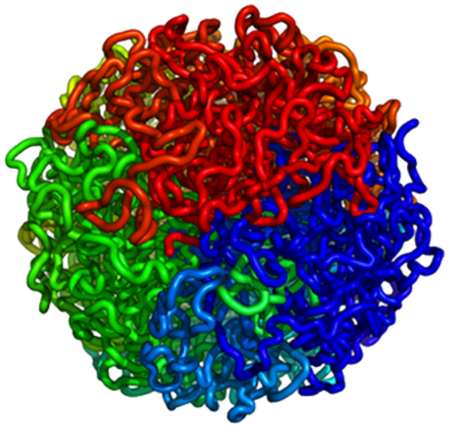
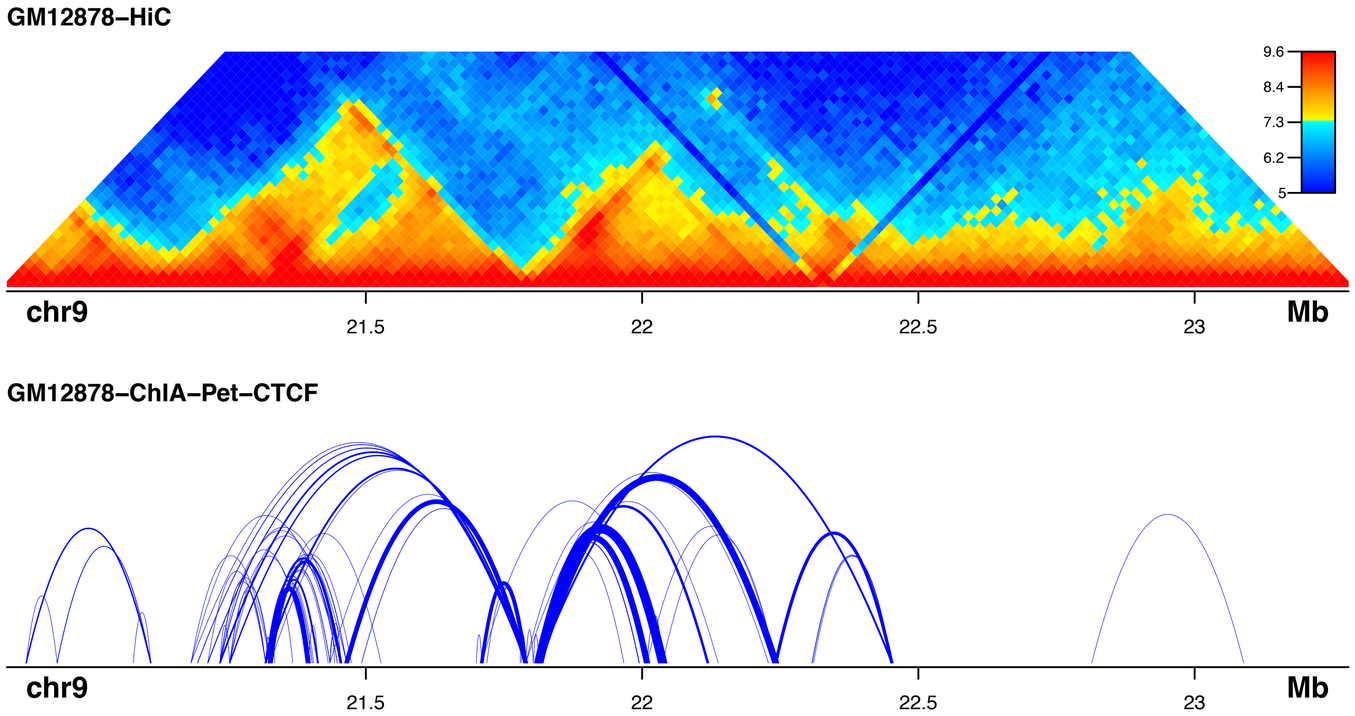
Analyzing 3-D genome organization

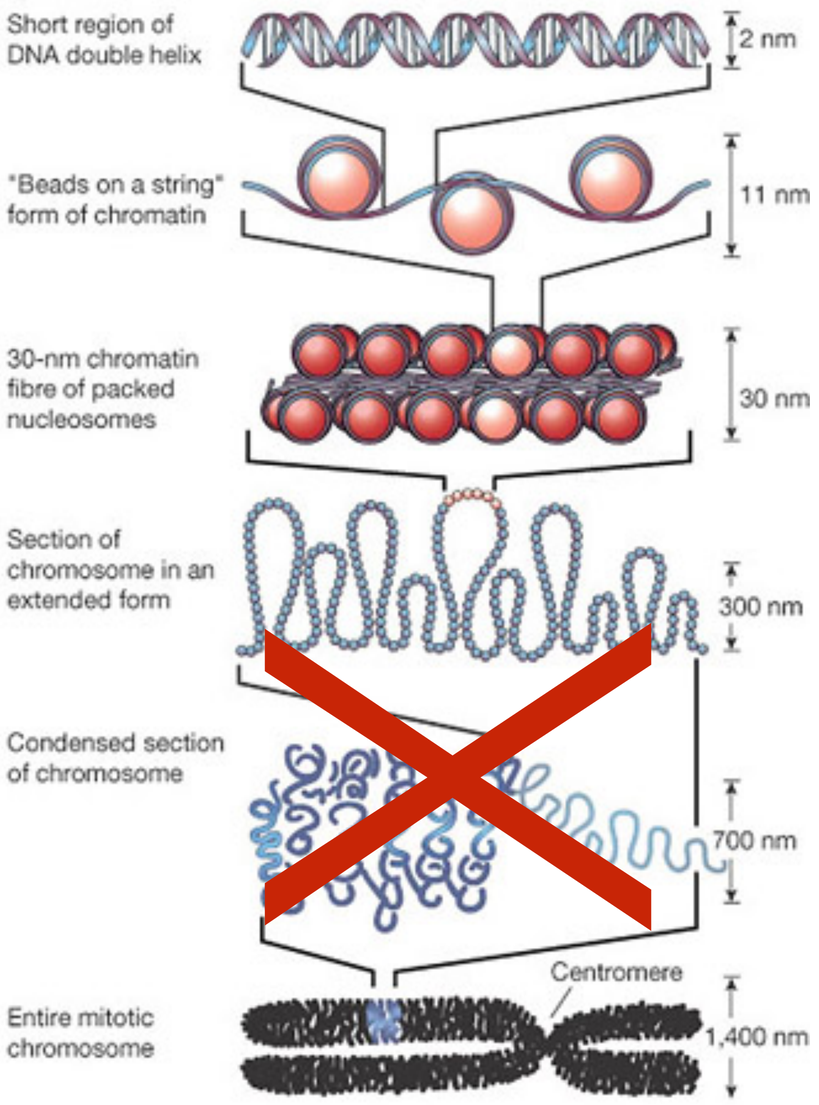
Martin Aryee

Massachusetts General Hospital
Harvard T.H. Chan School of Public Health
Broad Institute



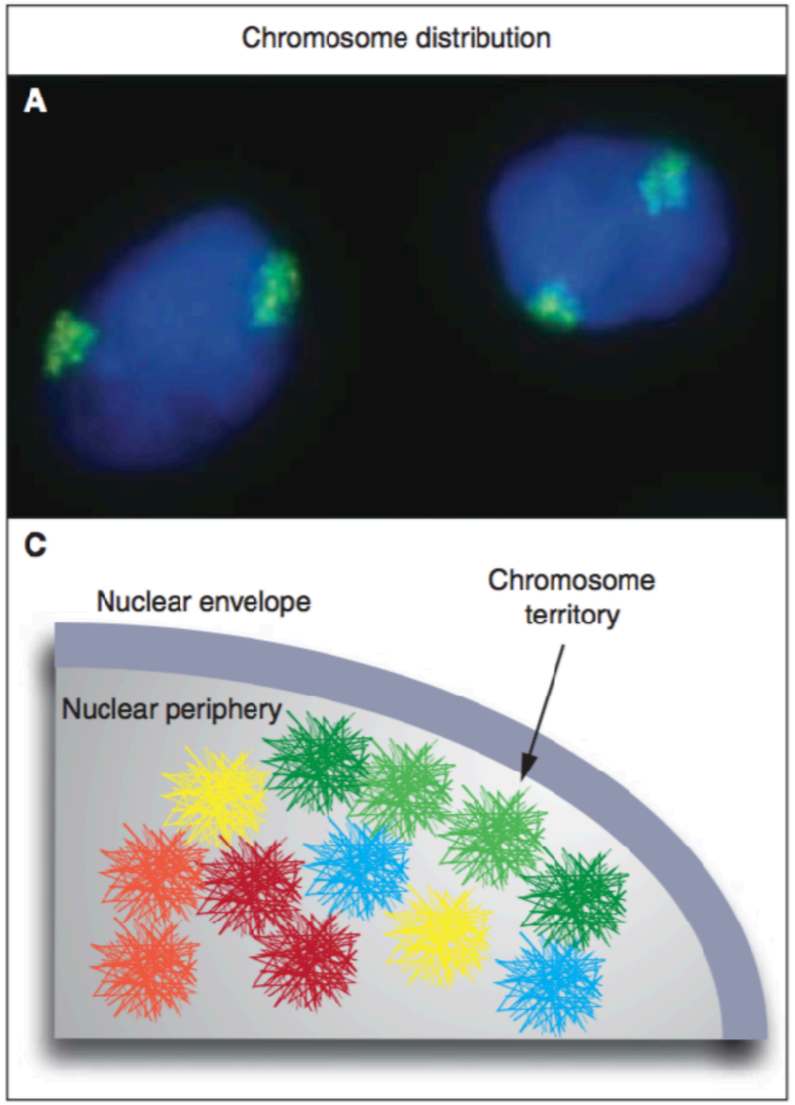
http://www.aidenlab.org





Felsenfeld and Groudine. Nature, 2003

Chromosomes exhibit non-random spatial organization

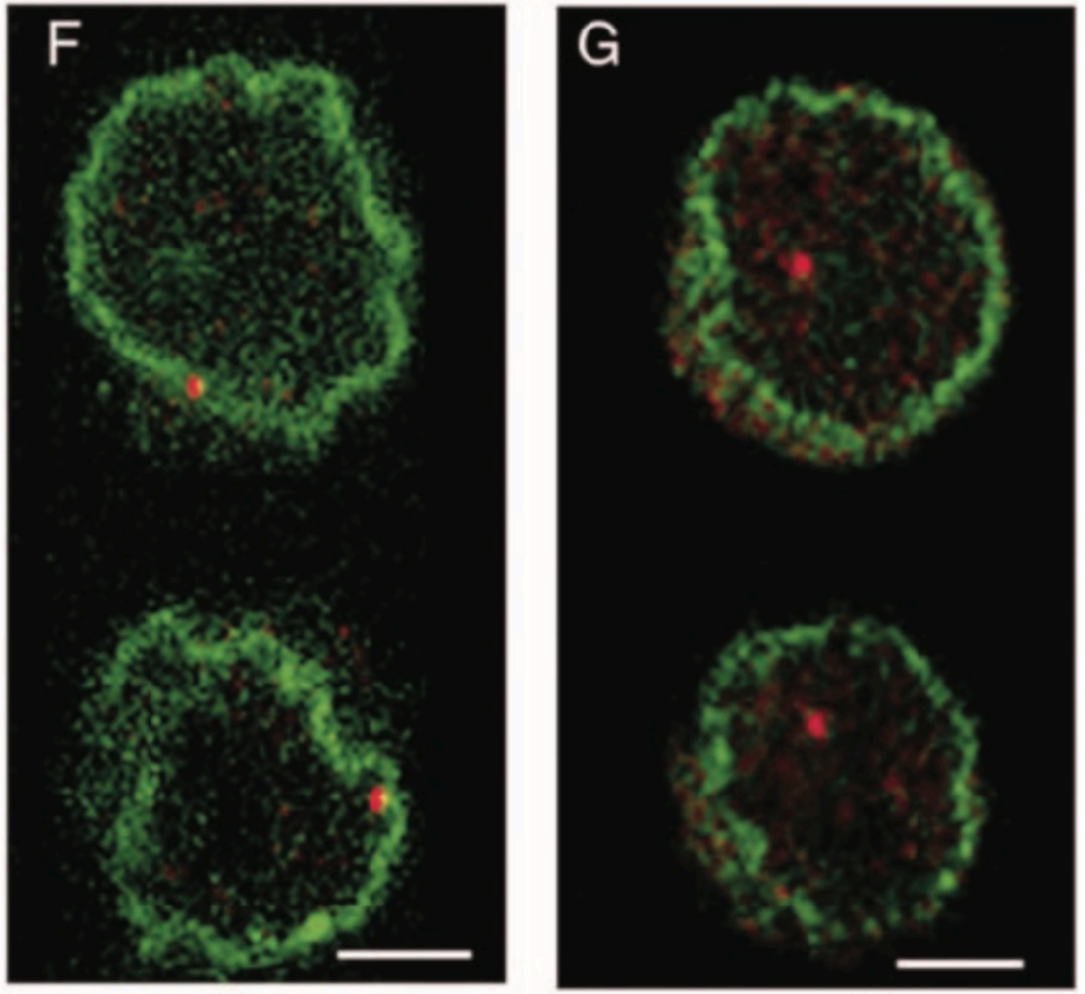


Dekker and Misteli, Cold Spring Harb Perspect Biol (2015)

Gene repositioning is associated with gene activation

Silenced gene

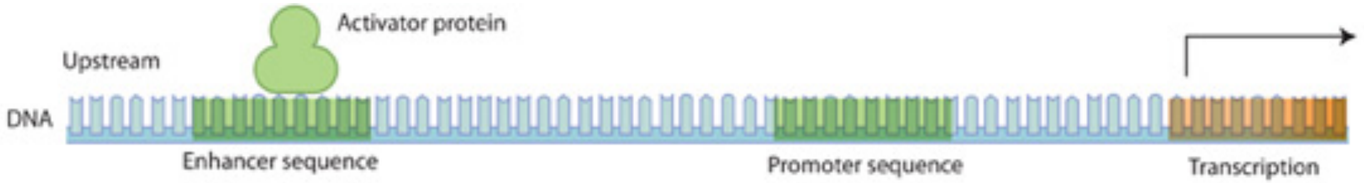
Active, expressed gene



Green = anti-lamin B
Red = IgH locus

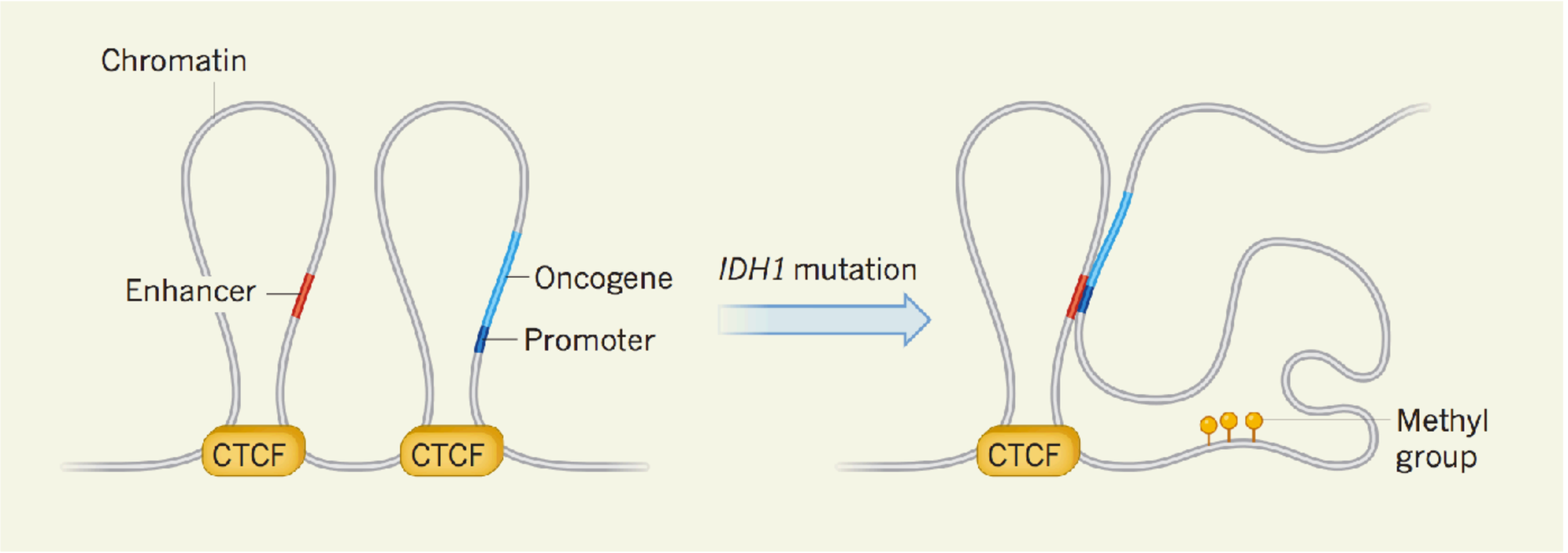
Kosak et al., Science (2002)

Enhancer - promoter looping



http://www.nature.com/scitable/topicpage/gene-expression-14121669

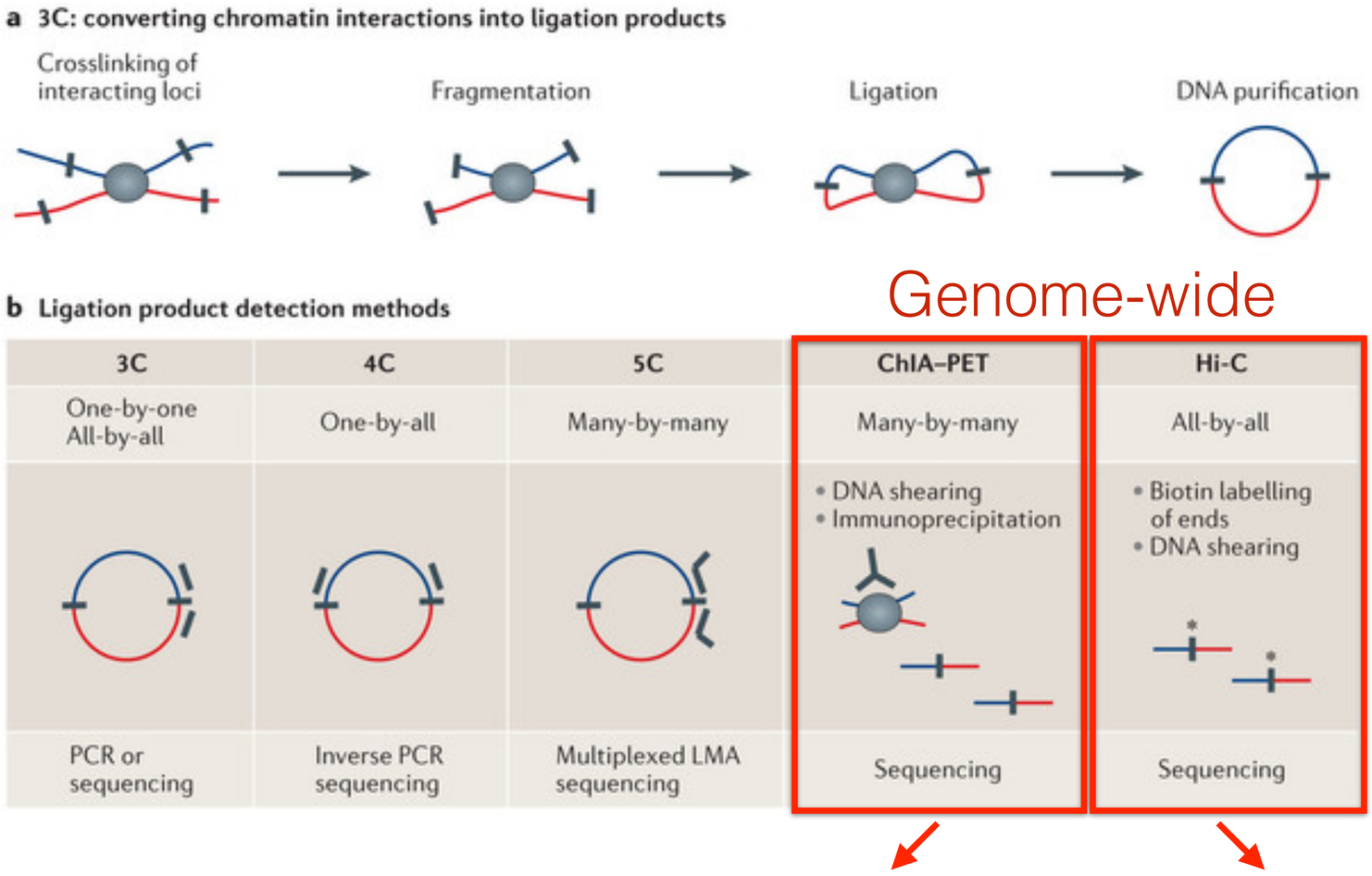
Disruption of genome topology in cancer



Flavahan and Drier et al., Nature (2015)
Image from Grimmer & Costello, Nature (2015)

“Structural” vs. “Functional” loops

Genome topology assays



Genome-wide

~200M-500M Reads

~1-5 Billion Reads

Nature Reviews | Genetics

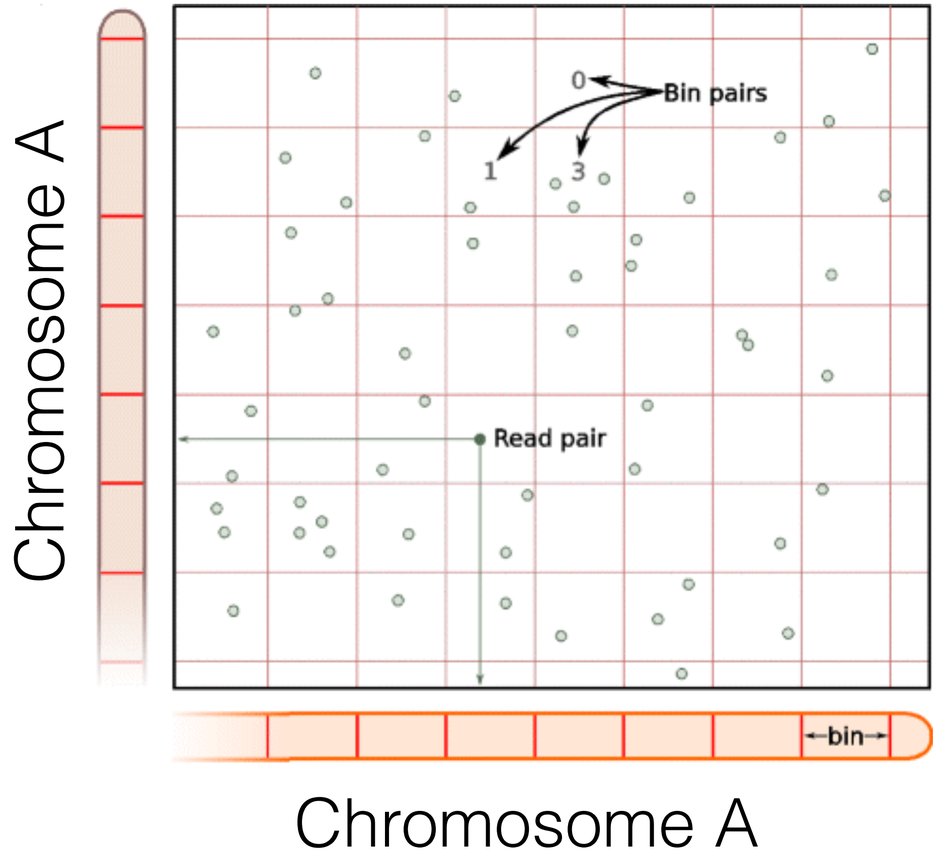
QC: Hi-C

- Key metric for both Hi-C and ChIA-PET:
  Fraction of paired-end read (“PETs”) supporting long-range interactions

https://www.bioinformatics.babraham.ac.uk/projects/hicup/scripts_description/

- Poor Hi-C library: <20% long-range (>20kb) pairs
- Good Hi-C library: ~40% long-range (>20kb) pairs
- A shallow sequencing run (~5M reads) is sufficient to assess library quality

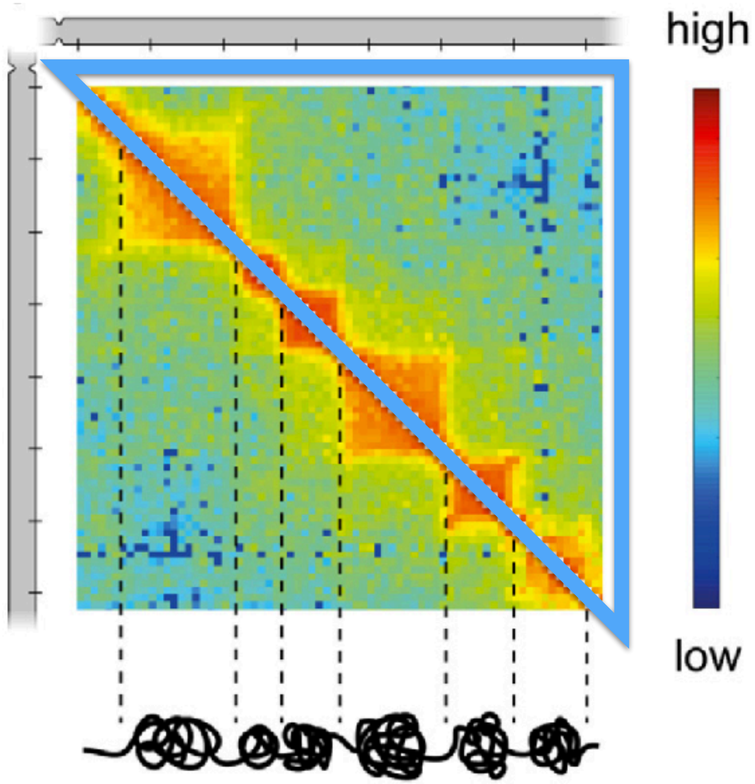
HiC data:
From read pairs to an interaction matrix

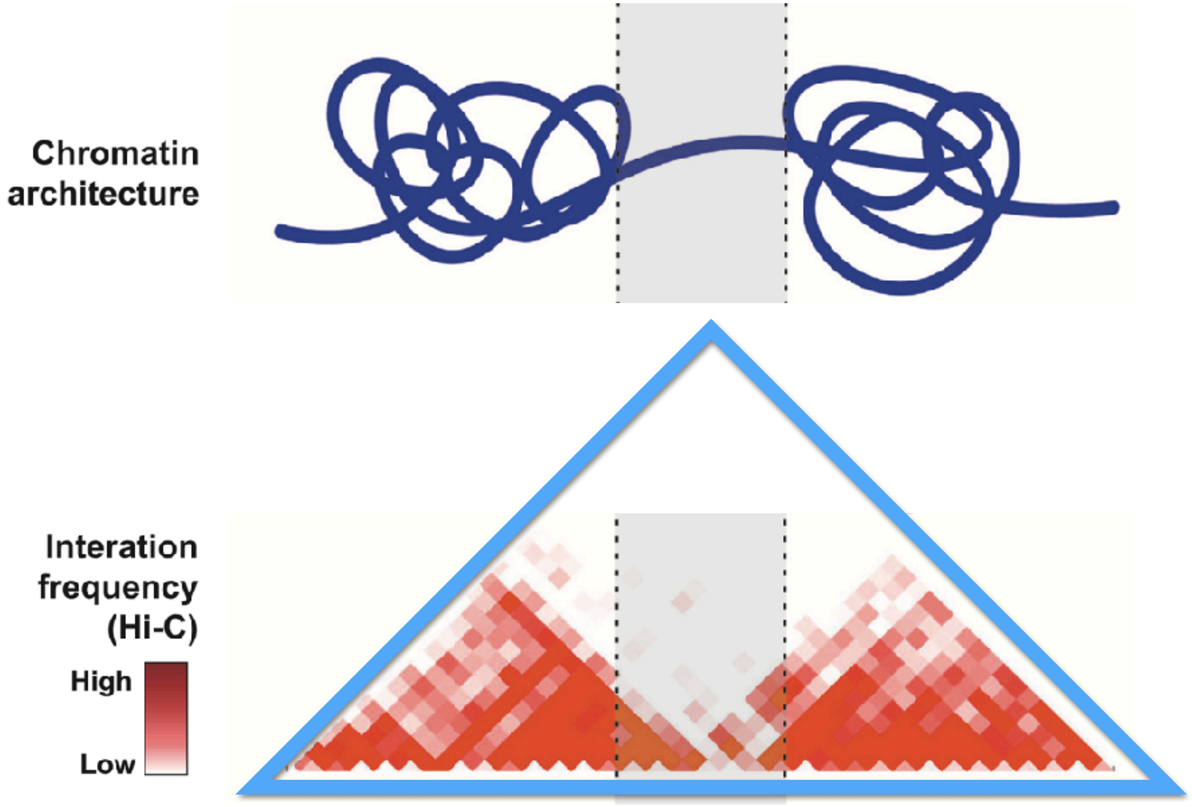


Resolution (bin size)	Number of matrix cells
1 Mbase	~10M
10 kbase	~100B

Adapted from Lun et al., BMC Bioinformatics, 2015, Volume 16, Number 1, Page 1

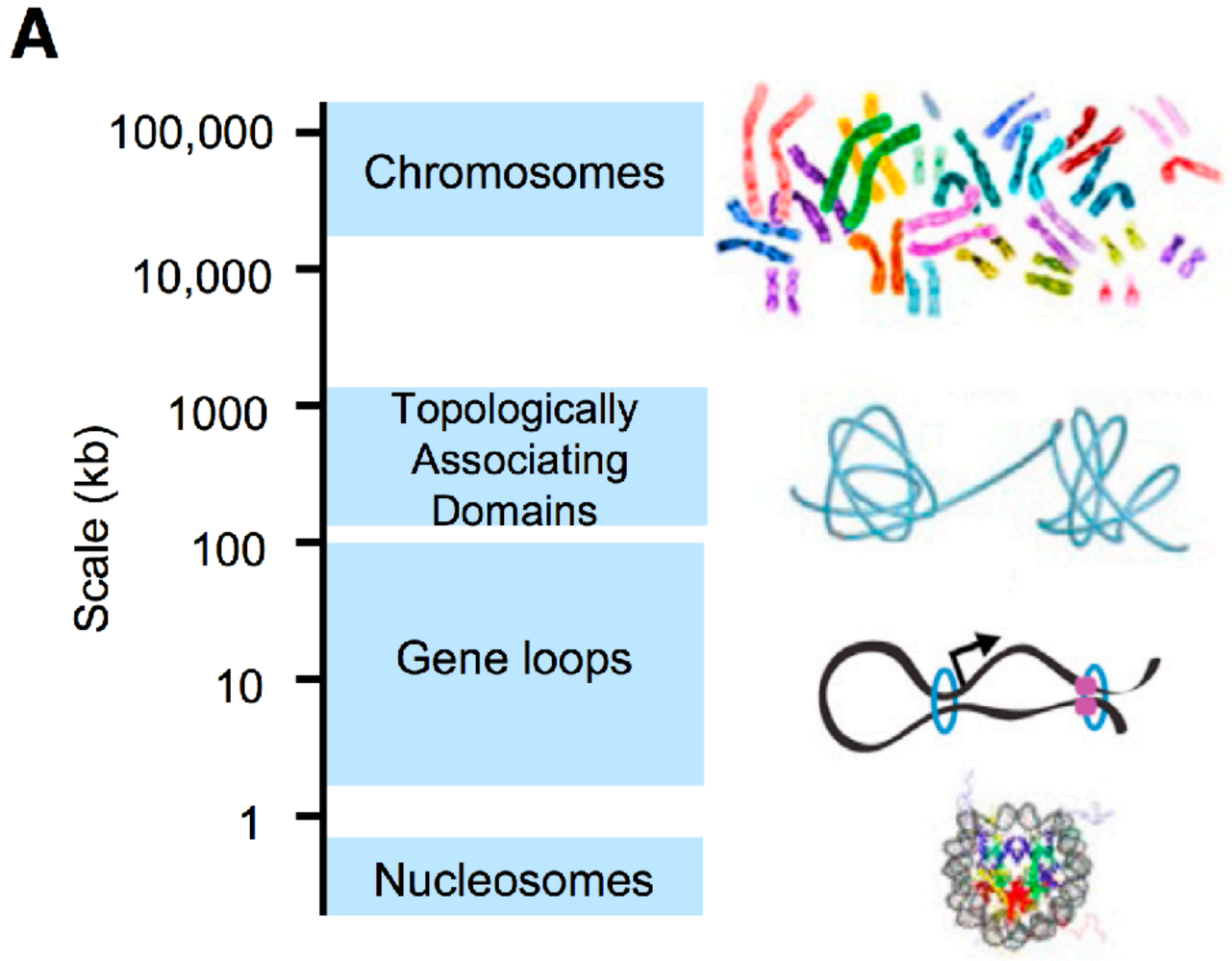
HiC data:
From read pairs to an interaction matrix





Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin Domains: The Unit of Chromosome Organization. Molecular Cell 62, 668–680 (2016).

3D genome organization at multiple scales

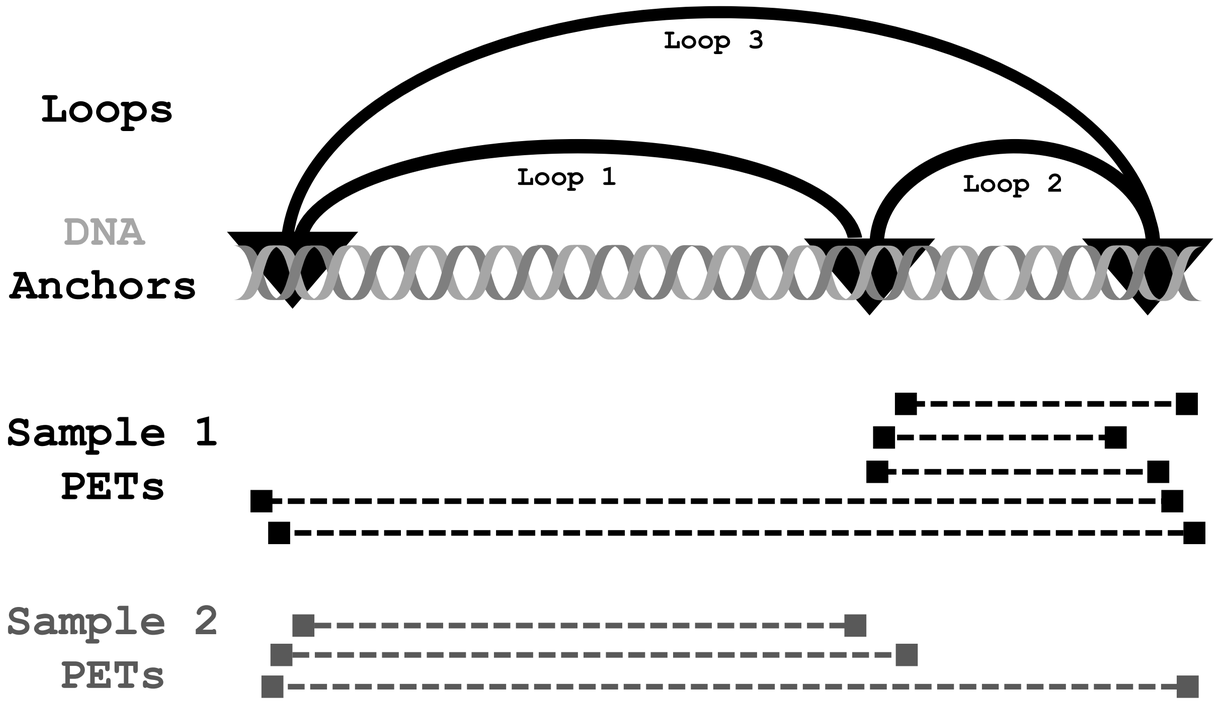


1. Dowen, J. M. et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. - PubMed - NCBI. Cell 159, 374–387 (2014).

HiChIP / ChIA-PET data:
From read pairs to an interaction matrix

Processing steps:

1. Identify loop anchors (i.e. ChIP peaks)
2. Anchor pairs joined by PETs are putative loops
3. Count PETs (paired-end reads) per putative loop
4. Determine loop significance (i.e. is the number of PETs higher than the background?)

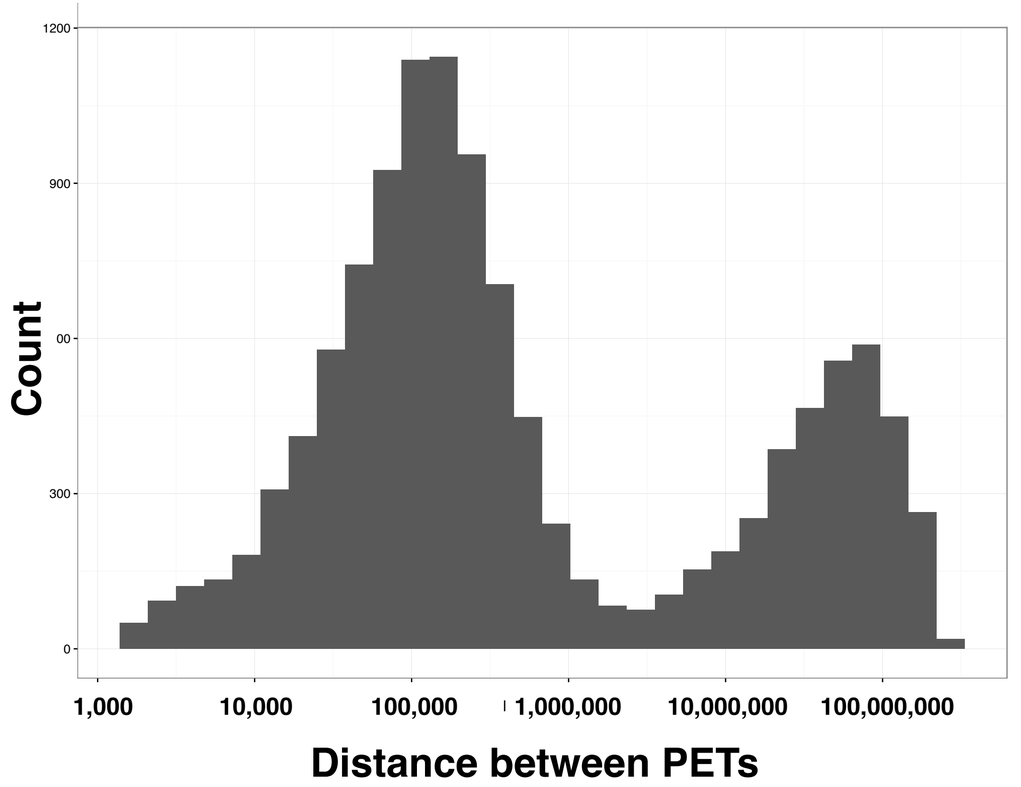


Counts Matrix

	Sample 1	Sample 2
Loop 1	0	2
Loop 2	3	0
Loop 3	2	1

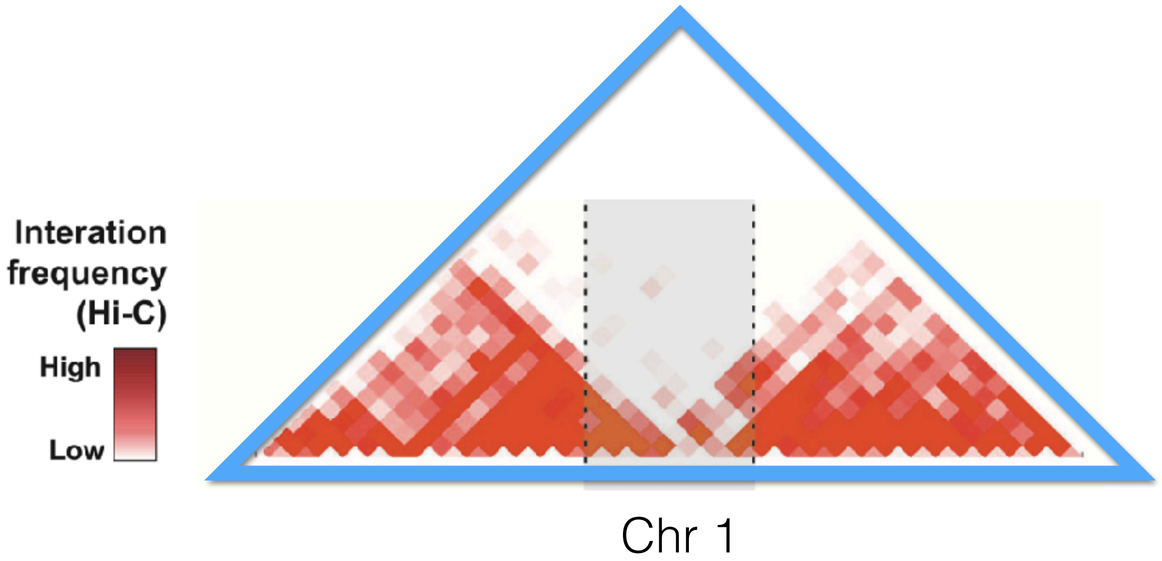
PET = Paired End Tag

How long are chromatin loops?



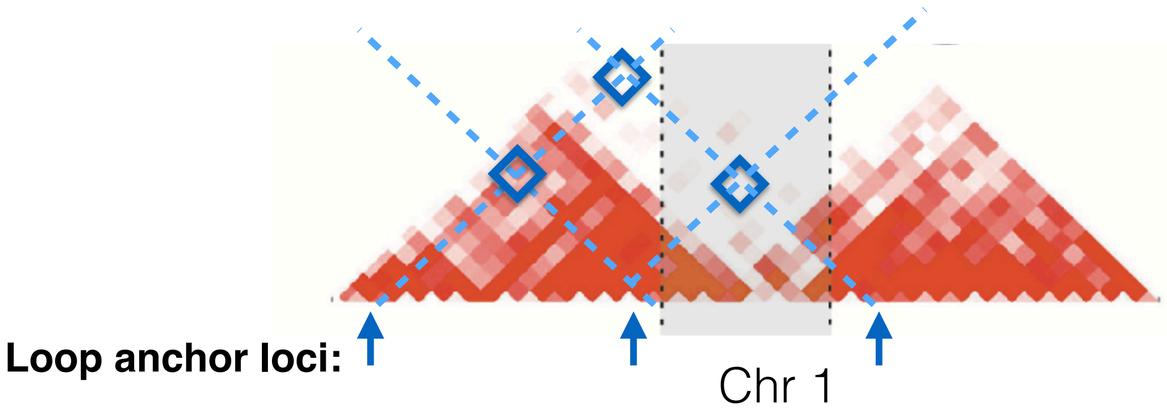
Hi-C

All pairwise interactions



ChIA-PET

Pairwise interactions between loop anchor loci



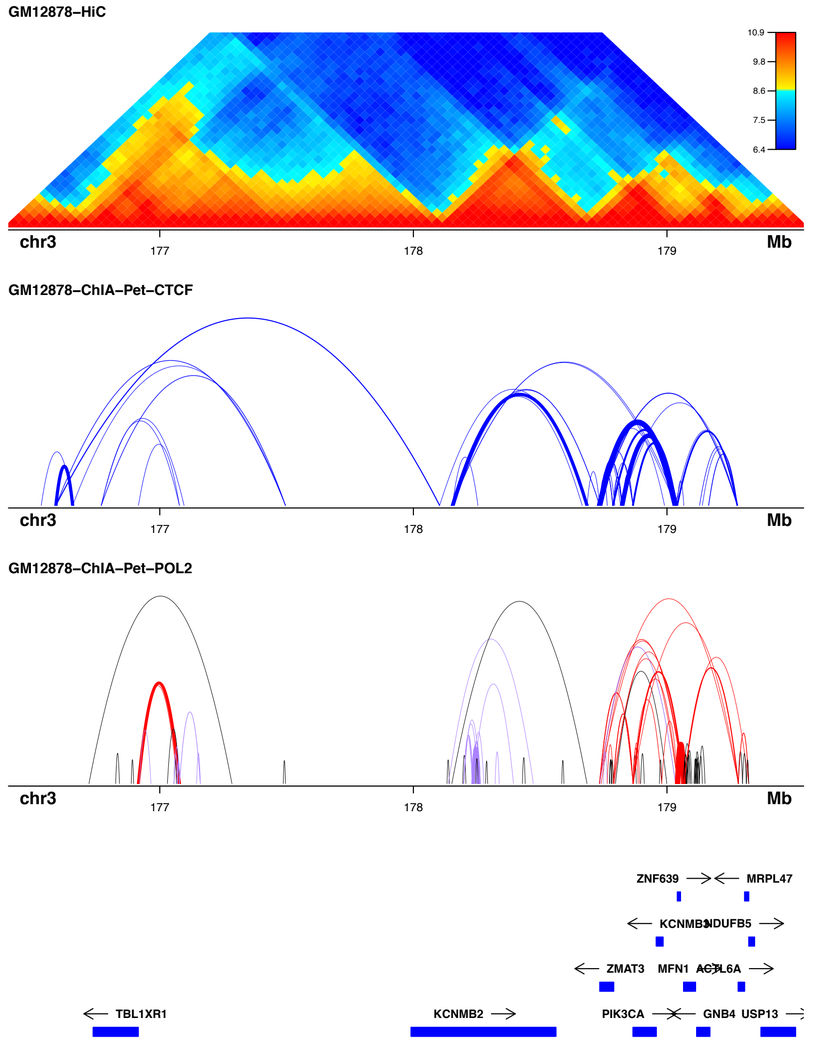
Adapted from: Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin Domains: The Unit of Chromosome Organization. Molecular Cell 62, 668–680 (2016).

Hi-C and ChIA-PET are complementary

HiC

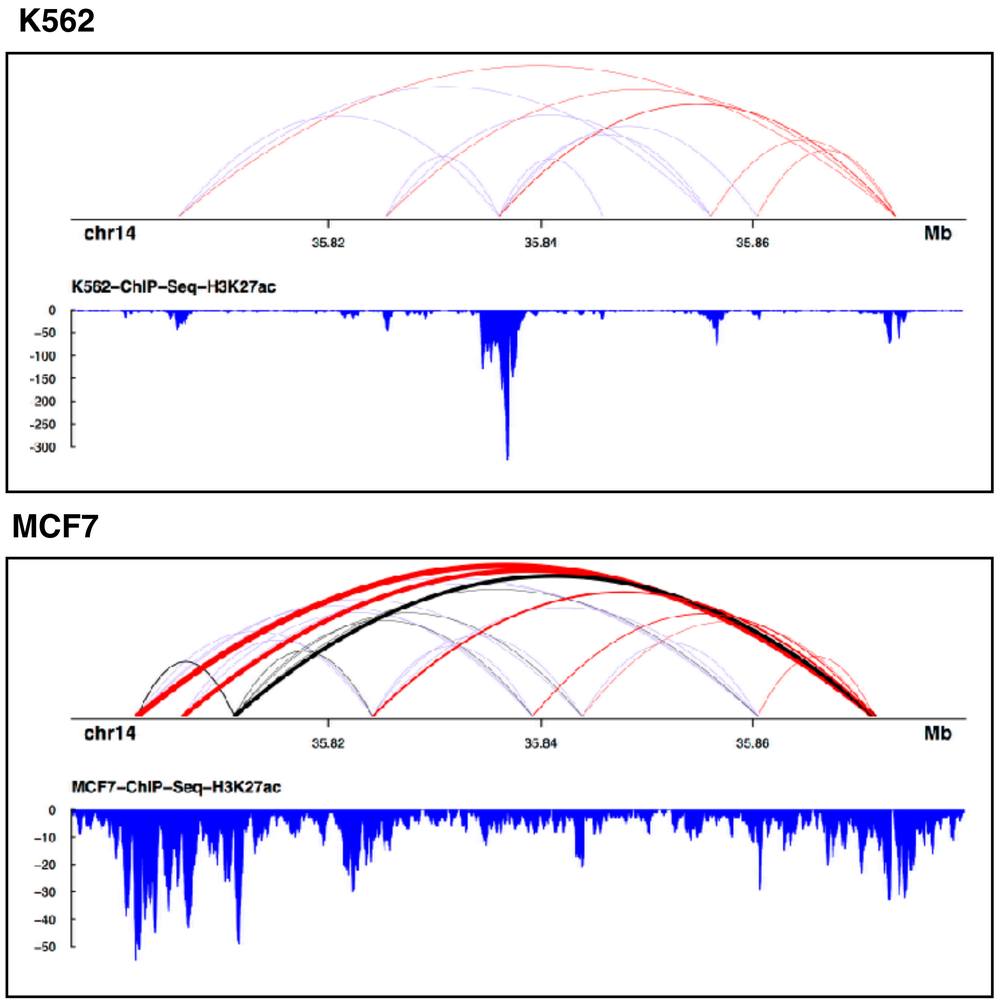
ChIA-PET: CTCF loops

ChIA-PET: POL2 loops



http://dnalandscaper.aryeelab.org

Differential topology analysis: POL2 ChIA-PET



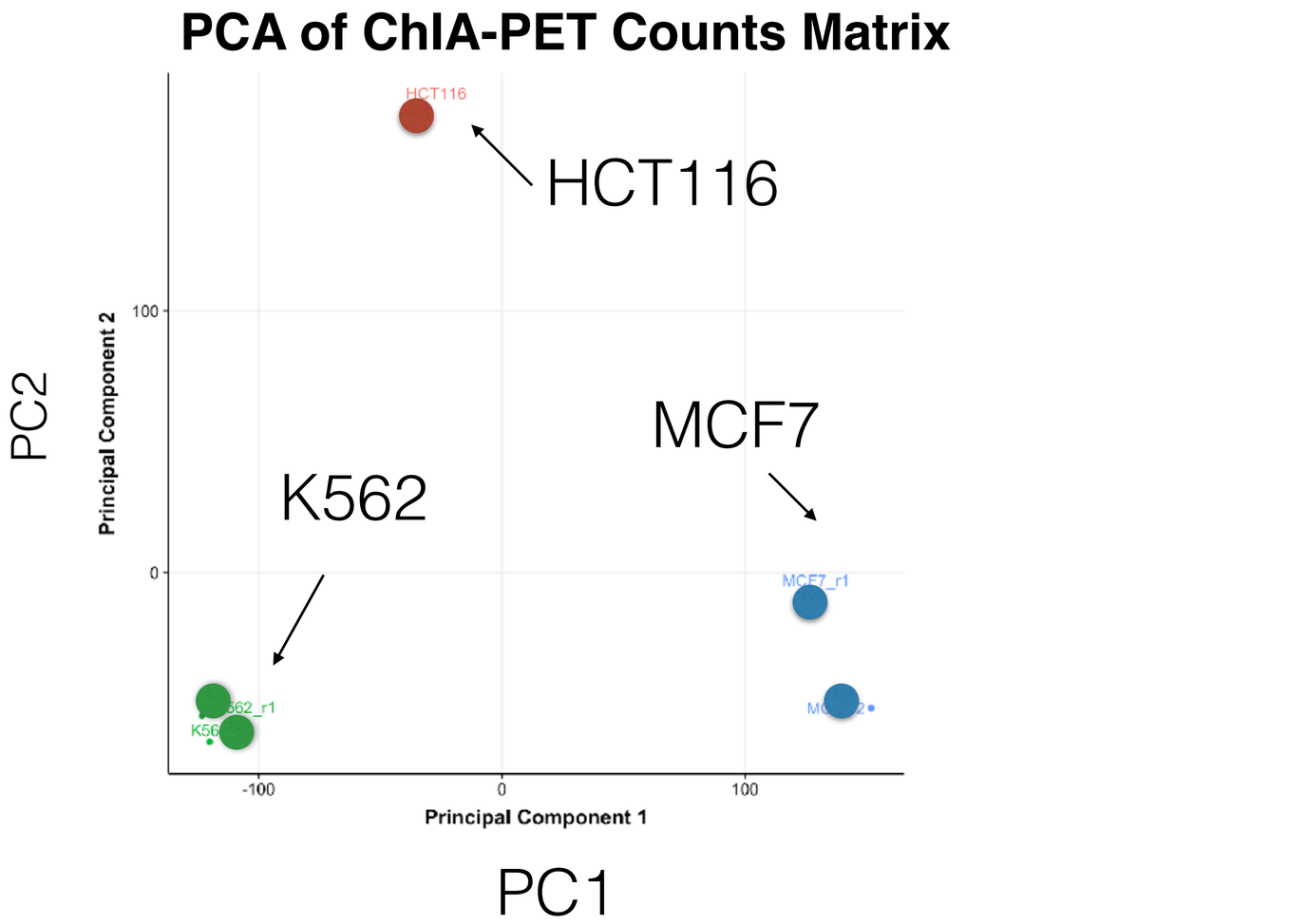
Legend

enhancer - promoter

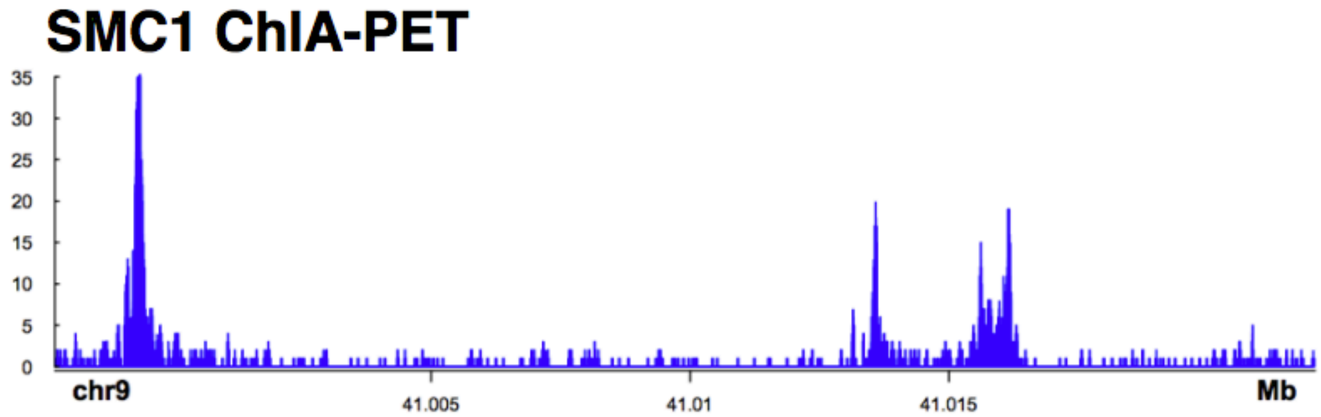
CTCF - CTCF

http://dnalandscaper.aryeelab.org

Pol2 ChIA-PET loops



ENCODE Project Data

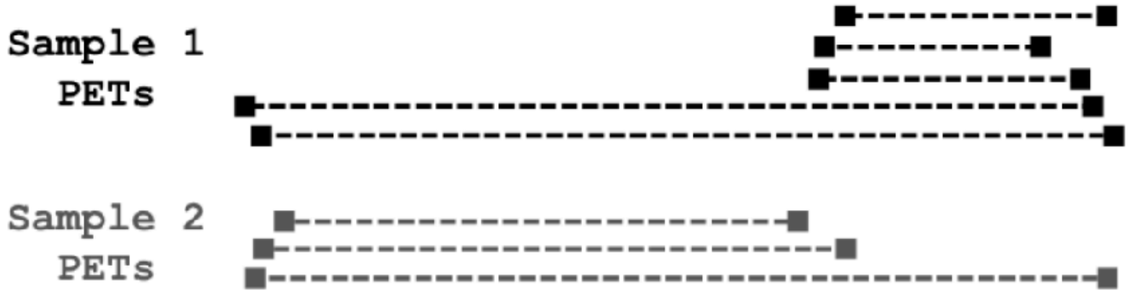


Sample 1

PEAK PEAK PEAK

Sample 2

PEAK PEAK



Need to define a common anchor set:

<PEAK> <PEAK> <PEAK>

Differential loops: diffloop

```
   chr start  end
[1] 1   10000 11000
[2] 1   12000 13000
[3] 1   17000 17600
[4] 1   21500 22000
          …
```

Anchors

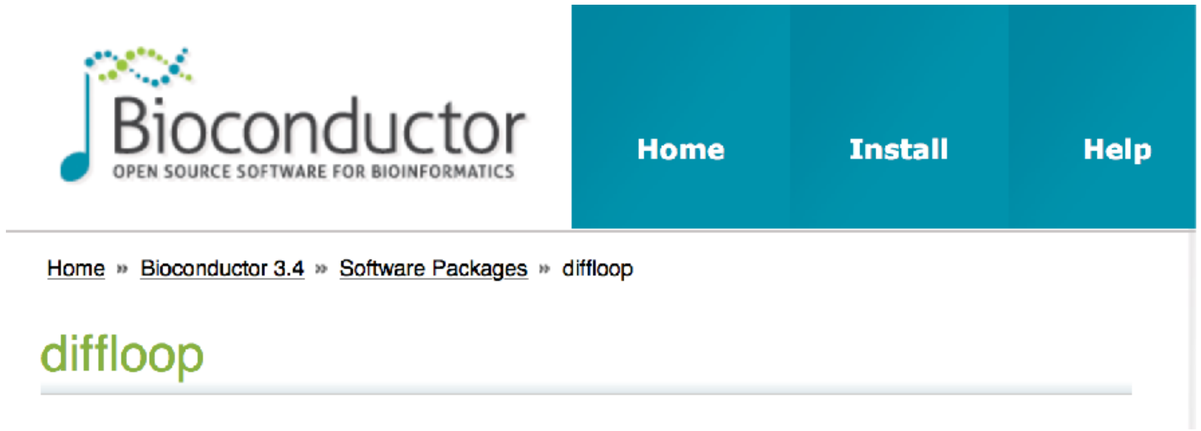
```
     left  right
[1]     1      2
[2]     1      3
[3]     2      3
[4]     2      4
          …
```

Loops

```
     k1  k2  m1  m2
[1]  4   3   1   0
[2]  1   2   6   2
[3]  2   0   7   8
[4]  0   1   1   0
          …
```

Loop 1
Loop 2
. . .

PET counts



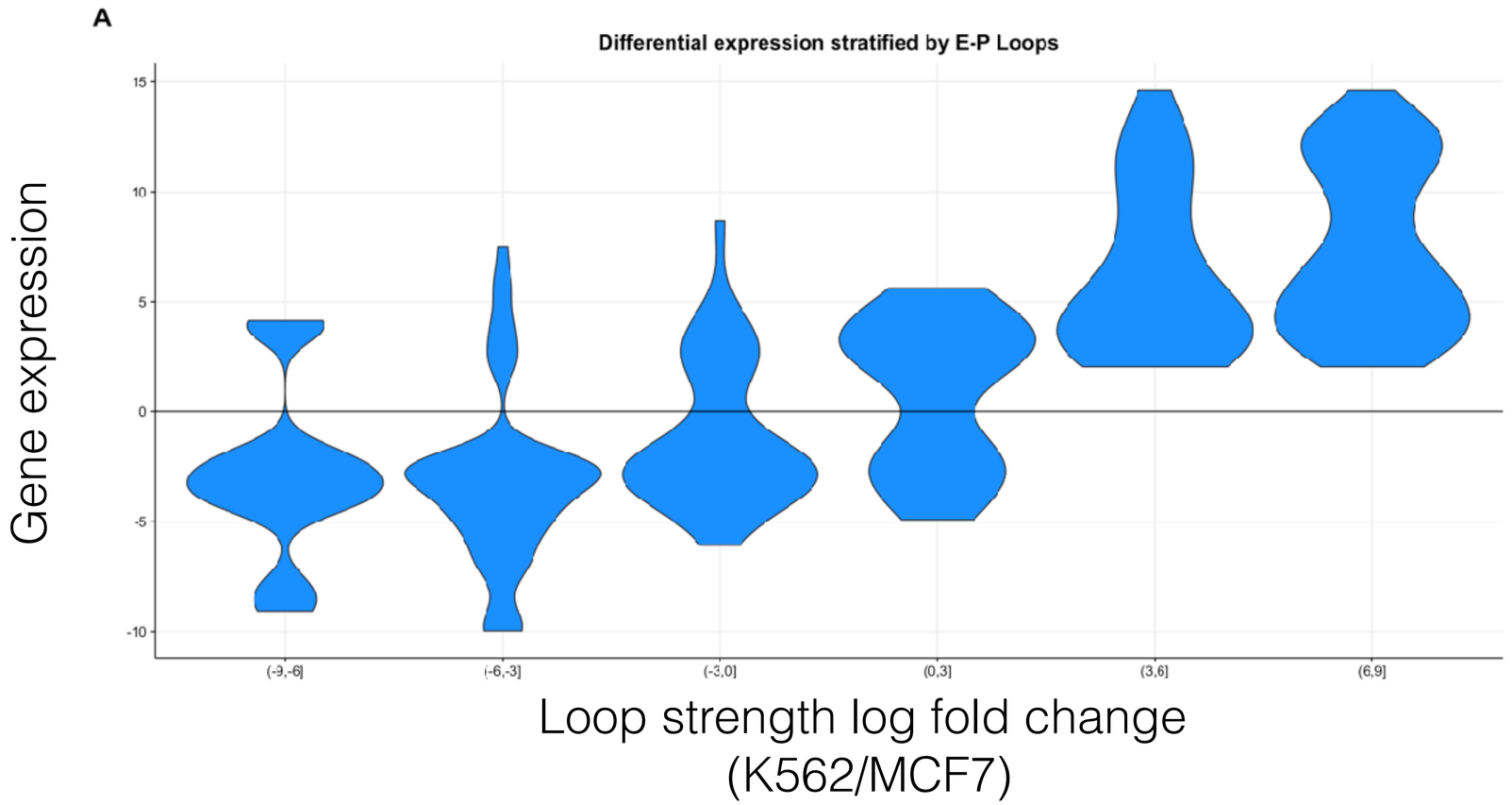
PET = paired end tag

http://bioconductor.org/packages/diffloop/

Genome-wide analysis of differential looping

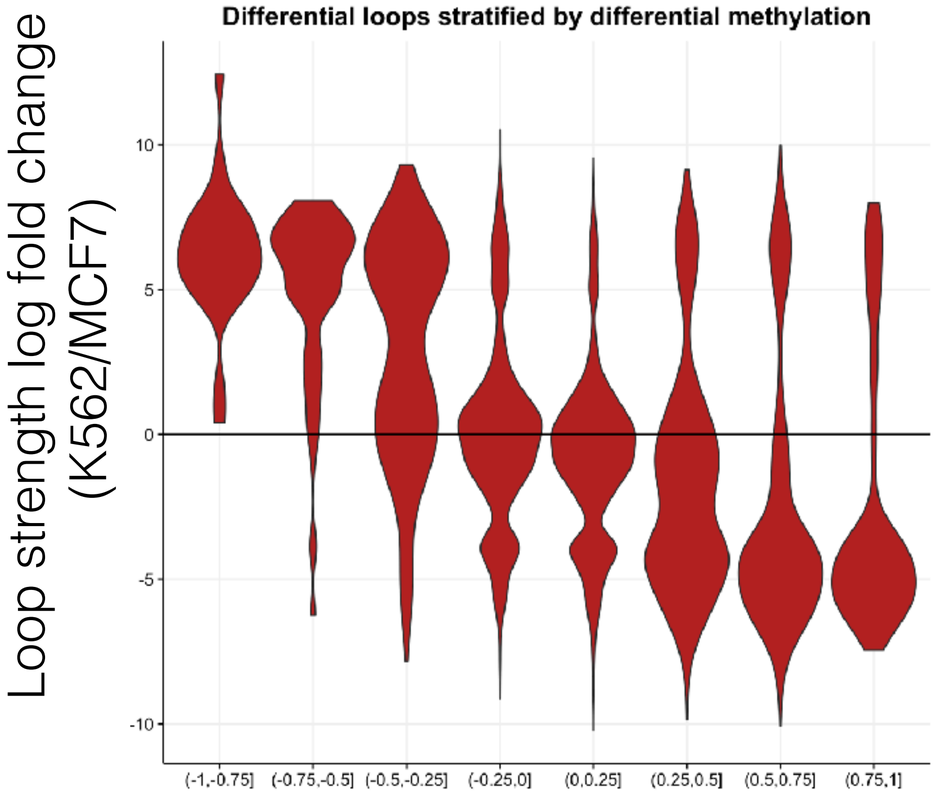
- Genome-wide identification of **POL2** ChIA-PET loops in MCF7 and K562
- At an **FDR of 1%**, there are **> 2,600** differential loops
- Nearly **2,000** were classified as **enhancer-promoter loops** (another 500 p-p)
- Target genes in MCF7 enriched for **Estrogen Response Pathways**

POL2 loop strength correlates with expression



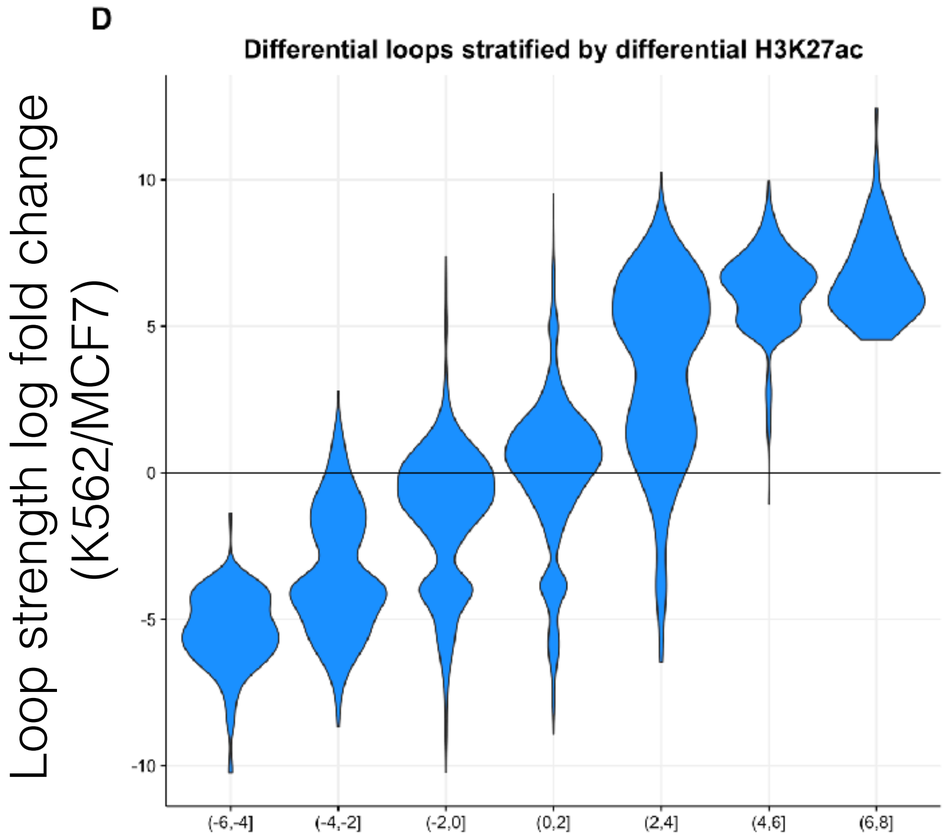
POL2 ChIA-PET

Epigenetic correlates of enhancer-promoter looping: DNA Methylation



POL2 ChIA-PET

Epigenetic correlates of enhancer-promoter looping: H3K27Ac



POL2 ChIA-PET

HiChIP

NATURE METHODS | BRIEF COMMUNICATION

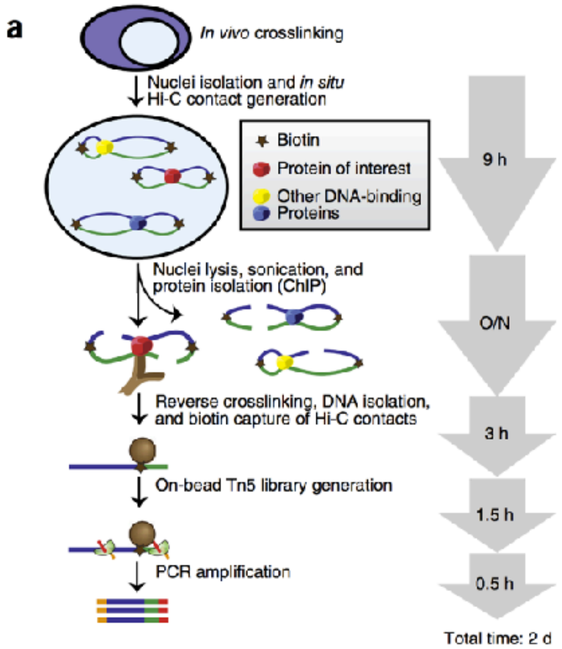
HiChIP: efficient and sensitive analysis of protein-directed genome architecture

Maxwell R Mumbach, Adam J Rubin, Ryan A Flynn, Chao Dai, Paul A Khavari, William J Greenleaf & Howard Y Chang

Affiliations | **Contributions** | **Corresponding author**

Nature Methods **13**, 919–922 (2016) | doi:10.1038/nmeth.3999

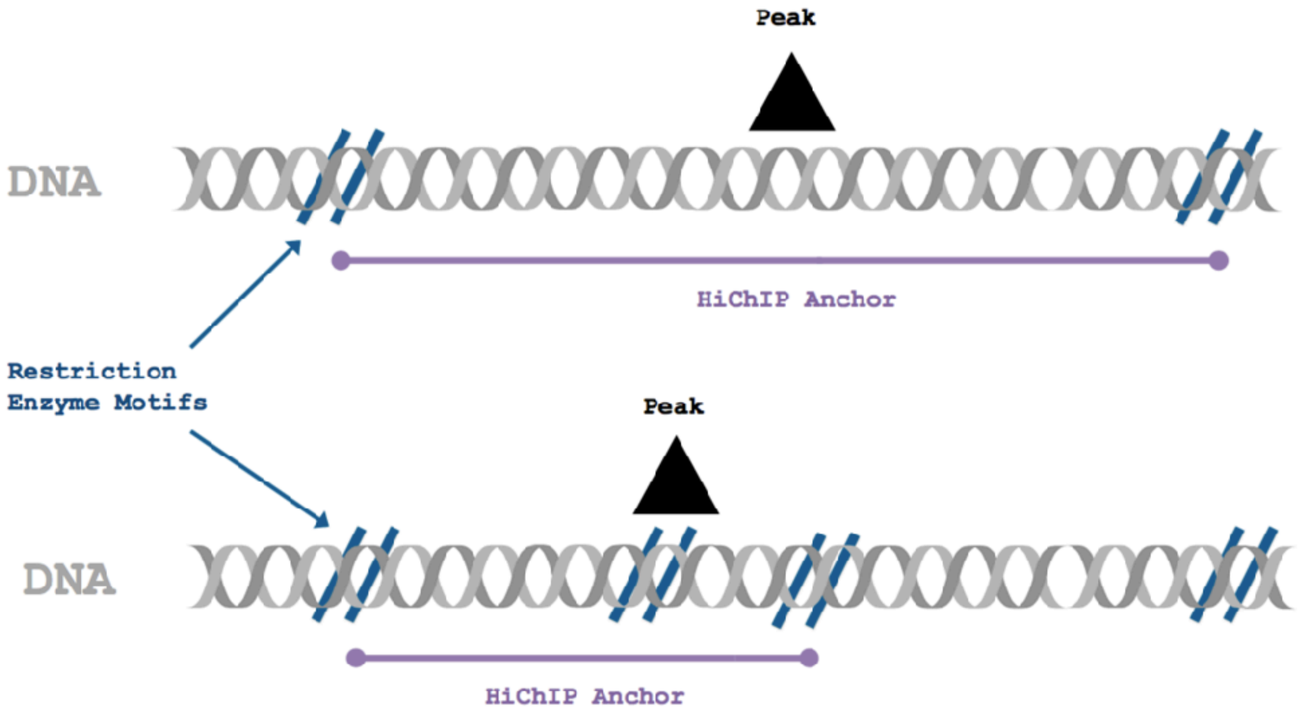
Received 02 May 2016 | Accepted 10 August 2016 | Published online 19 September 2016



- Protocol allows for < 1 million cells
- Higher read efficiency than ChIA-PET
- Shorter protocol

hichipper

This package is maintained by Caleb Lareau in the Aryee Lab. Source code is made freely available here and a packaged install version is provided through PyPi.



Acknowledgements

Caleb Lareau

Zack McCaw

Jose Malagon Lopez

Brad Bernstein

Sarah Johnstone

Esmat Hegazi

Rafael Irizarry

Alejandro Reyes