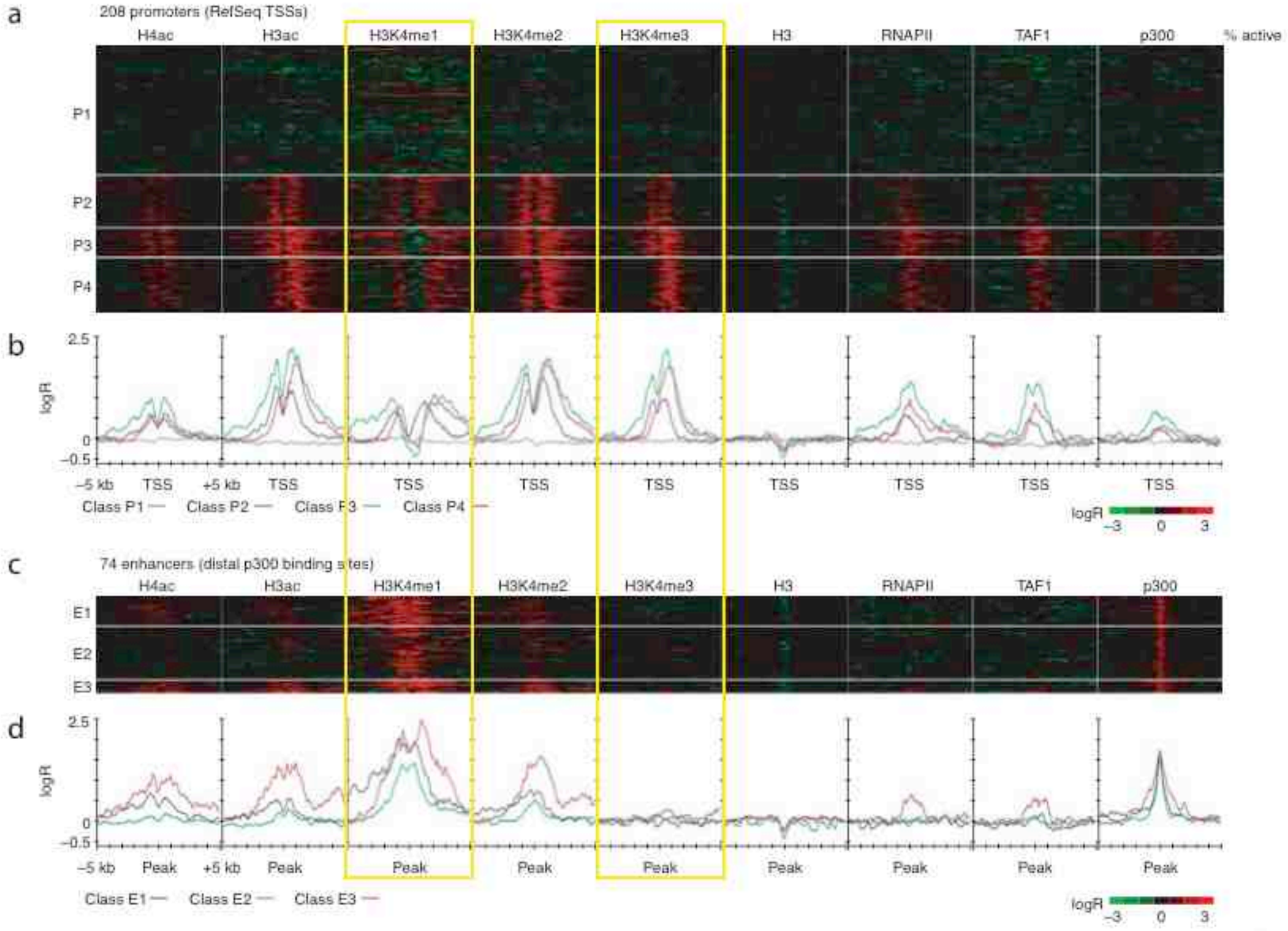
- *gene activity and histone modification*
  - *genes activity can be predicted from histone modification signature*
  - *direction of transcription can be determined from histone modification pattern*
- *RNA polymerase is paused at the promoter - transcription elongation (not polymerase recruitment) might be the rate limiting step for gene expression*

Barski et al Cell 129:823 (2007)

enhancer mapping

- *find binding sites for sequence specific transcription factors*
- *catalog of transcription factor binding sites is beginning to be assembled*
  - *p53, ER, REST/NRSF, TCF7L2, STAT1, ...*
- *a large number of binding sites scattered everywhere in the genome (a couple of thousand to tens of thousand enhancers)*
- *associated with DNase I hypersensitive sites*
- *challenges*
  - *sorting functional/critical sites from nonfunctional/redundant sites*
  - *assigning target genes that are directly regulated by each binding site*

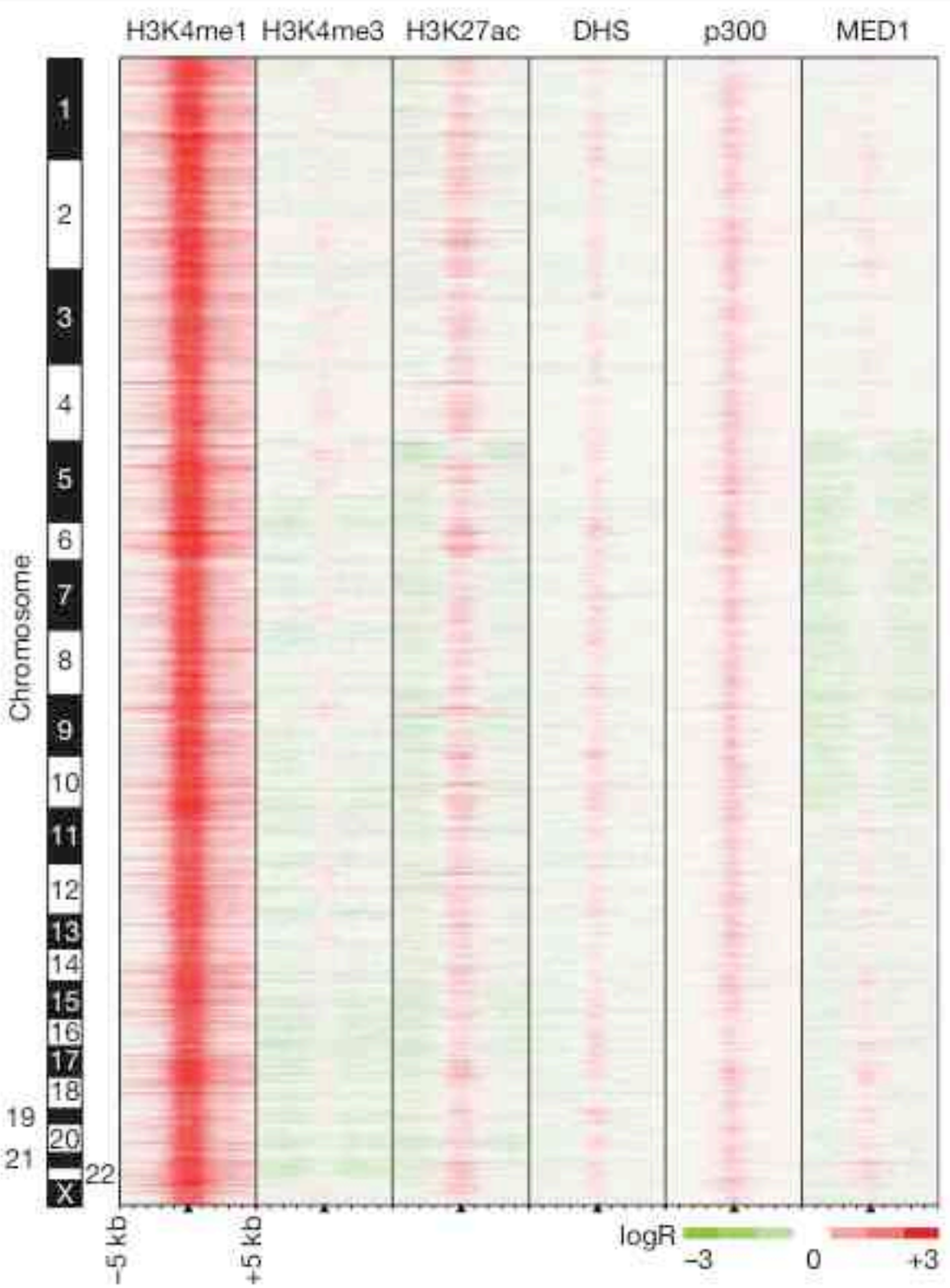
differential pattern of histone H3 methylation at promoters and enhancers

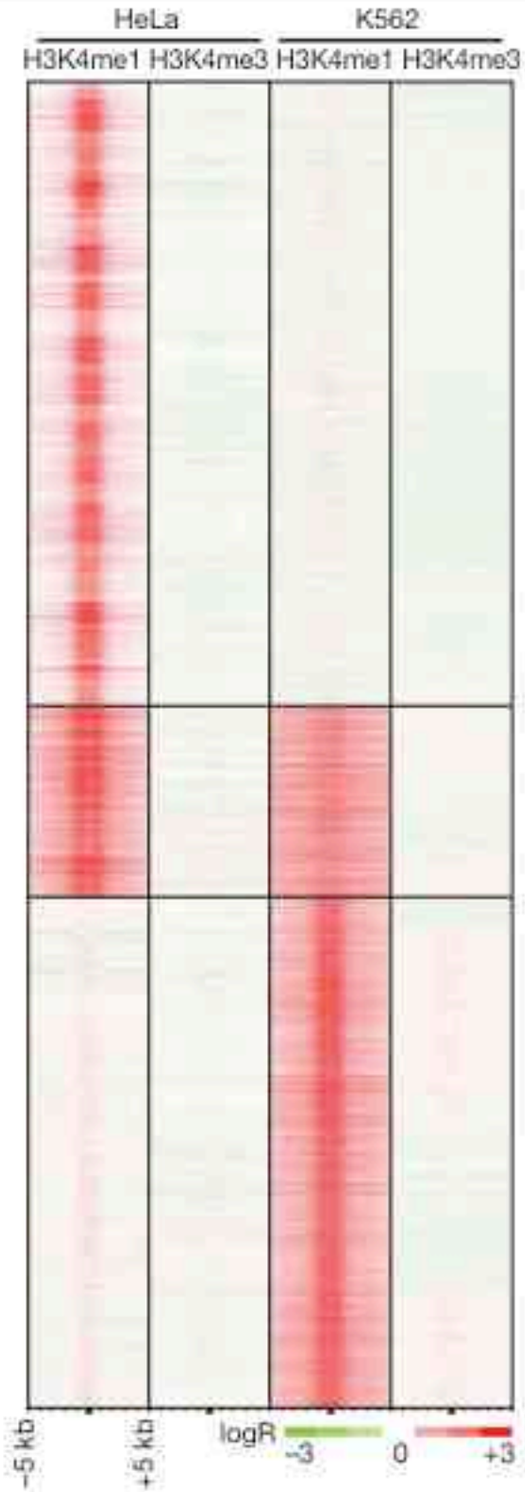


Heintzman et al Nat Genet 39:311 (2007)

genome-wide identification of enhancers

- *over 36,000 enhancers in HeLa cells*
- *over 24,000 enhancers in K562 cells*
- *mostly cell type specific*
- *mapping their domain of action*
- *modified prior to functional response or use of enhancers*
- *how to restrict enhancer function?*

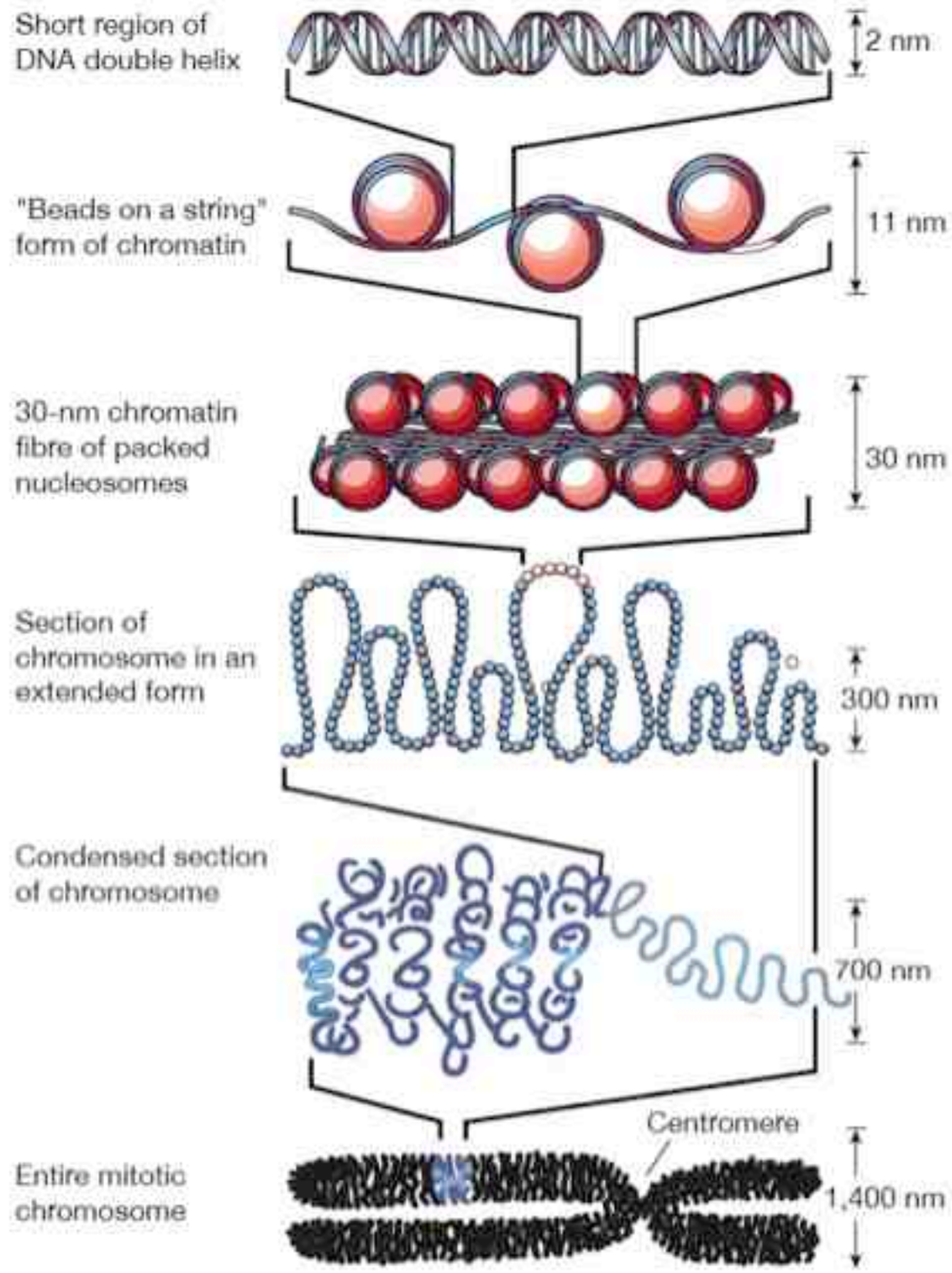




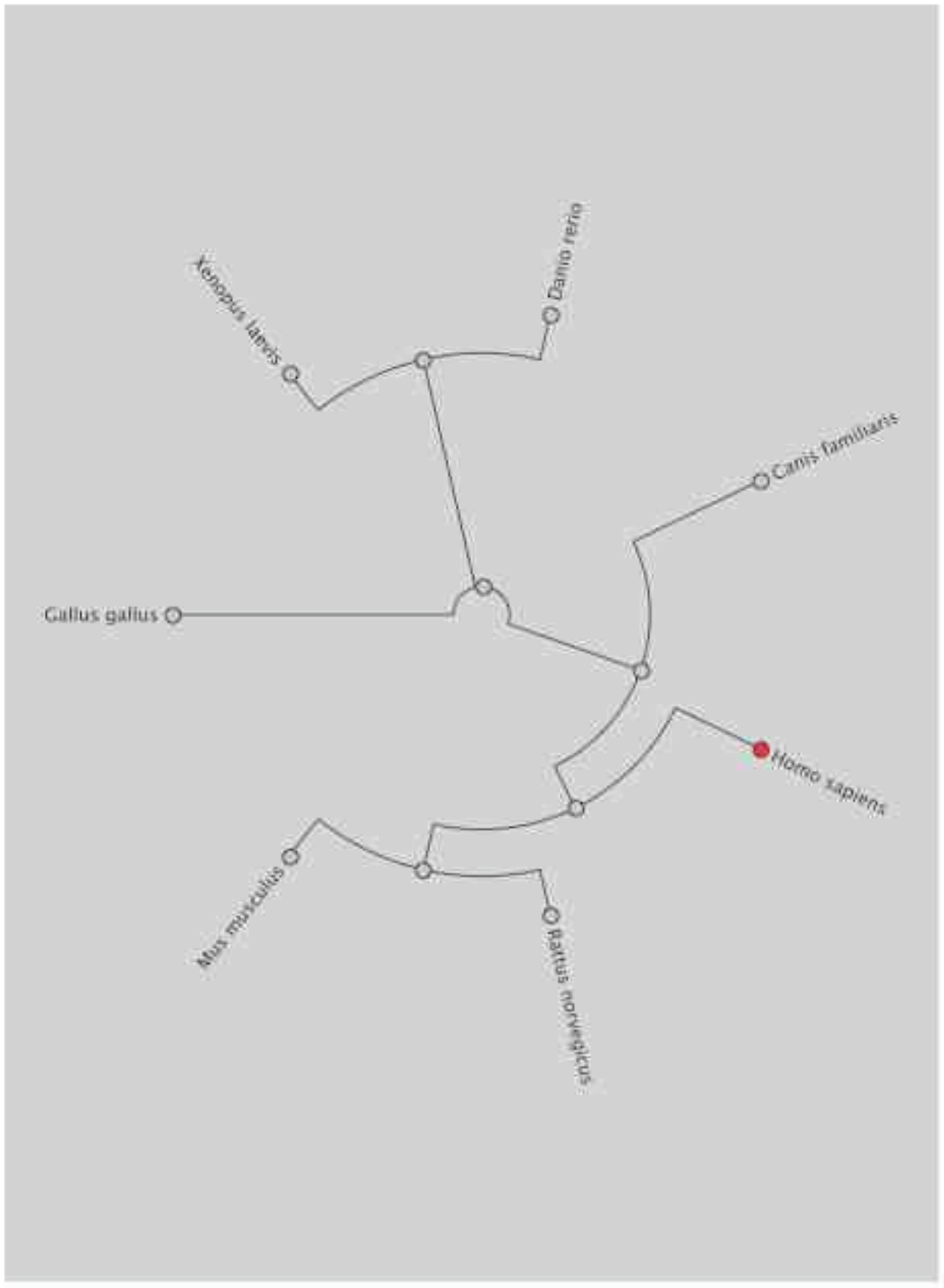
Heintzman et al, Nature 2009

chromatin and regulatory domains

- *cytologically distinguishable structures*
  - *telomeres, centromeres, nucleolus*
- *hetero- and euchromatin, transcription factories and compartments*
  - *depending on the position, an identical transgene can exhibit eight-fold change in expression*
- *these distinct domains are critical for many processes, including:*
  - *gene regulation*
  - *X-inactivation*
  - *recombination, replication*
- ***how are these domains established and maintained?***



Felsenfeld & Groudine, Nature 421:453 (2003)



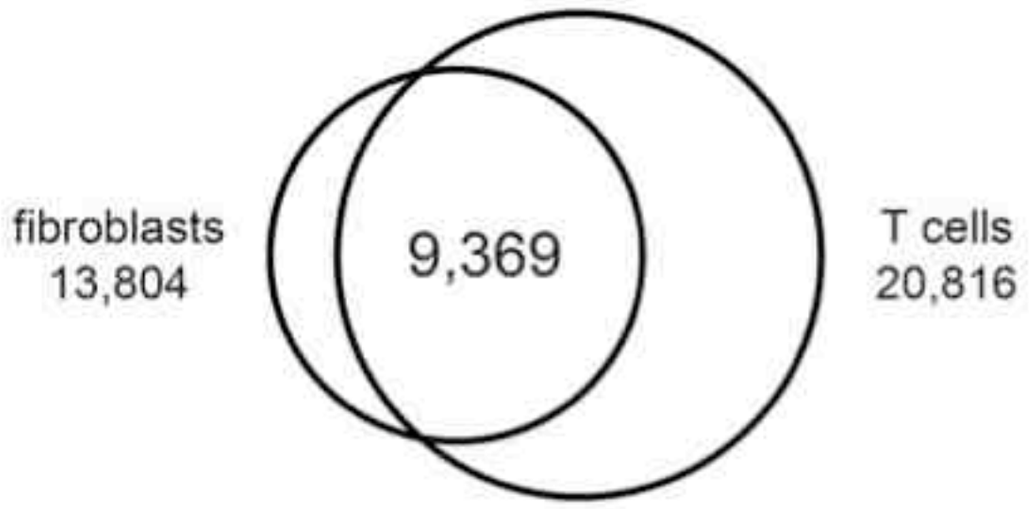

```
human     kkkgvkktfqcelcsytcprrsnldrhmkshtderphkchlcgrafrtvtllrnhlntht
dog       kkkgvkktfqcelcsytcprrsnldrnmkshtderphkchlcgrafrtvtllrnhlntht
rat       kkkgvkktfqcelcsytcprrsnldrhmkshtderphkchlcgrafrtvtllrnhlntht
mouse     kkkgvkktfqcelcsytcprrsnldrhmkshtderphkchlcgrafrtvtllrnhlntht
chicken   kkkgvkktfqcelcsytcprrsnldrnmkshtderphkchlcgrafrtvtllrnhlntht
frog      kkkgvkktfqcelcsytcprrsnldrhmkshtderphkchlcgrafrtvtllrnhlntht
fish      kkkgvkktfqcelcsytcprrsnldrhmkshtderphkchlcgrafrtvtllrnhlntht

human     gtrphkcpdcdmafvtsgelvrhrrykhthekpfkcsmcdyasvevsklkrhirshtger
dog       gtrphkcpdcdmafvtsgelvrhrrykhthekpfkcsmcdyasvevsklkrhirshtger
rat        gtrphkcpdcdmafvtsgelvrhrrykhthekpfkcsmcdyasvevsklkrhirshtger
mouse     gtrphkcpdcdmafvtsgelvrhrrykhthekpfkcsmcdyasvevsklkrhirshtger
chicken   gtrphkcpdcdmafvtsgelvrhrrykhthekpfkcsmcdyasvevsklkrhirshtger
frog      gtrphkcpdcdmafvtsgelvrhrrykhthekpfkcsmcdyasvevsklkrhirshtger
fish      gtrphkctdcdmafvtsgelvrhrrykhthekpfkcsmcdyasvevsklkrhirshtger

human     pfqcslcsyasrdtyklkrhmrthsgekpyecyicharftqsgtmkmhilqkhtenvakf
dog       pfqcslcsyasrdtyklkrhmrthsgekpyecyicharftqsgtmkmhilqkhtenvakf
rat       pfqcslcsyasrdtyklkrhmrthsgekpyecyicharftqsgtmkmhilqkhtenvakf
mouse     pfqcslcsyasrdtyklkrhmrthsgekpyecyicharftqsgtmkmhilqkhtenvakf
chicken   pfqcslcsyasrdtyklkrhmrthsgekpyecyicharftqsgtmkmhilqkhtenvakf
frog      pfqcslcsyasrdtyklkrhmrthsgekpyecyicharftqsgtmkmhilqkhtenvakf
fish      pfqcslcsyasrdtyklkrhmrthsgekpyecyicharftqsgtmkmhilqkhtenvakf

human     hcphcdtviarksdlgvhlrkqhsyieqgkkcrycdavfheryaliqhqkshknekrfkc
dog       hcphcdtviarksdlgvhlrkqhsyieqgkkcrycdavfheryaliqhqkshknekrfkc
rat       hcphcdtviarksdlgvhlrkqhsyieqgkkcrycdavfheryaliqhqkshknekrfkc
mouse     hcphcdtviarksdlgvhlrkqhsyieqgkkcrycdavfheryaliqhqkshknekrfkc
chicken   hcphcdtviarksdlgvhlrkqhsyieqgkkcrycdavfheryaliqhqkshknekrfkc
frog      hcphcdtviarksdlgvhlrkqhsyieqgkkcrycdtvfheryaliqhqkshknekrfkc
fish      hcphcdtviarksdlgvhlrkqhsyieqgrkcrycdavfheryaliqhqkshknekrfkc

human     dqcdyacrqerhmimhkrthtgekpyacshcdktfrqkglldmhfkryhdpnfvpaafvc
dog       dqcdyacrqerhmimhkrthtgekpyacshcdktfrqkglldmhfkryhdpnfvpaafvc
rat       dqcdyacrqerhmimhkrthtgekpyacshcdktfrqkglldmhfkryhdpnfvpaafvc
mouse     dqcdyacrqerhmimhkrthtgekpyacshcdktfrqkglldmhfkryhdpnfvpaafvc
chicken   dqcdyacrqerhm mhkrthtgekpyacshcdktfrqkglldmhfkryhdpnfvpaafvc
frog      dqc yacrqerhmimhkrthtgekpyacshcdktfrqkglldmhfkryhdp fvpaafvc
fish      dqcdyacrqerhm mhkrthtgekpyacs c ktfrqkglldmhf ryhdpnfvp  fvc

human     skcgktftrrntmarhadncqgpdgvegengge----tkkskrgrkrkmrskkedssdse
dog       skcgktftrrntmarhadncqgpdgvegengge----tkkskrgrkrkmrskkedssdse
rat        skcgktftrrntmarhadncqgpdgvegengge----tkkskrgrkrkmrskkedssdse
mouse     skcgktftrrntmarhadncqgpdgvegengge----tkkskrgrkrkmrskkedssdse
chicken   skcgktftrrntmarhadncsgldggegengge----tkkgkrgrkrkmrskkedssdse
frog      skcgktftrrntm rhad cqgpdgtdgengegevthkkgkrgrkrkmrskkegstdse
fish      kcgktftrrntmarha ncqgndsddgengtp----pkrgrggrkrkmrsrkddddddd
```

CTCF is highly conserved in all vertebrates, **>95%** aa identity in its DNA Binding Domain

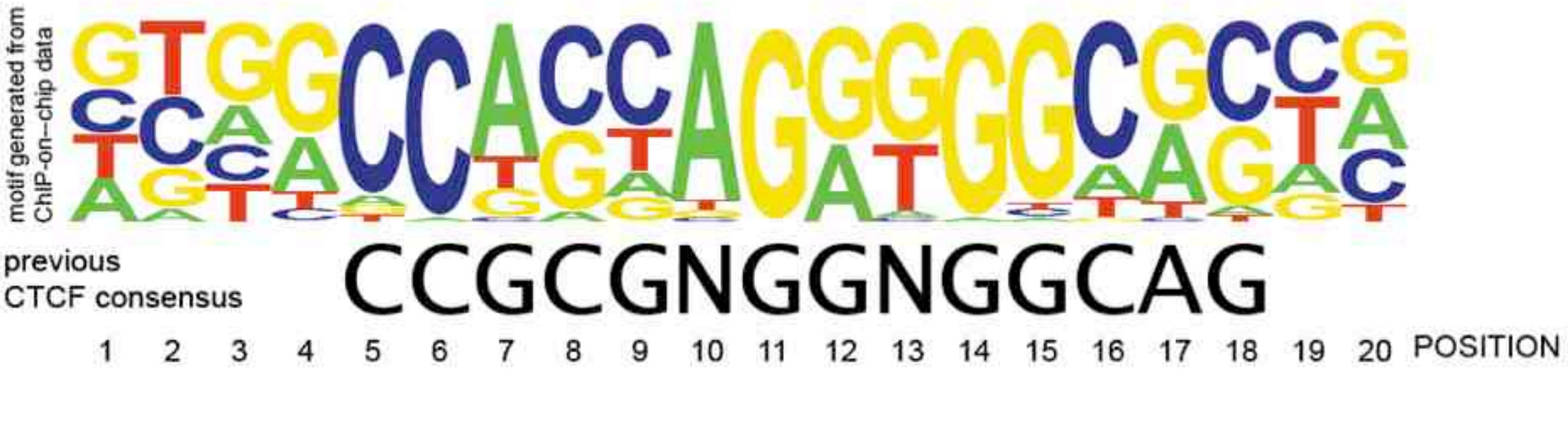
mapping CTCF binding sites in the human genome

- *CTCF ChIP-on-chip (Kim et al, 2007)*
  - *primary human fibroblasts*
  - *38 arrays covering the entire human genome*
    - *50-mer oligo at every 100bps*
    - *14.5 million 50-mer oligos*
      - *1.45 billion bases of **non-repetitive** human genomic DNA covered*
      - *100bp resolution*
      - *strategy described in Kim et al, Nature 436:876-880 (2005)*
  - *custom-designed condensed array for ChIP-chip validation*
  - *identified **13,804** CTCF binding sites*
    - *specificity >98%, sensitivity >88%*

- *CTCF ChIP-Seq (Barski et al, 2007)*
  - *primary human T-cells*
  - *Solexa/Illumina platform*
    - *2.9 million tags - ~20,000 sites*
    - *recovered a number of CTCF binding sites in the repeat regions*
    - *at 8 tags or greater, about 70% (~10,000) of CTCF ChIP-chip sites can be recovered*



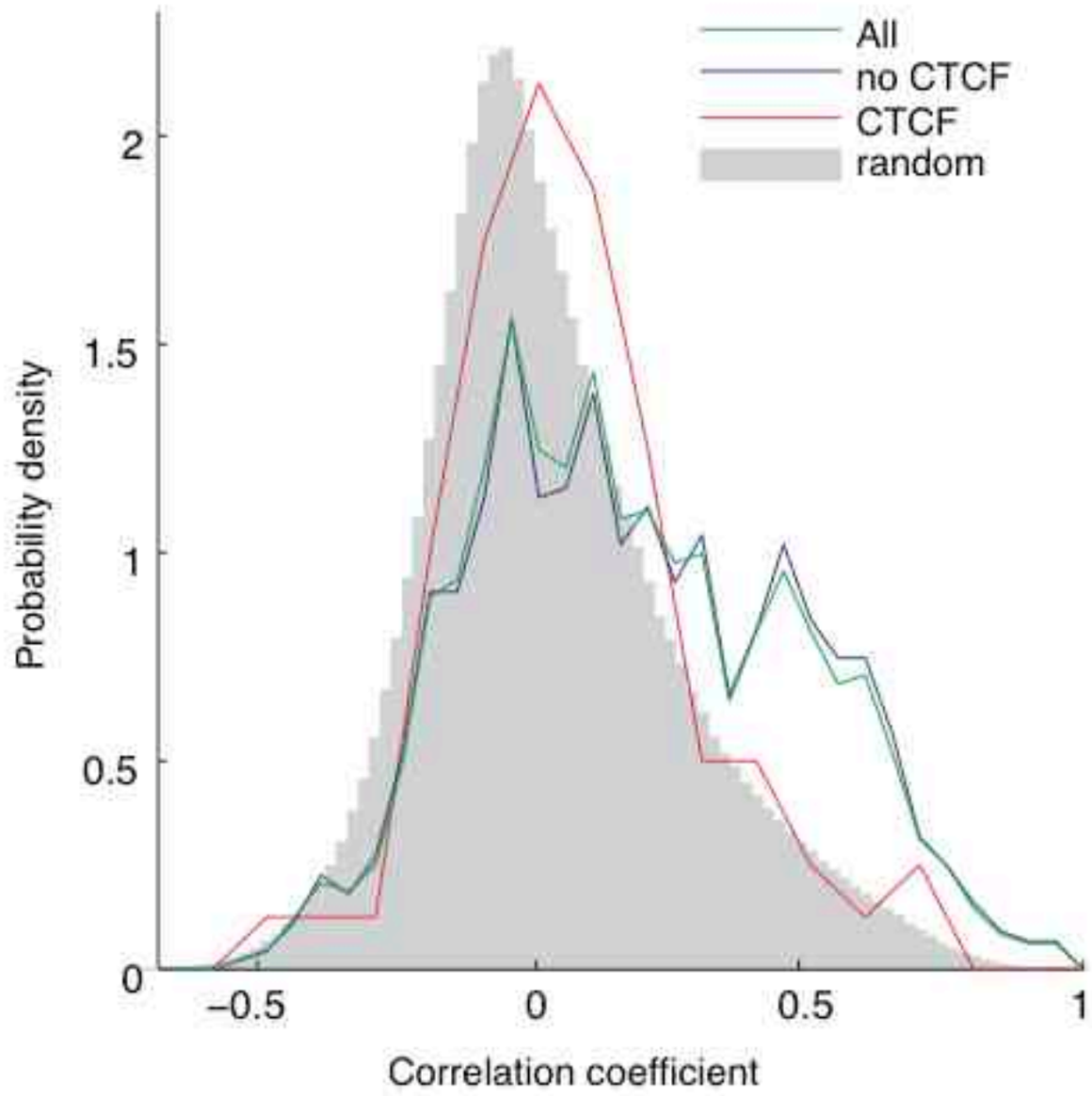
20-mer motif determines CTCF binding throughout the genome



- *a vast majority of CTCF binding sites characterized by a specific 20-mer motif*
  - *present in >95% CTCF binding sites*
- *most highly conserved noncoding sequence element in the genome*

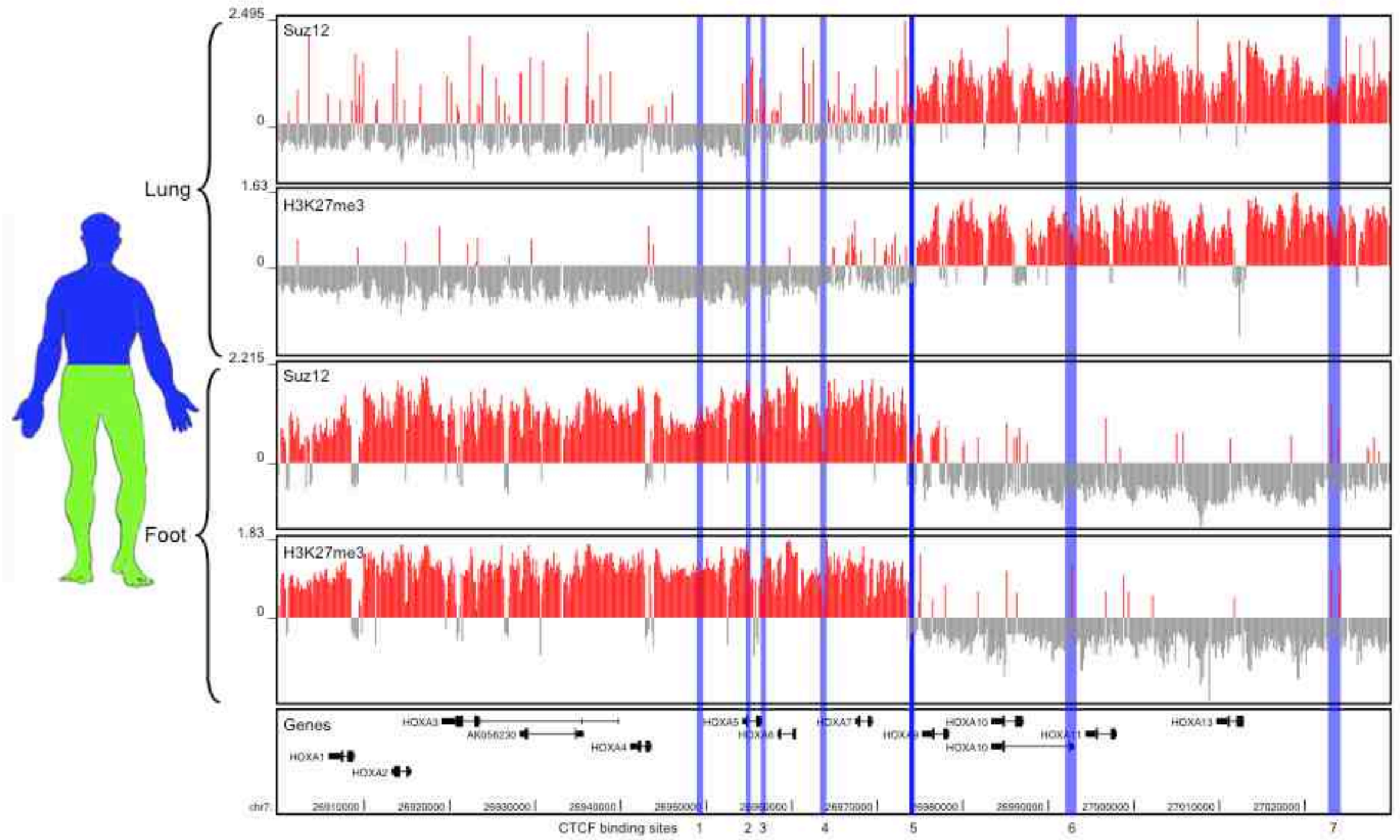
Kim et al Cell 128:1231 (2007)

CTCF binding break correlated expression of neighboring genes



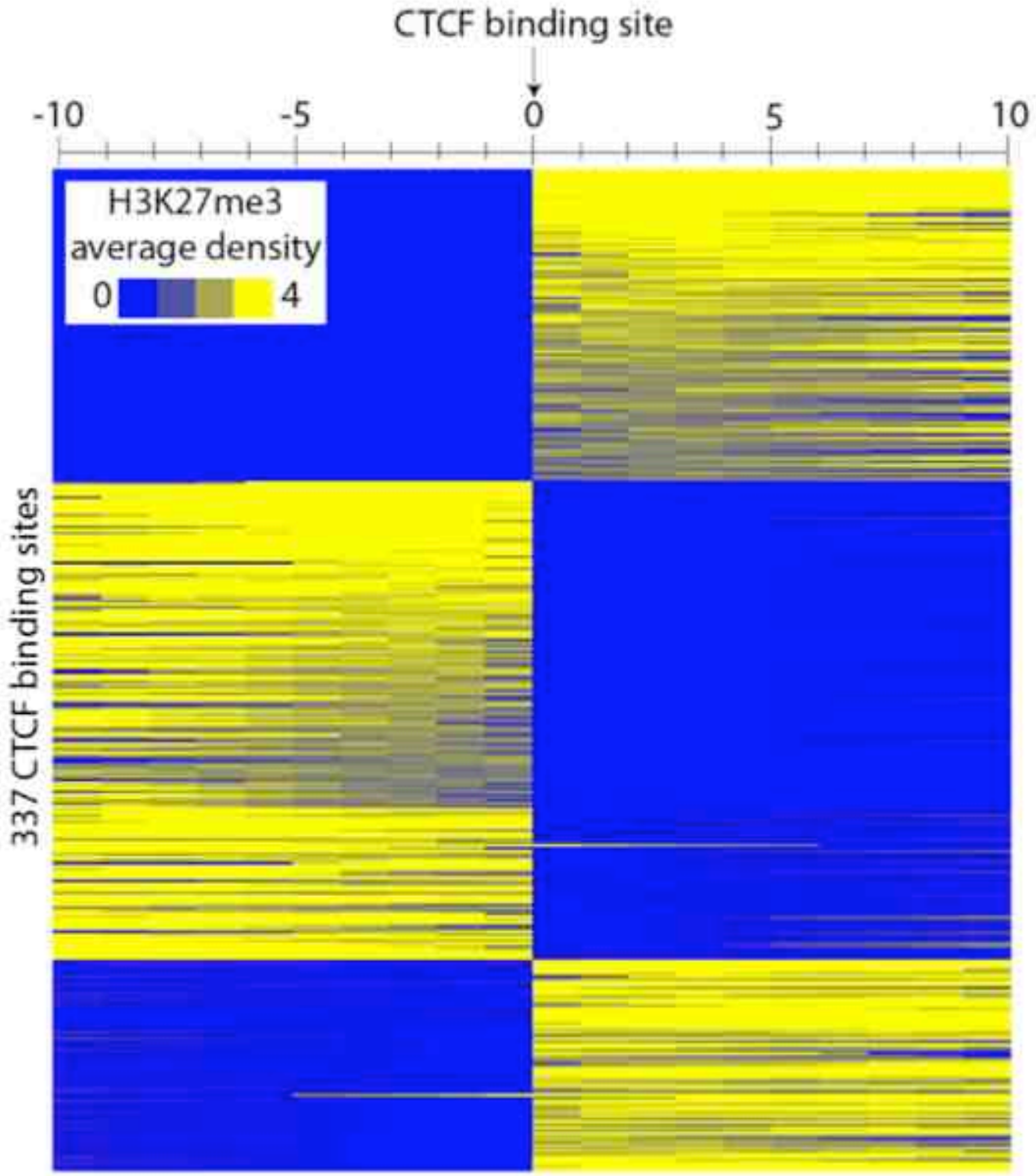
Xie et al, PNAS 104:7145 (2007)

CTCF binding site serves to restrict heterochromatin at the HOXA locus

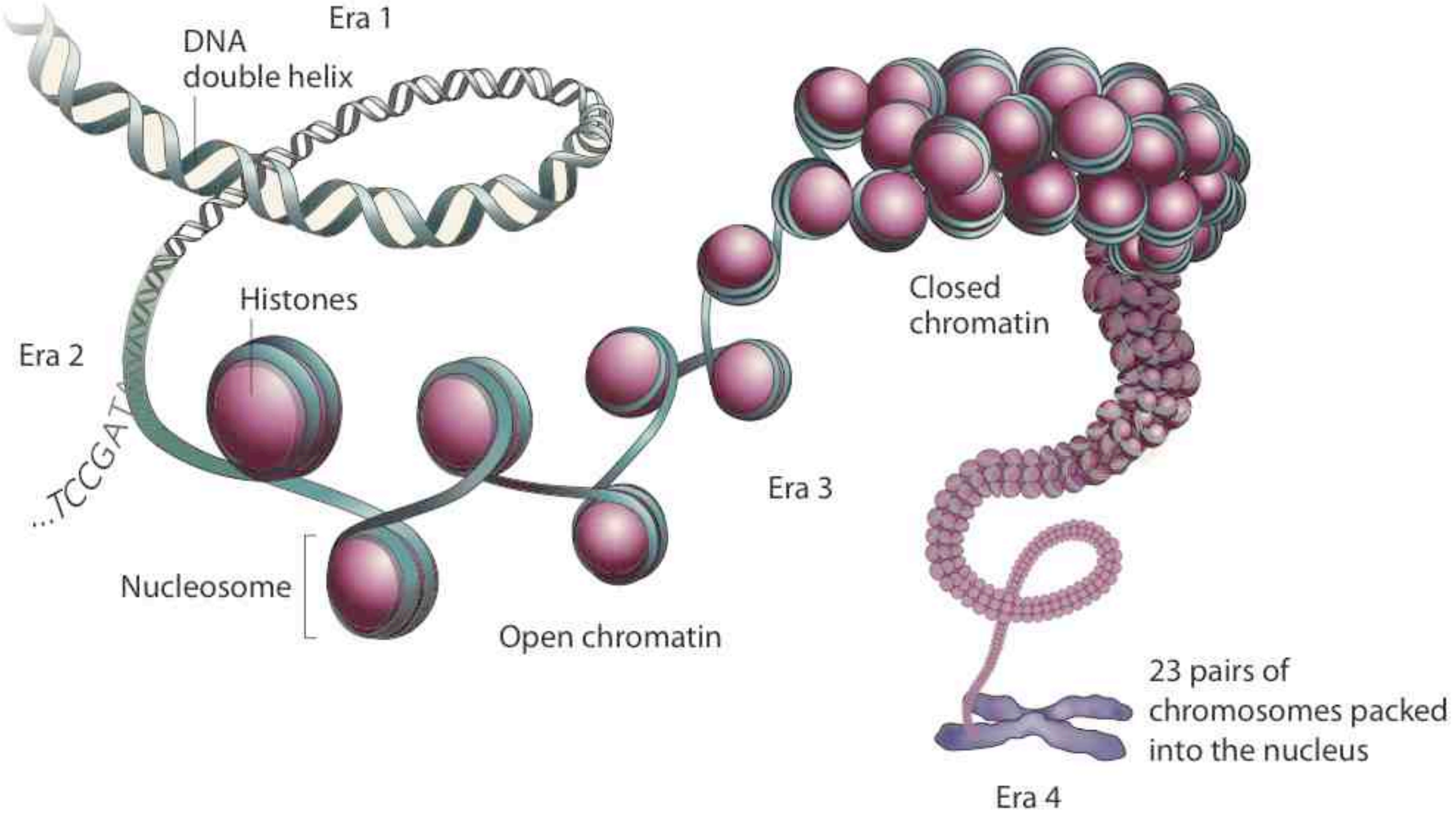


Kim et al Cell 128:1231 (2007); Rinn et al PLoS 2:119 (2007)

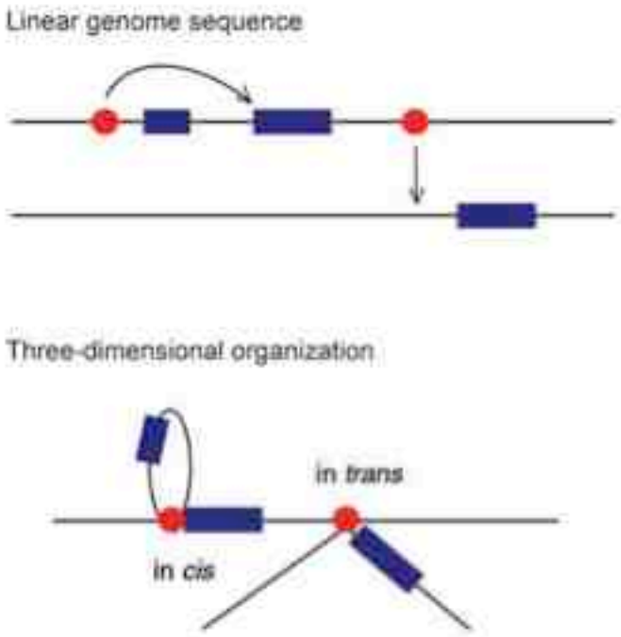
a large number of CTCF-associated chromatin barriers in the human genome

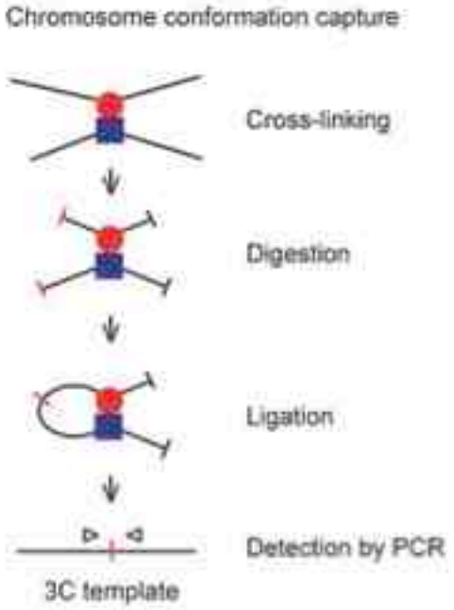


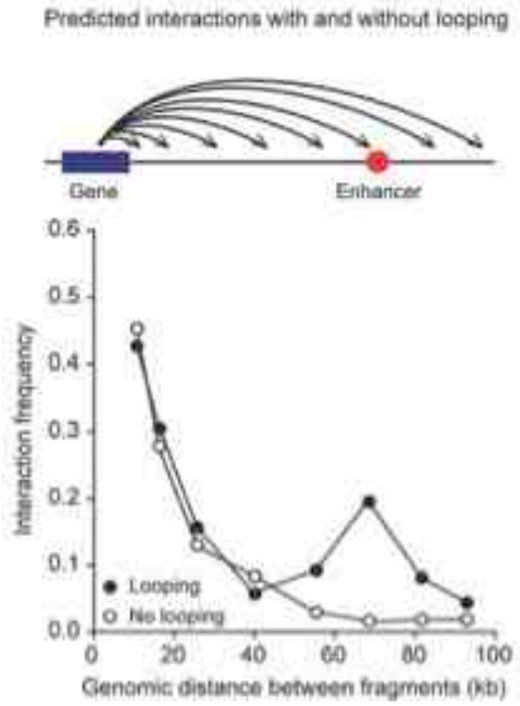
how to analyze higher order chromatin



chromosome conformation capture (3C)







- *immobilize protein-DNA & protein-protein interactions in the nuclei with formaldehyde*
- *restriction digestion; dilute to prevent intermolecular ligation; ligate the compatible ends*
- *PCR to interrogate the possible ligated junctions*

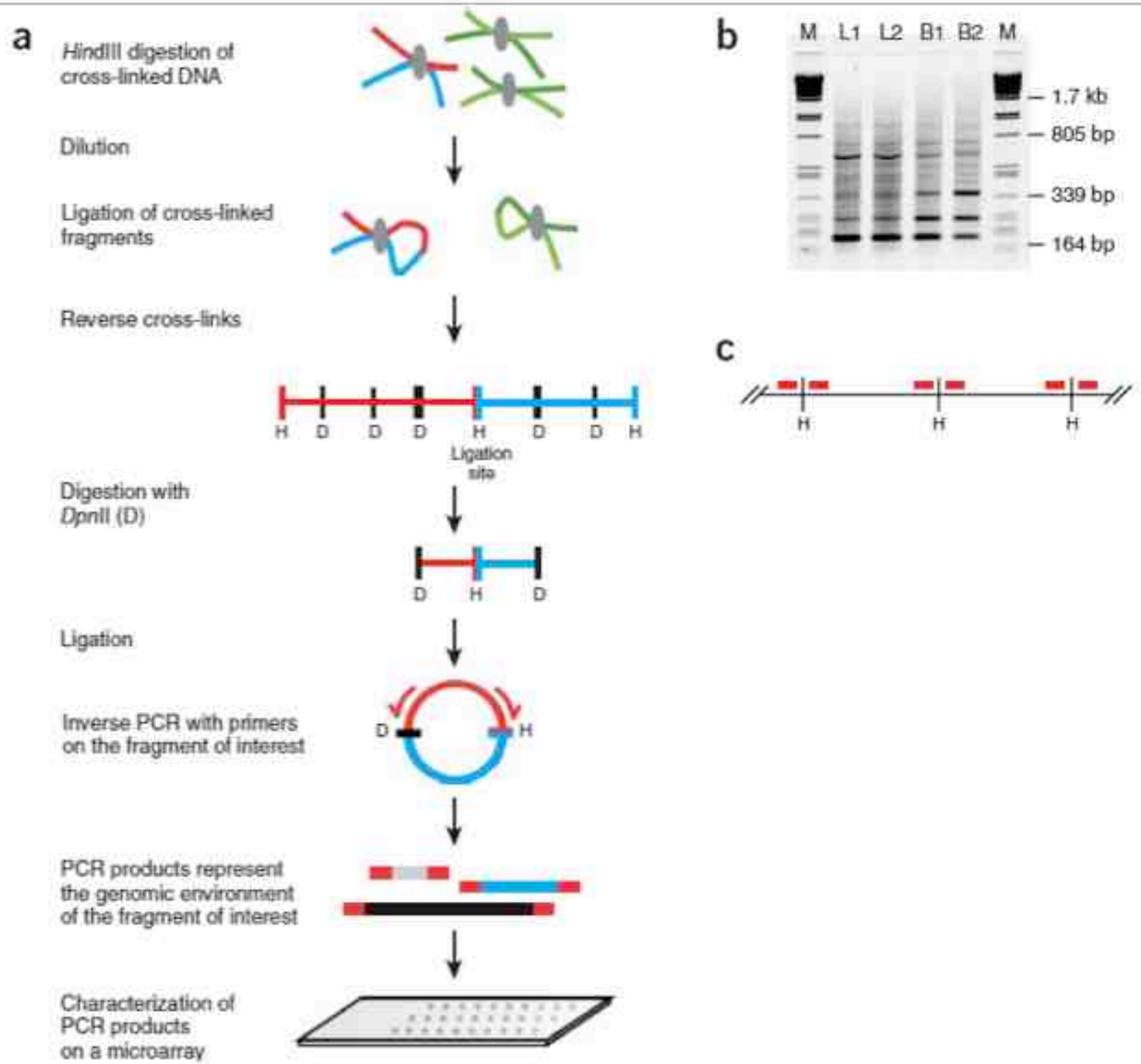
assembling long range interactions for gene regulation

- *enhancers can mediate long range interactions both in cis and in trans*
  - *LCR interacts with the active genes in the β-globin locus (de Laat)*
  - *interchromosomal interaction between IFNG and IL4 loci in a poised chromatin hub for rapid and high level induction of these genes (Flavell)*
  - *Androgen Receptor regulated expression - AR bound enhancer loops to contact the target promoter to facilitate gene expression (Brown)*
  - *Olfactory Receptor gene expression - ensure expression of single OR gene (from >2000); a single enhancer engages one OR gene promoter (both in cis & trans) and excludes others (Axel)*
- *insulators can mediate long range interactions*
  - *H19/IGF2 locus on chr 7 interacts with Wsb1/Nf1 on chr 11; deletion of CTCF binding site at H19/IGF2 abrogate imprinted expression of both loci (Hoffman)*
- *long interactions affect gene expression*
- *mechanisms not known: by exclusion from generally repressive environment or recruitment to transcriptionally permissive environment? how? motorized movements?*

genomic approaches to 3C

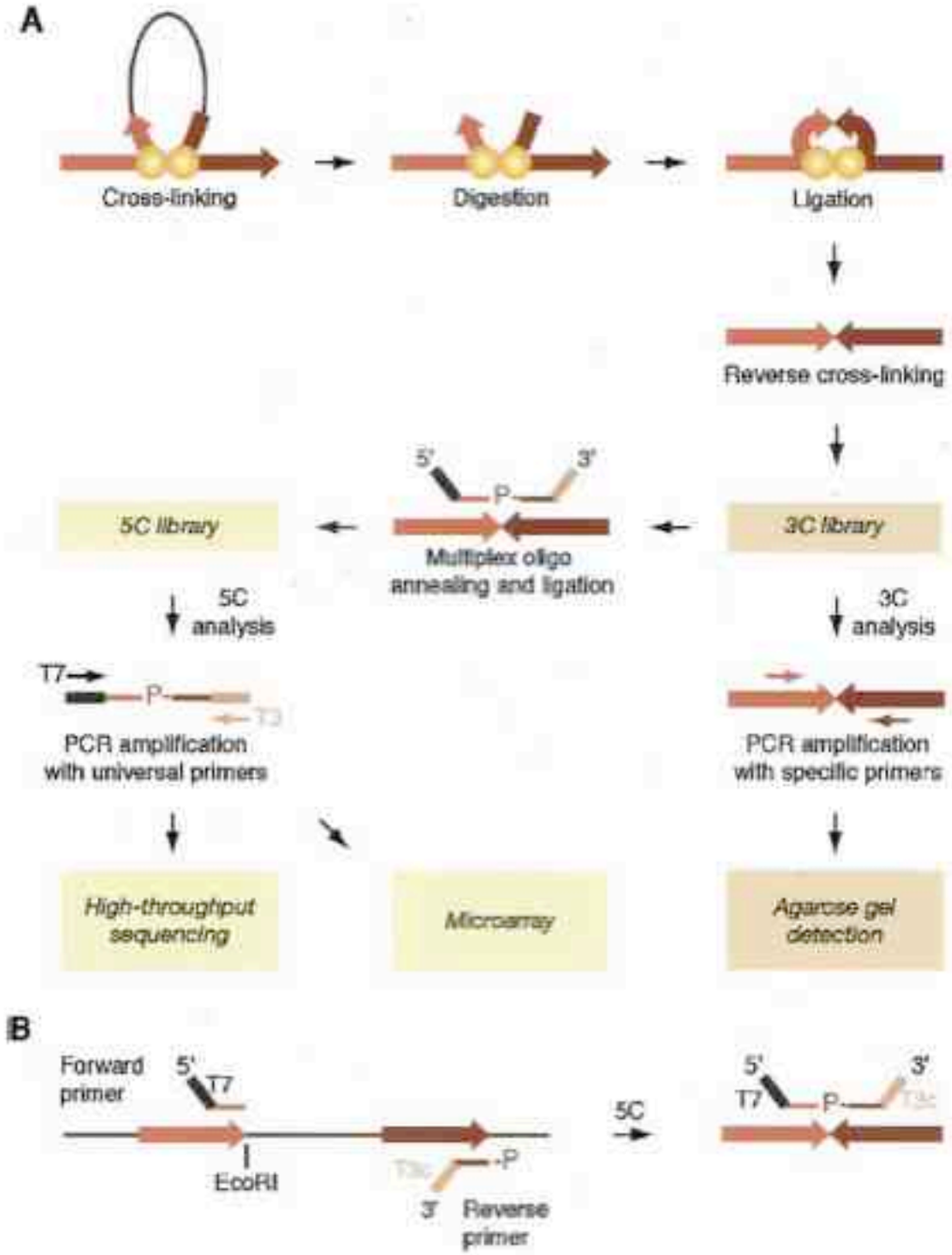
- *4C - circular 3C, microarrays or sequencing to detect interactions from a single site*
- *5C - 3C on microarray*
- *6C - ChIP-3C*

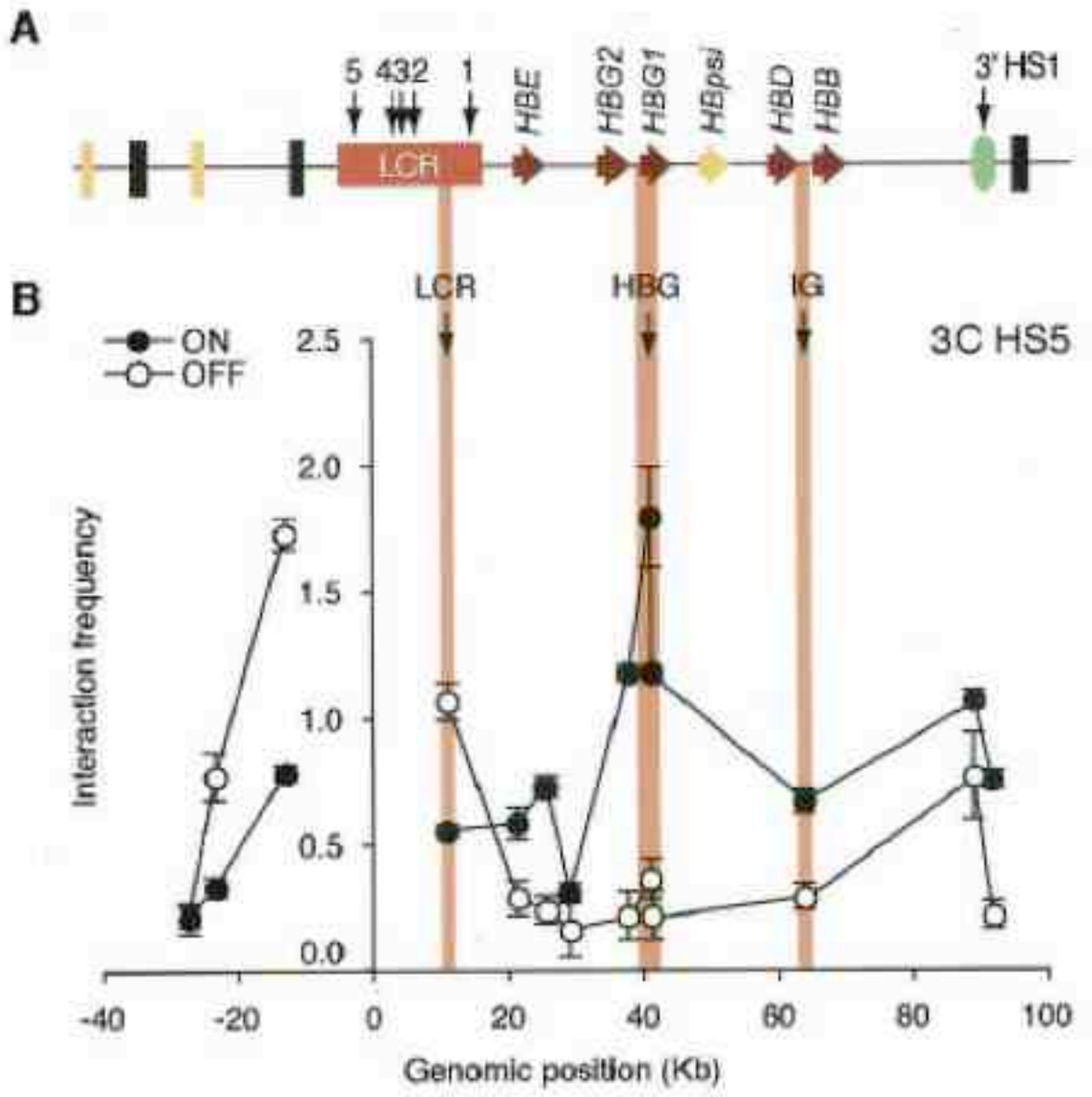
4C - circular 3C



Simonis et al Nat Genet 38:1348 (2006); Gondor et at Nat Protocols 3:303 (2008)

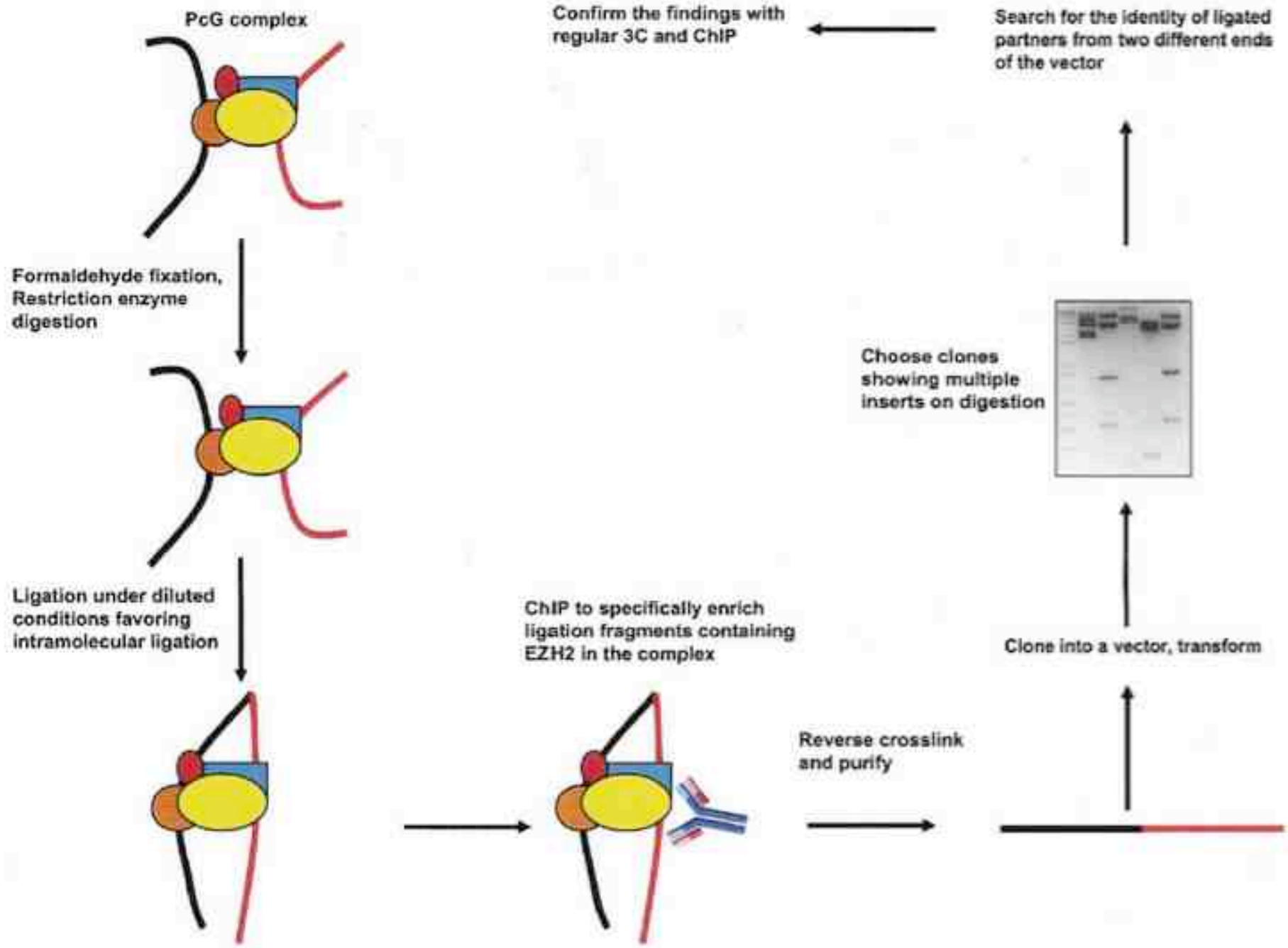
5C





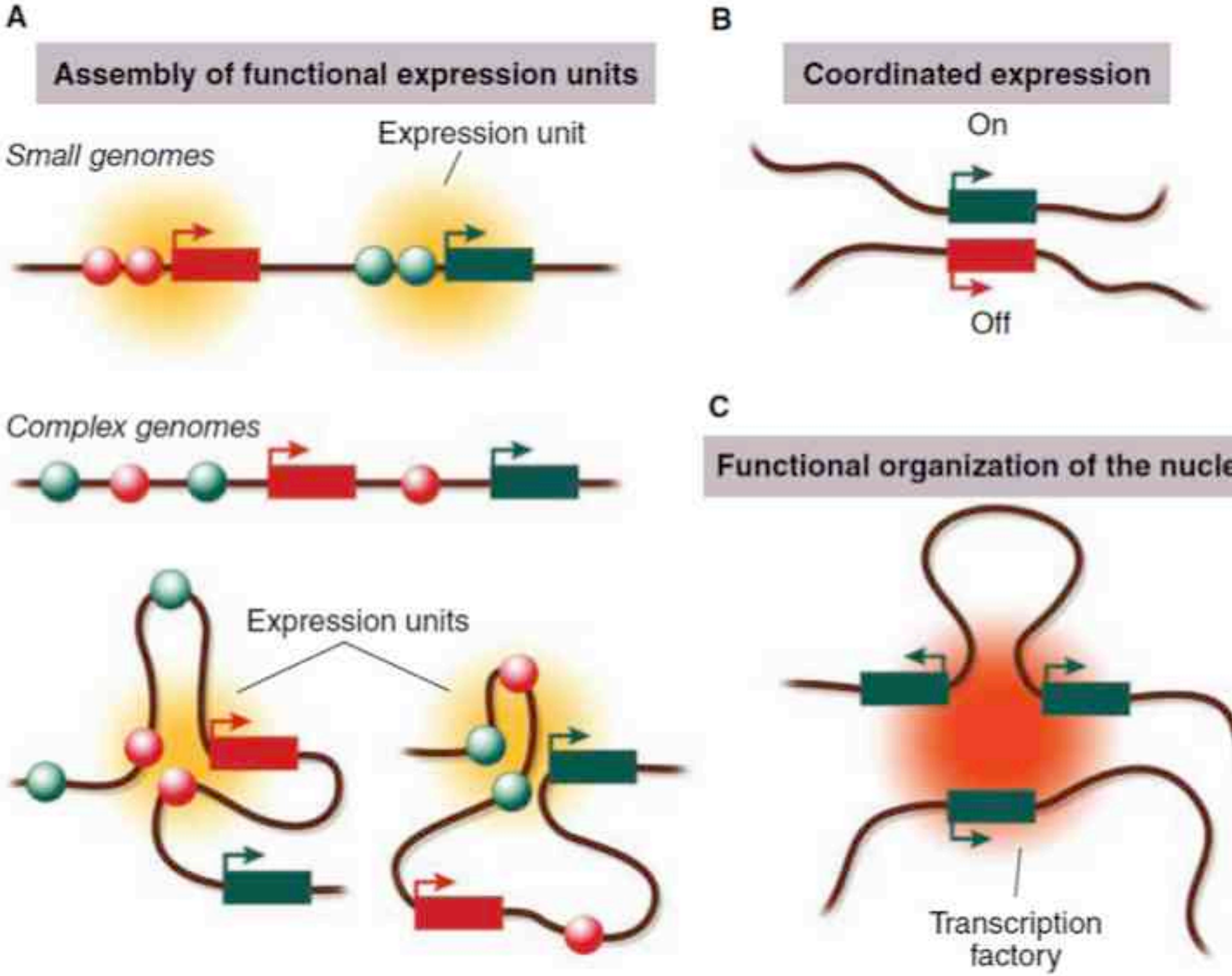
Dostie et al Genome Research 16:1299 (2006)

6C - ChIP-3C



Tiwari et al Genome Research 18:1171 (2008)

3D organization of the genome has functional consequences



Dekker Science 319:1793 (2008)

summary

- *crosslinking approaches have been critical defining regulatory events that occur in the genome*
- *distinct classes of noncoding elements can be determined from distinct patterns of histone modification - promoters from enhancers*
- *insulators are used to define regulatory and chromatin domains in the genome*
- *genome is organized into higher order structures containing loops and long range interactions that influence gene activity and a number of assays have been developed to identify these interactions*