

Lab StandUp: 21 june

SUCCESS

- Lab Retreat!
- Joanna Wolthuis got an ePoster at Metabolomics conference next week!

CALLS FOR

- Make a list of discussion points for Deep Democracy consideration
- final PD candidates will visit to give seminars (June)



ACTION

- Clean up the hpc storage
 - See next slide
- Why jobs takes a long time to run? Is HPC too crowded?

Who Date

All

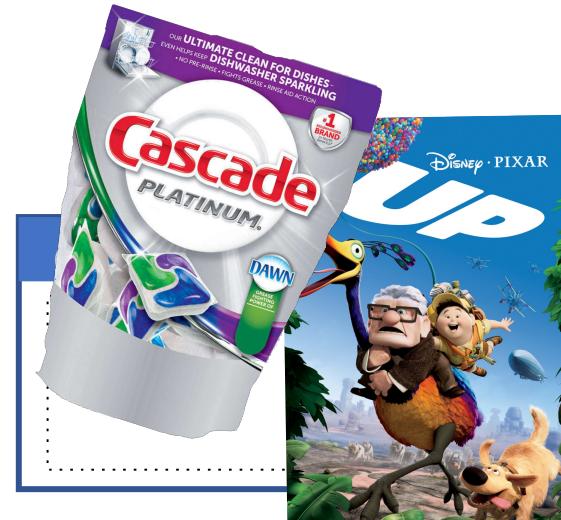
SURPRISES

surprise



OoO

- JvB holiday 22 June-7 July
- Jeroen @ISMB Basel 21-25 July
- Jeroen Holidays 10 - 26 Aug
- No (default) meetings 15-July - 1 Sept
- JW @Metabolomics2019 June 22-27, holiday/moving July 8-19
- Joske @ISMB Basel 21-25 July



Who Date

size;owner;path;;;	
43.7T;inijman;./ridder	166G;aallahyar;./ridder/users/dstoker
26.5T;mroosmalen;./ridder/nanopore	2.5T;jvonberg;./ridder/users/jvonberg
3.2T;aallahyar;./ridder/nanopore/mchc_recalled	1.1T;jvonberg;./ridder/users/jvonberg/LMM-comparison
123G;inijman;./ridder/nanopore/mchc_recalled/ESCs-1xDpnII	452G;spulit;./ridder/users/jvonberg/forJoanna
2.6T;inijman;./ridder/nanopore/mchc_recalled/ESCs-V211-350ng	900G;jvonberg;./ridder/users/jvonberg/time-to-event-gwas
435G;inijman;./ridder/nanopore/mchc_recalled/PromethION-R2	2T;jwolthuis;./ridder/users/jwolthuis
10.5T;rstraver;./ridder/nanopore/processed_data	648G;jwolthuis;./ridder/users/jwolthuis/Data
10.5T;rstraver;./ridder/nanopore/processed_data/RCA	1.3T;jwolthuis;./ridder/users/jwolthuis/MBX
12.7T;rstraver;./ridder/nanopore/raw_data	2.1T;lsantuari;./ridder/users/lsantuari
12.7T;rstraver;./ridder/nanopore/raw_data/RCA	1.9T;lsantuari;./ridder/users/lsantuari/Datasets
17.1T;aallahyar;./ridder/users	102G;lsantuari;./ridder/users/lsantuari/Processed
8.8T;aallahyar;./ridder/users/aallahyar	209G;mnieboer;./ridder/users/mnieboer
8.8T;aallahyar;./ridder/users/aallahyar/My_Works	163G;mnieboer;./ridder/users/mnieboer/from_hpc_tmp
140G;adanyi;./ridder/users/adanyi	340G;rstraver;./ridder/users/rstraver
622G;akuzniar;./ridder/users/akuzniar	117G;smehrem;./ridder/users/smehrem
262G;akuzniar;./ridder/users/akuzniar/COLO829	100G;smehrem;./ridder/users/smehrem/breakpoint-pairs
214G;akuzniar;./ridder/users/akuzniar/NA12878	Total space reserved: 300T
	Total space occupied: 280.2T

What causes loops?



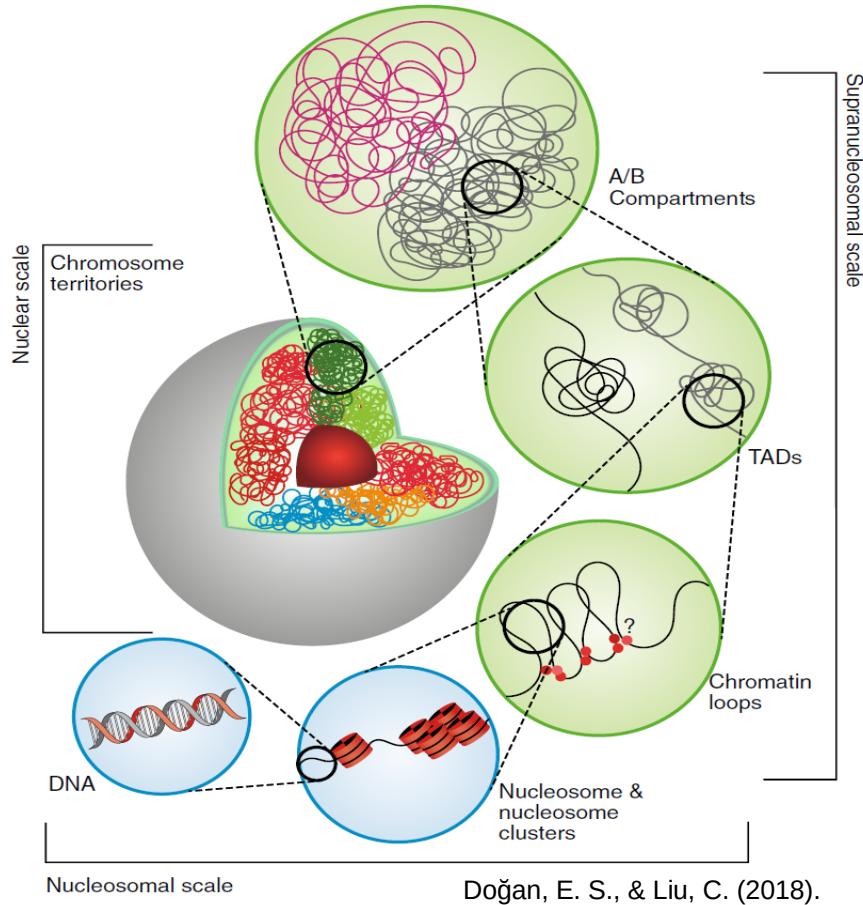
Content

- Introduction
- Data
- Research questions and results
- Conclusions

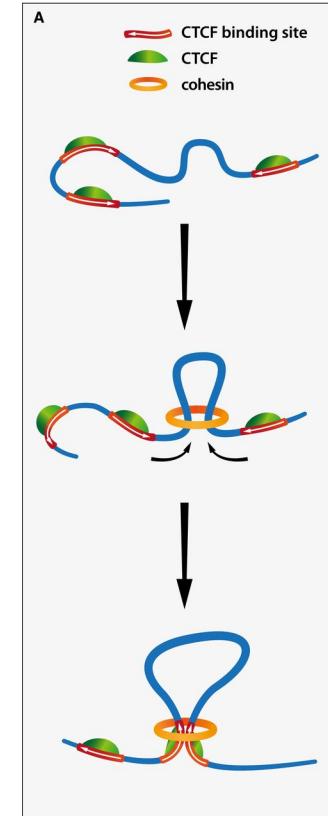
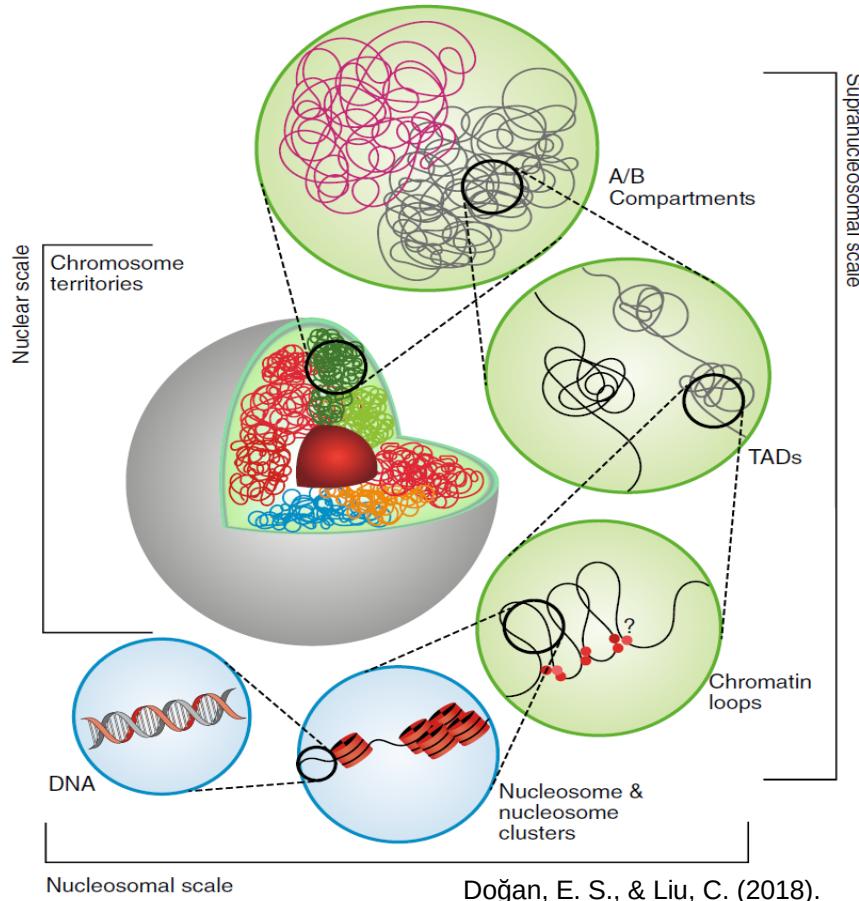
Introduction

- 3D genome conformation
- HiC
- Classification of HiC data

3D genome conformation

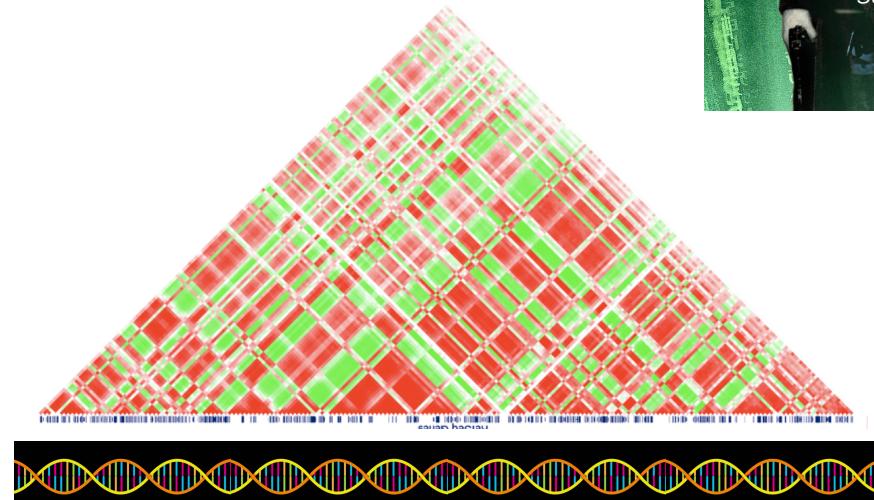
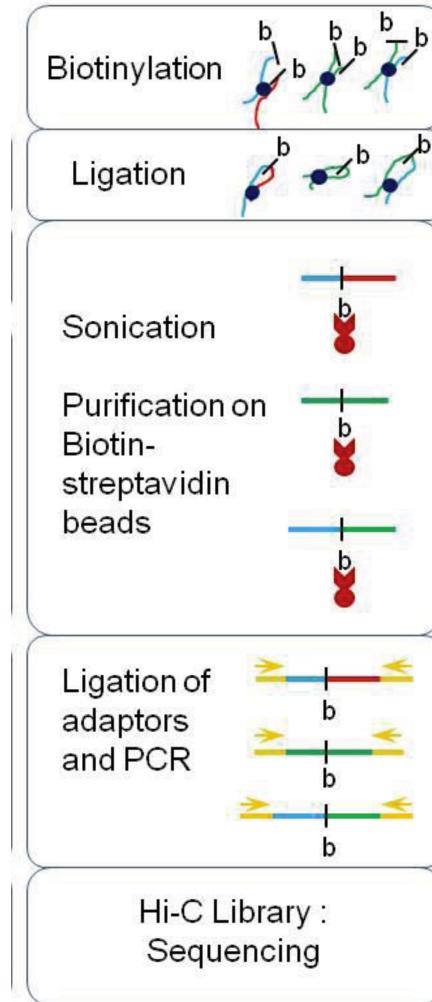


3D genome conformation

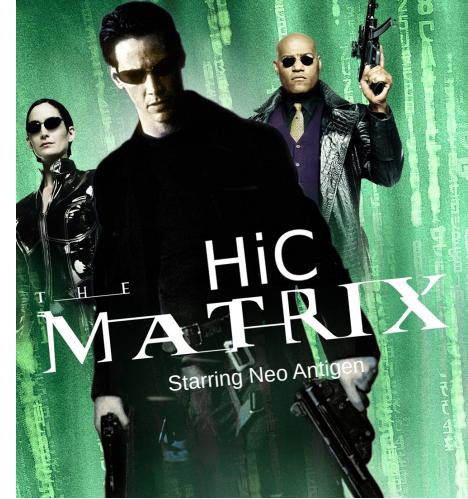


Wutz, G., et al. (2017)

HiC



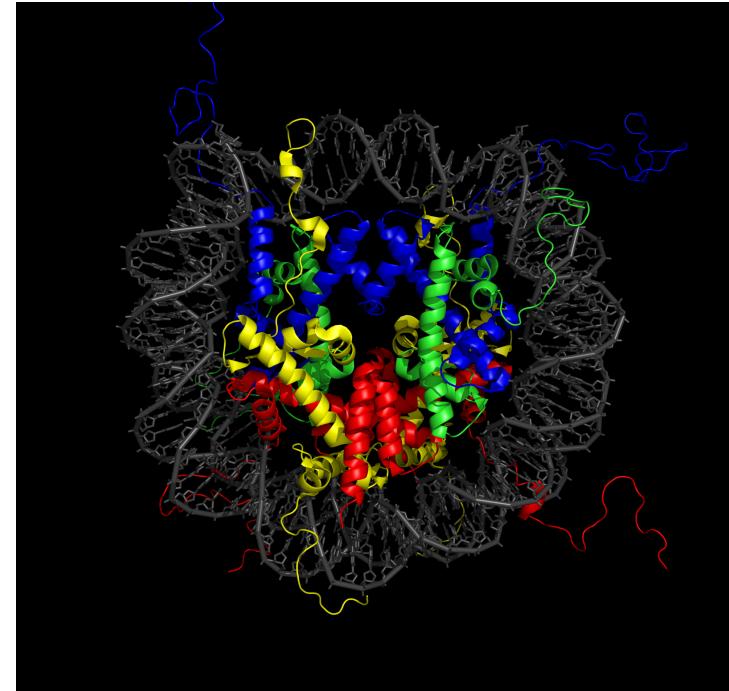
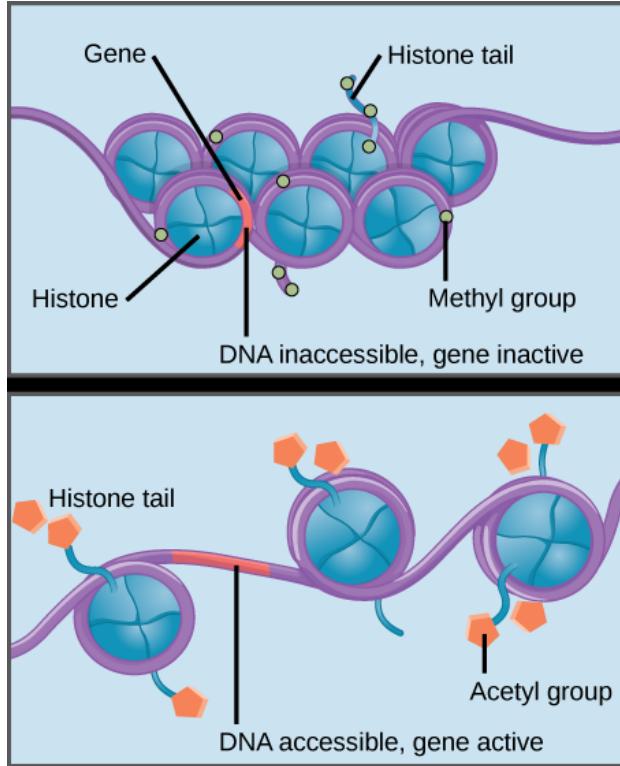
Ea, V., Baudement, M. O., Lesne, A., & Forné, T. (2015)



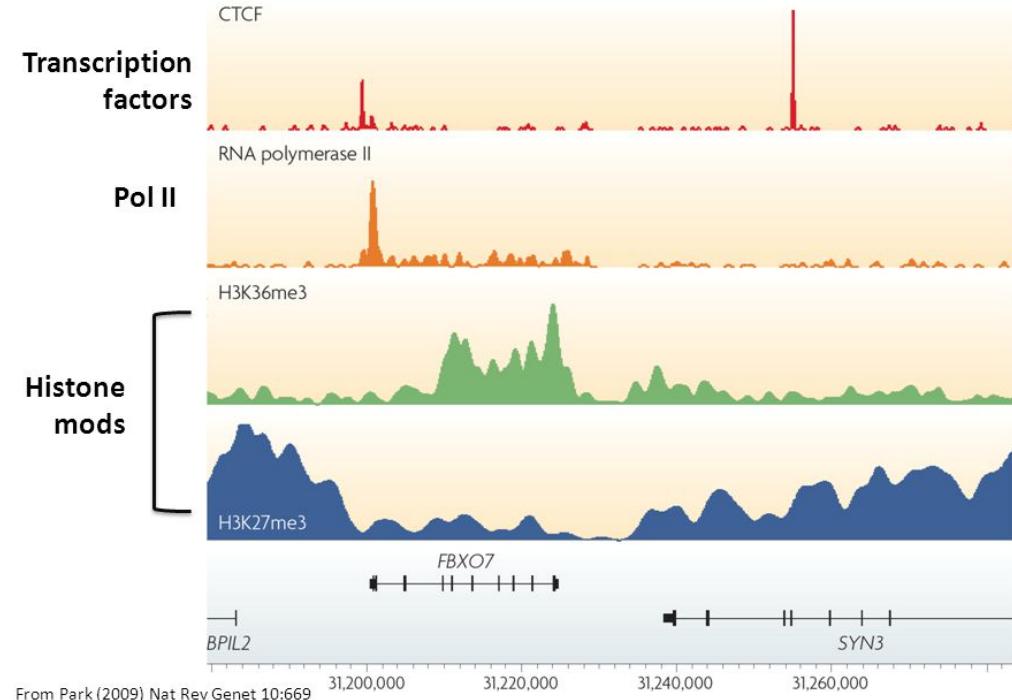
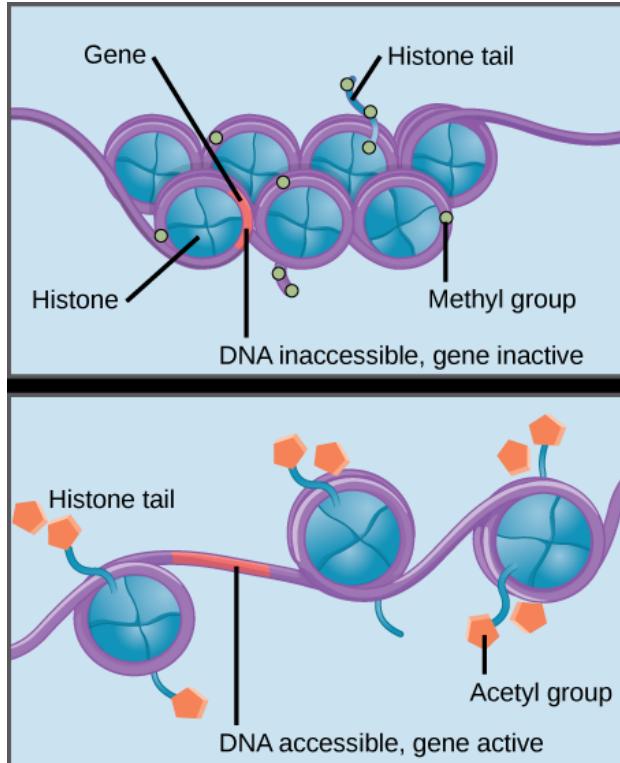
What we want to know

- What causes two loop anchors to interact?
- Can we use machine learning for this?

Variety of chromatin mark profiles

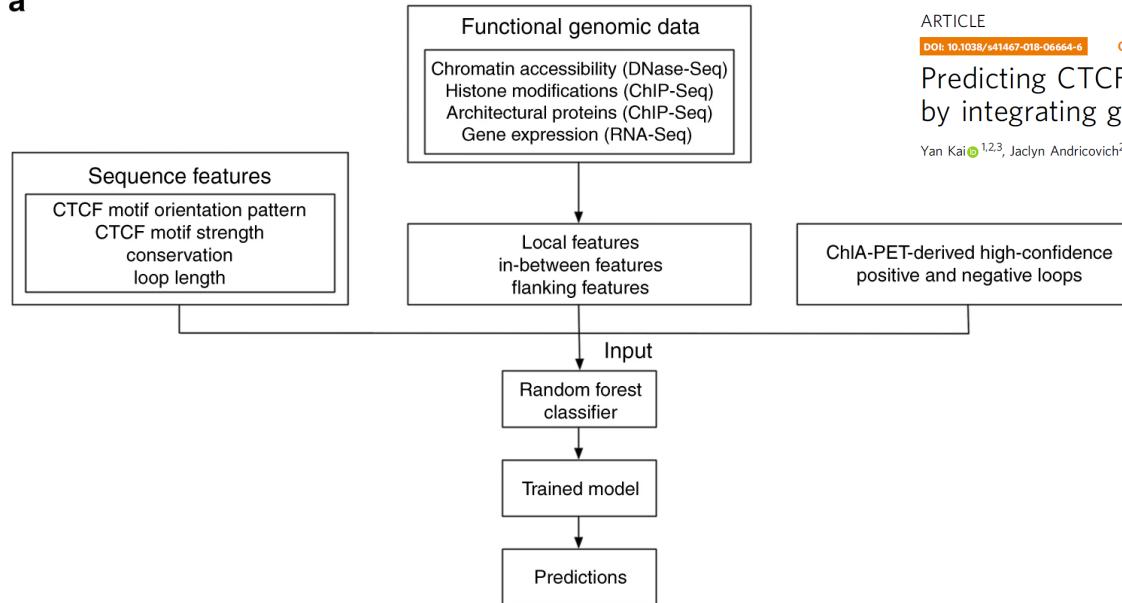


Variety of chromatin mark profiles



Classification of HiC data II: Predicting whether given CTCF sites will loop

a



ARTICLE

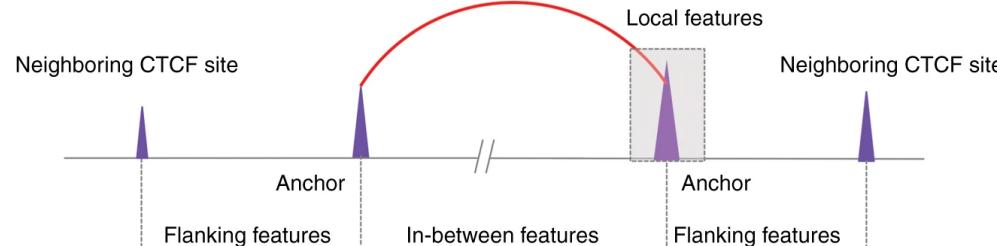
DOI: 10.1038/s41467-018-06664-6

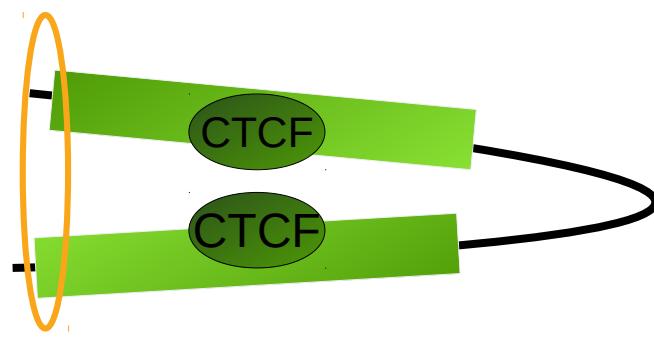
OPEN

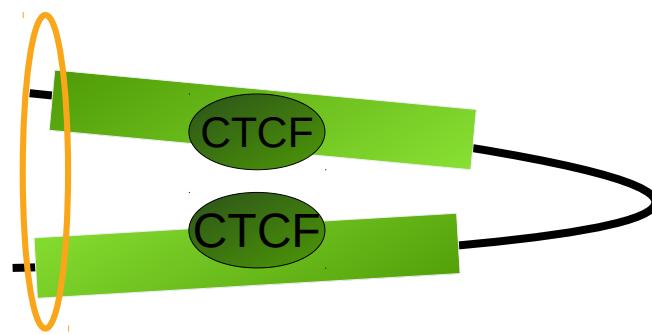
Predicting CTCF-mediated chromatin interactions by integrating genomic and epigenomic features

Yan Kai^{1,2,3}, Jaclyn Andricovich^{2,3}, Zhouhao Zeng¹, Jun Zhu^{1,4}, Alexandros Tzatsos^{2,3} & Weiqun Peng¹

b

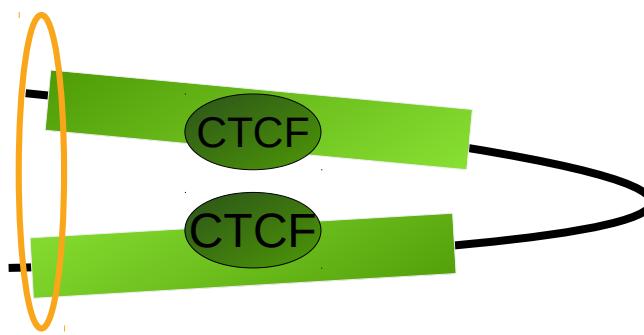






HiC information (loop data)

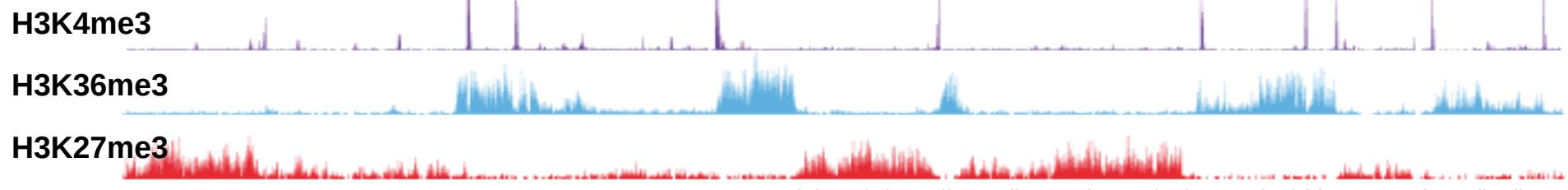




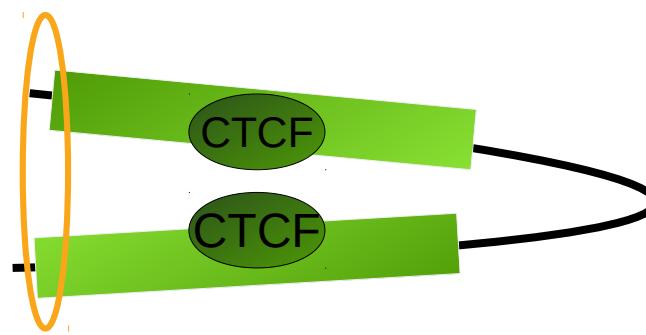
HiC information (loop data)



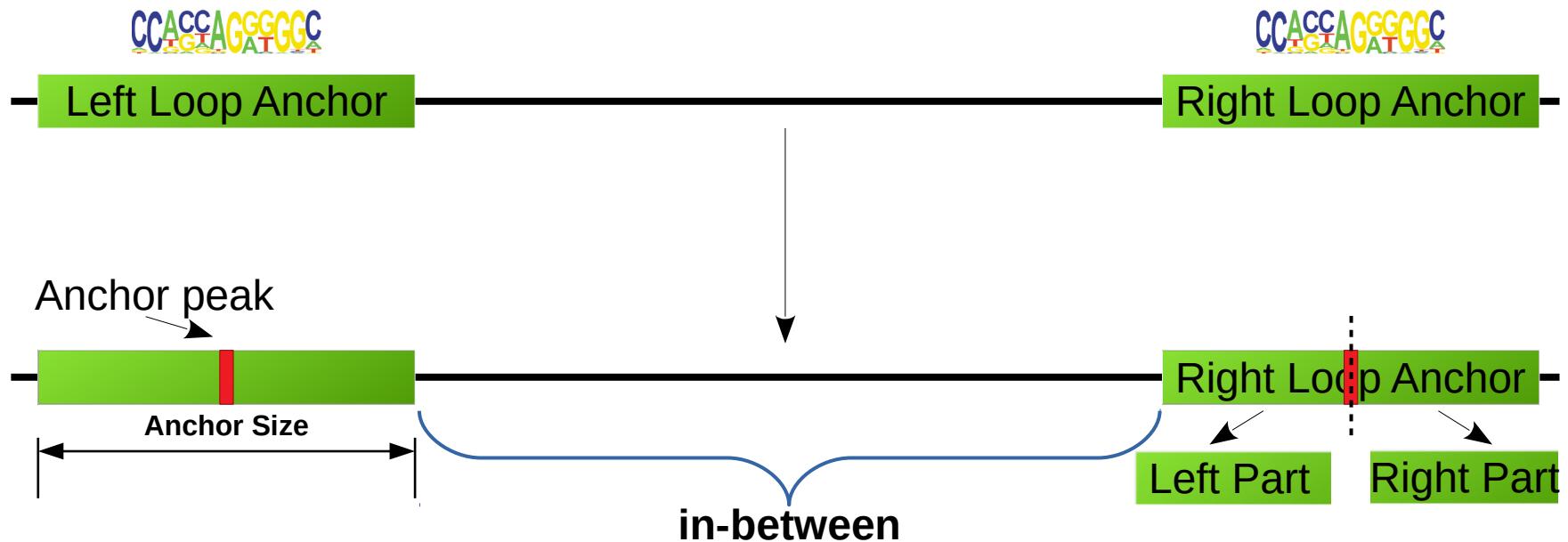
ChIP-Seq information

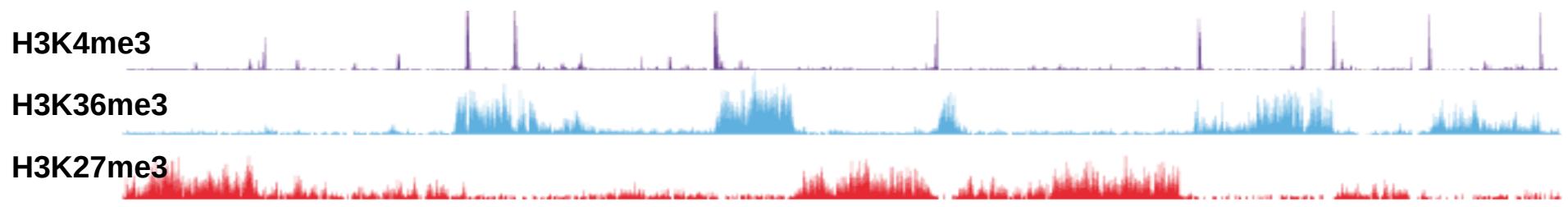
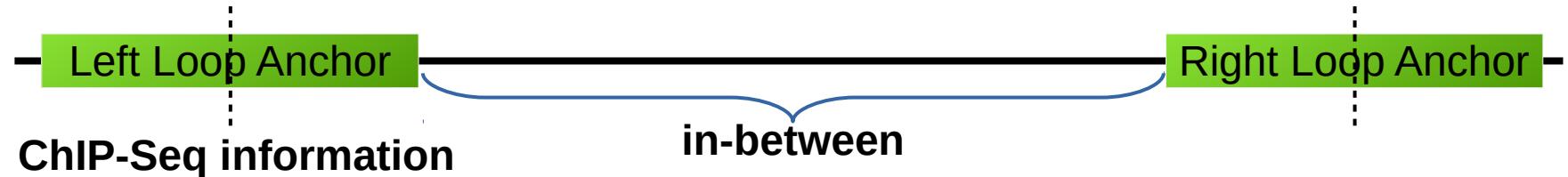


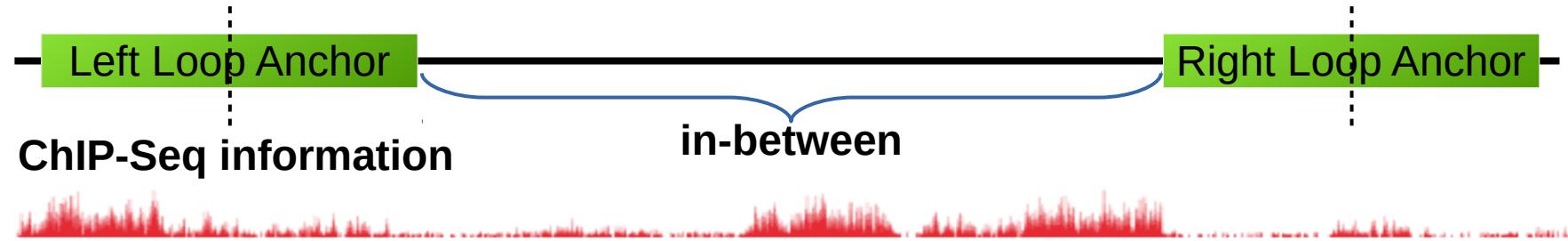
Wong, T. C., et al. (2011); <https://www.diagenode.com/en/categories/chip-seq-grade-antibodies>.

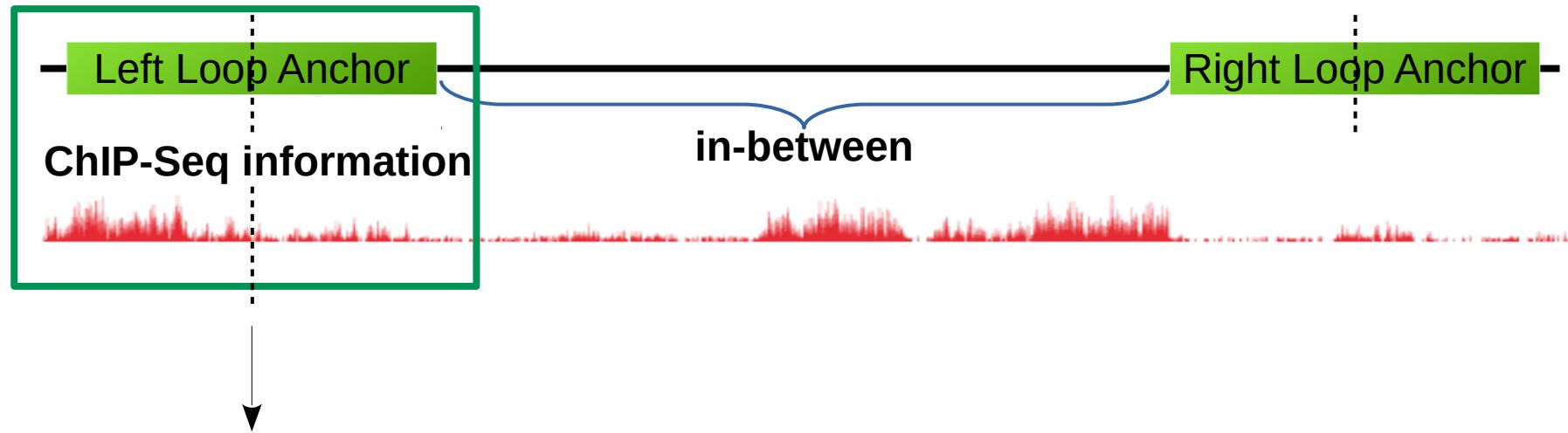


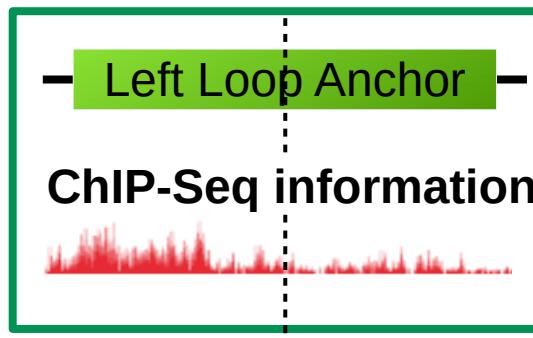
HiC information (loop data)





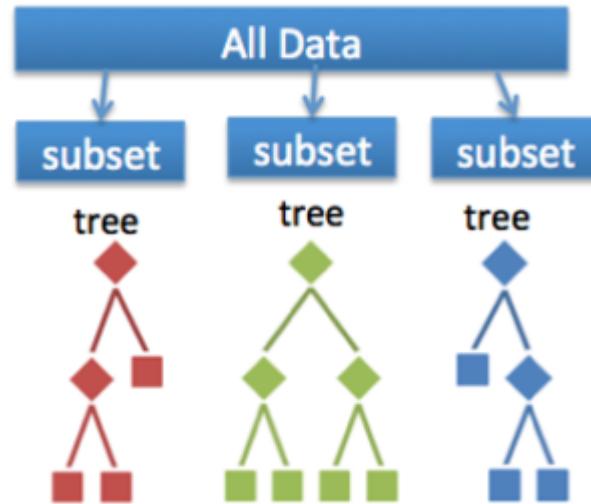




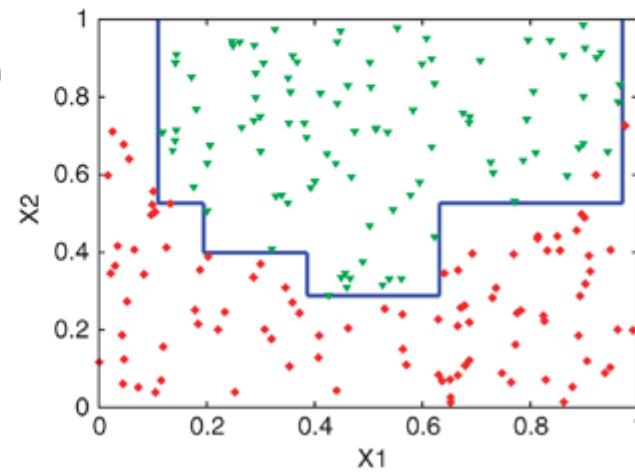
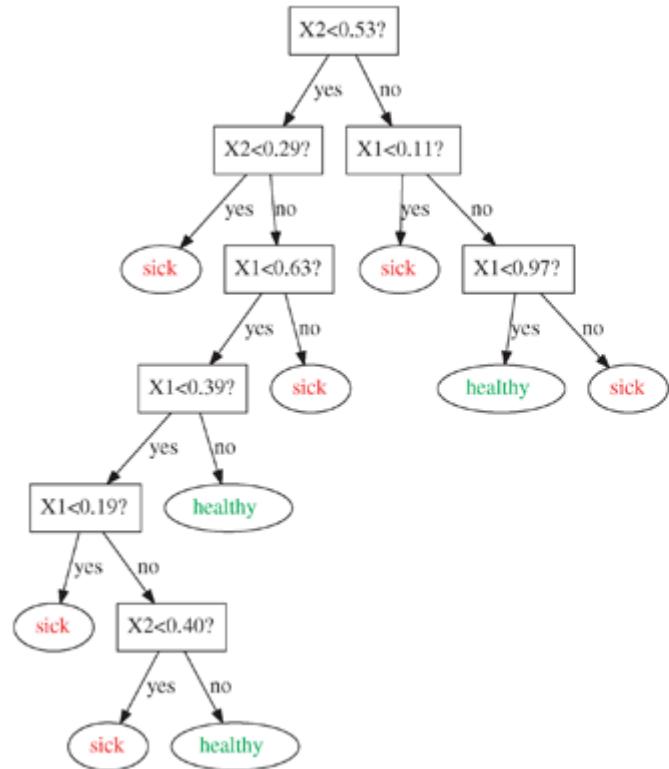


Location	ChIP-Seq mark	Feature name	Feature value
Left Anchor Left	RED	Significance	0.005
Left Anchor Left	RED	Coverage fraction	0.7

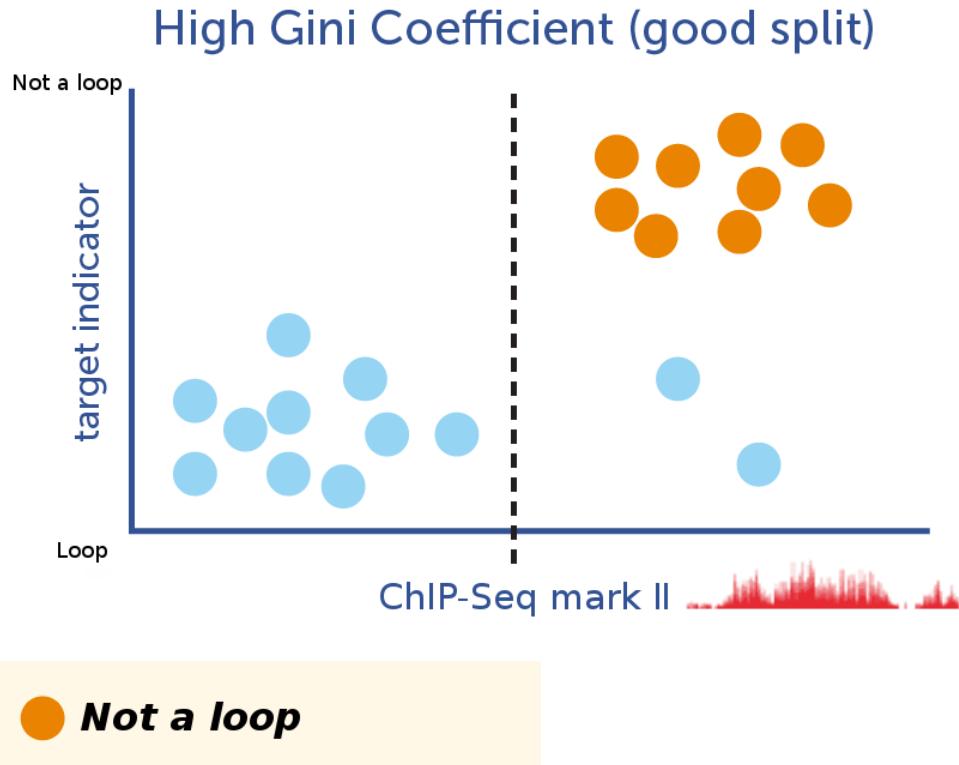
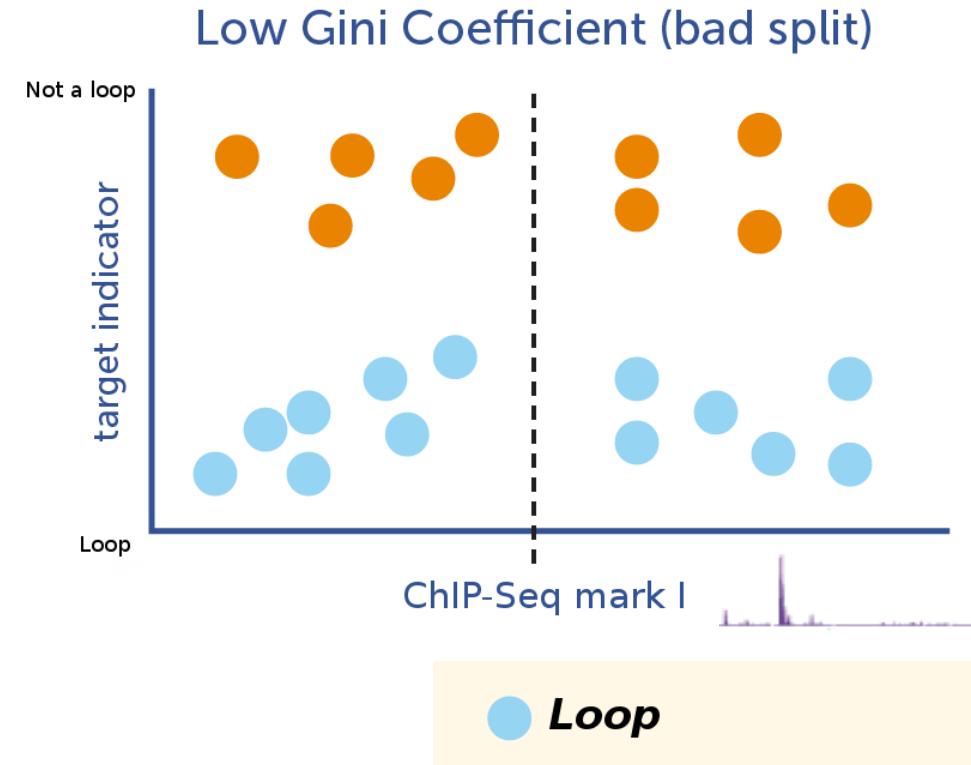
Classification with Random Forest



Classification with Random Forest



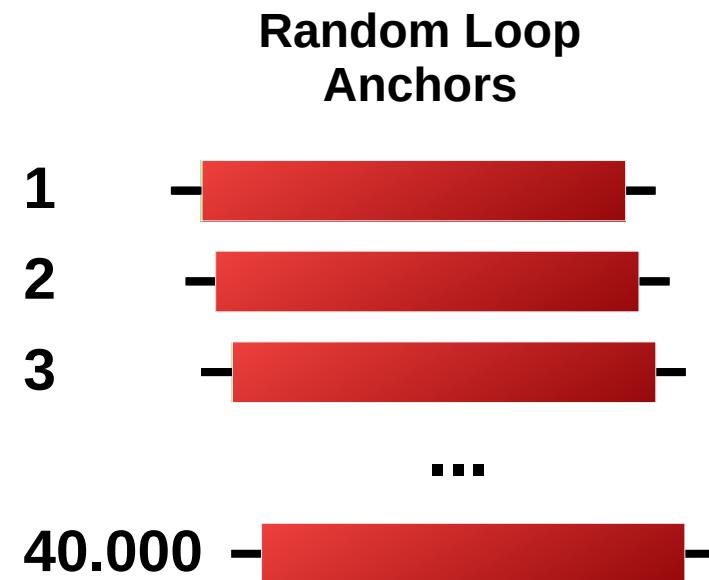
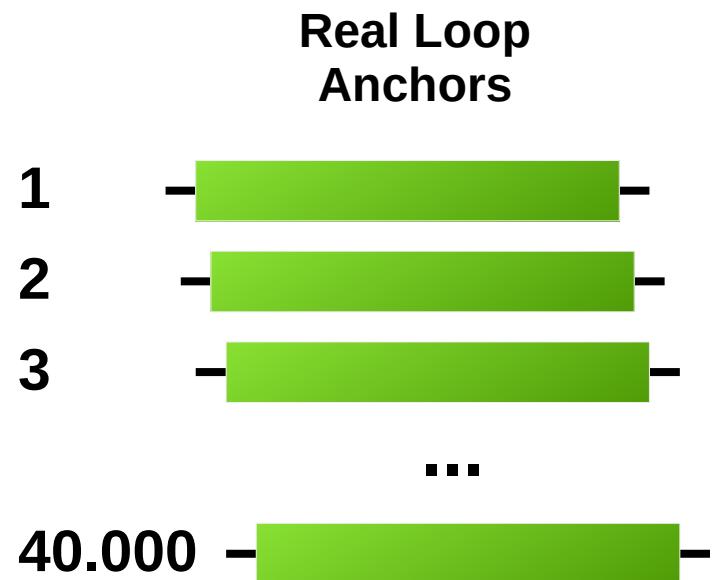
Classification with Random Forest



Questions

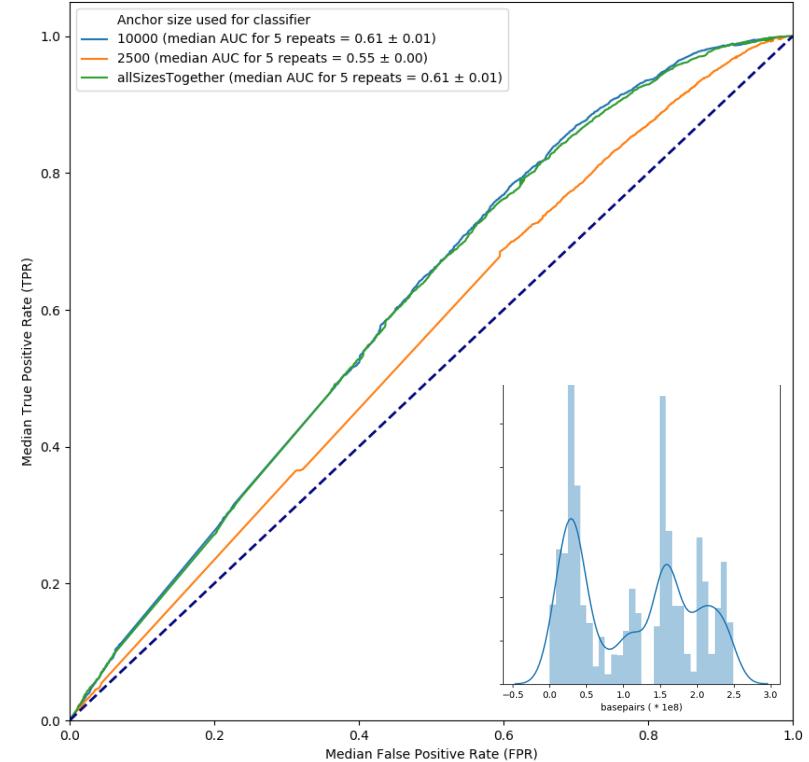
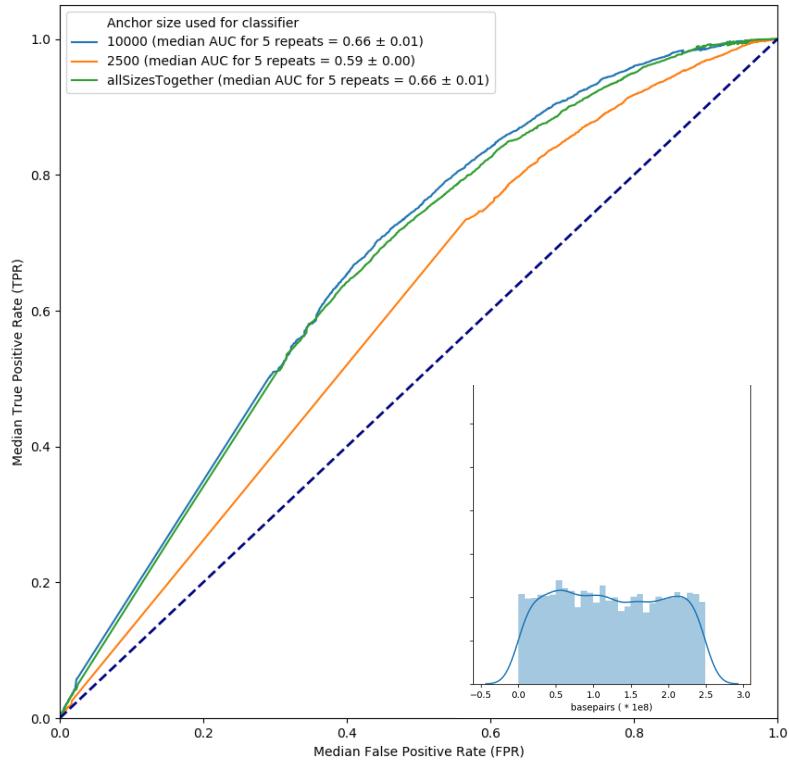
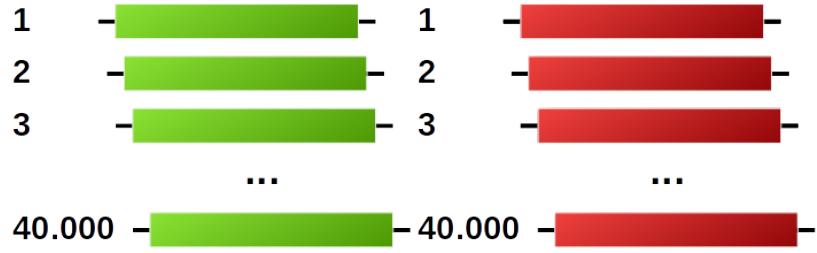
- What marks make you a loop anchor?
 - CTCF
 - Accessibility

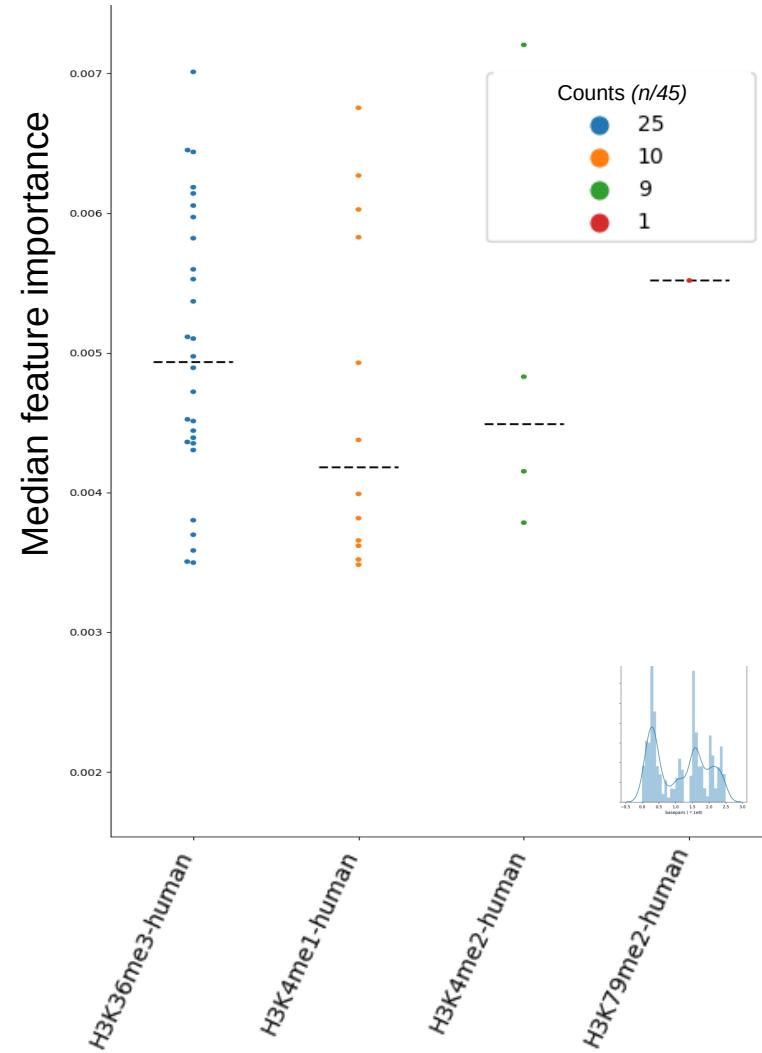
Questions



**Real Loop
Anchors**

**Random Loop
Anchors**



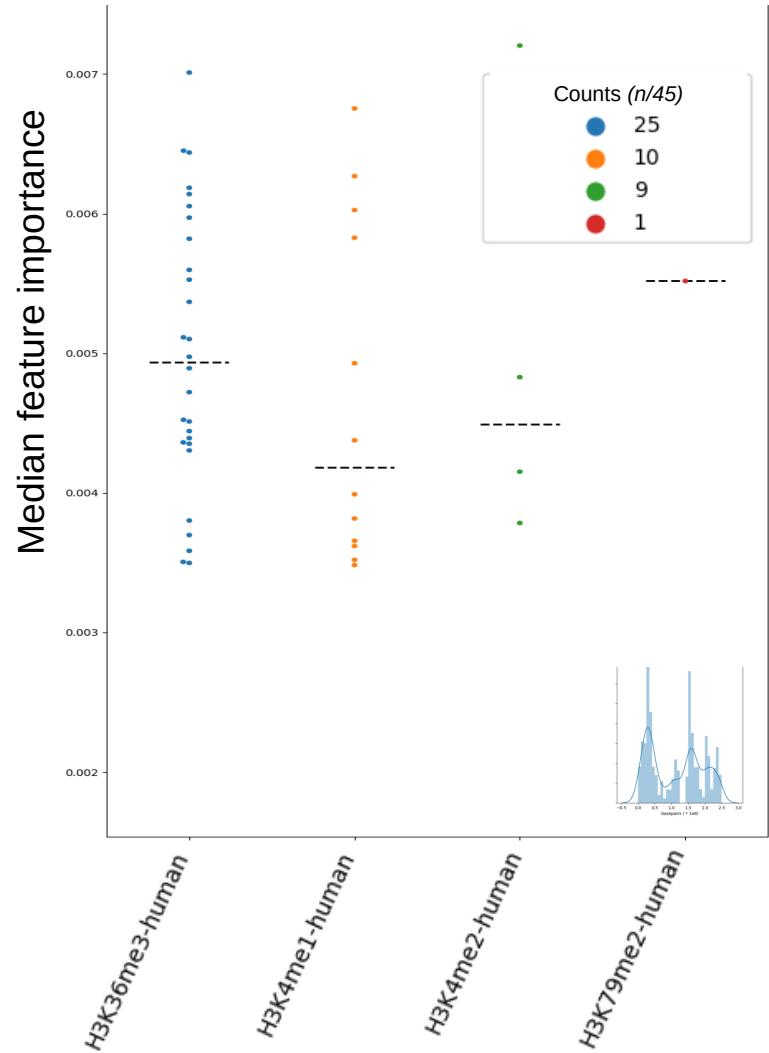




- H3K36me3

- Marks exons but also heterochromatin

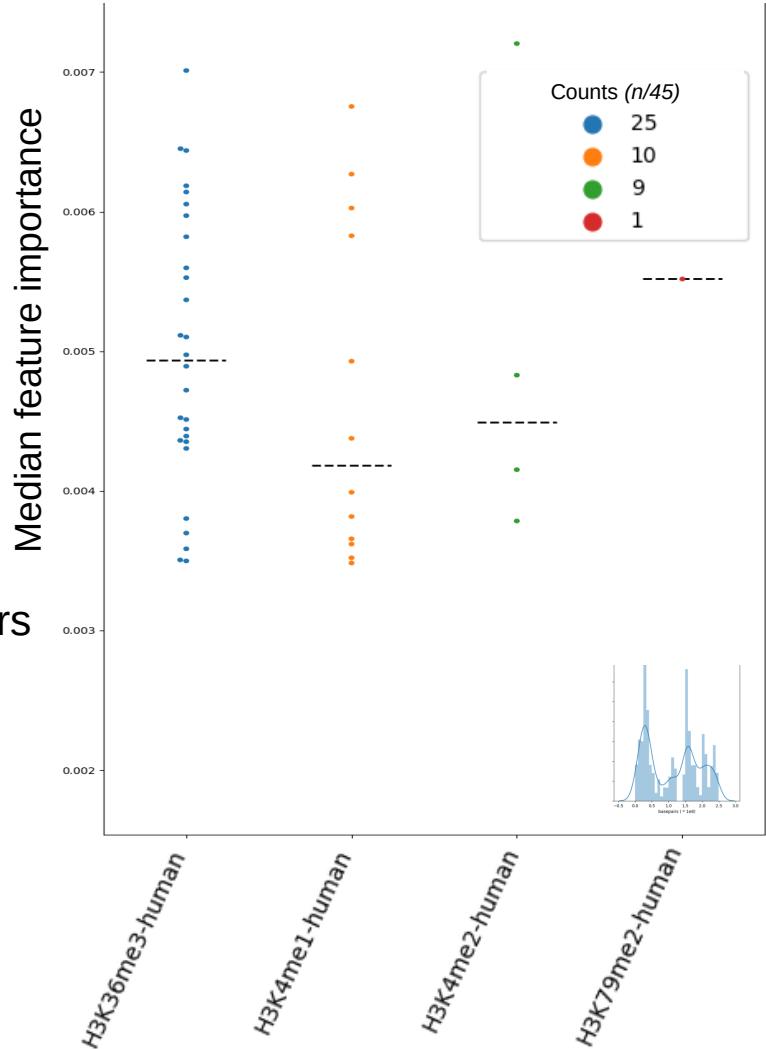
De Almeida, S. F., & Carmo-Fonseca, M. (2012); Wagner, E. J., & Carpenter, P. B. (2012); Chantalat, S., et al. (2011)





- H3K36me3
 - Marks exons but also heterochromatin
De Almeida, S. F., & Carmo-Fonseca, M. (2012); Wagner, E. J., & Carpenter, P. B. (2012); Chantalat, S., et al. (2011)
- H3K4me1
 - Enriched at enhancers. Distinguishes between enhancers and proximal promoters (together with H3K36me3 depletion)

Rada-Iglesias, A. (2018); Heintzman, et al. (2009)



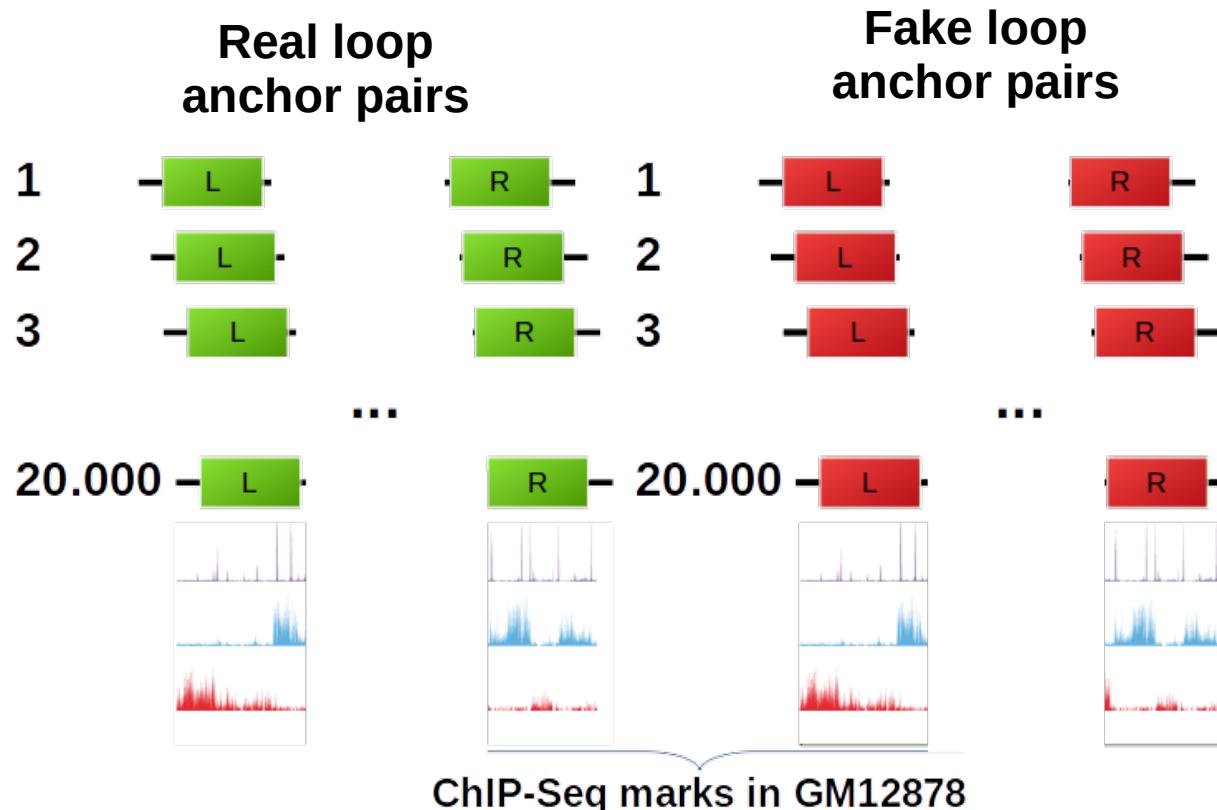
What marks you as a loop anchor?

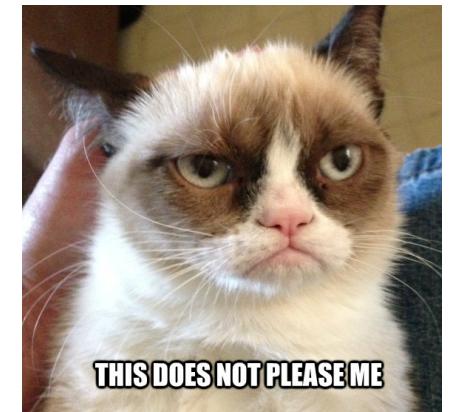
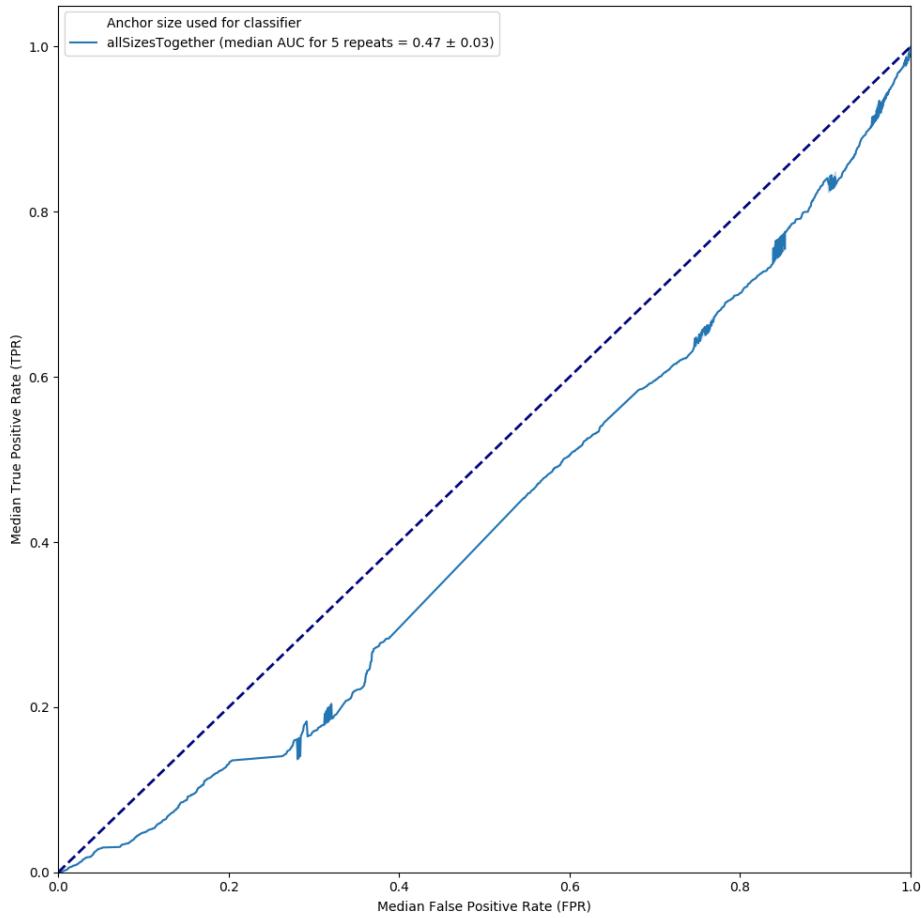
- Information on exons/enhancers/heterochromatin
 - More difficult to distinguish if from the same genomic area

Questions

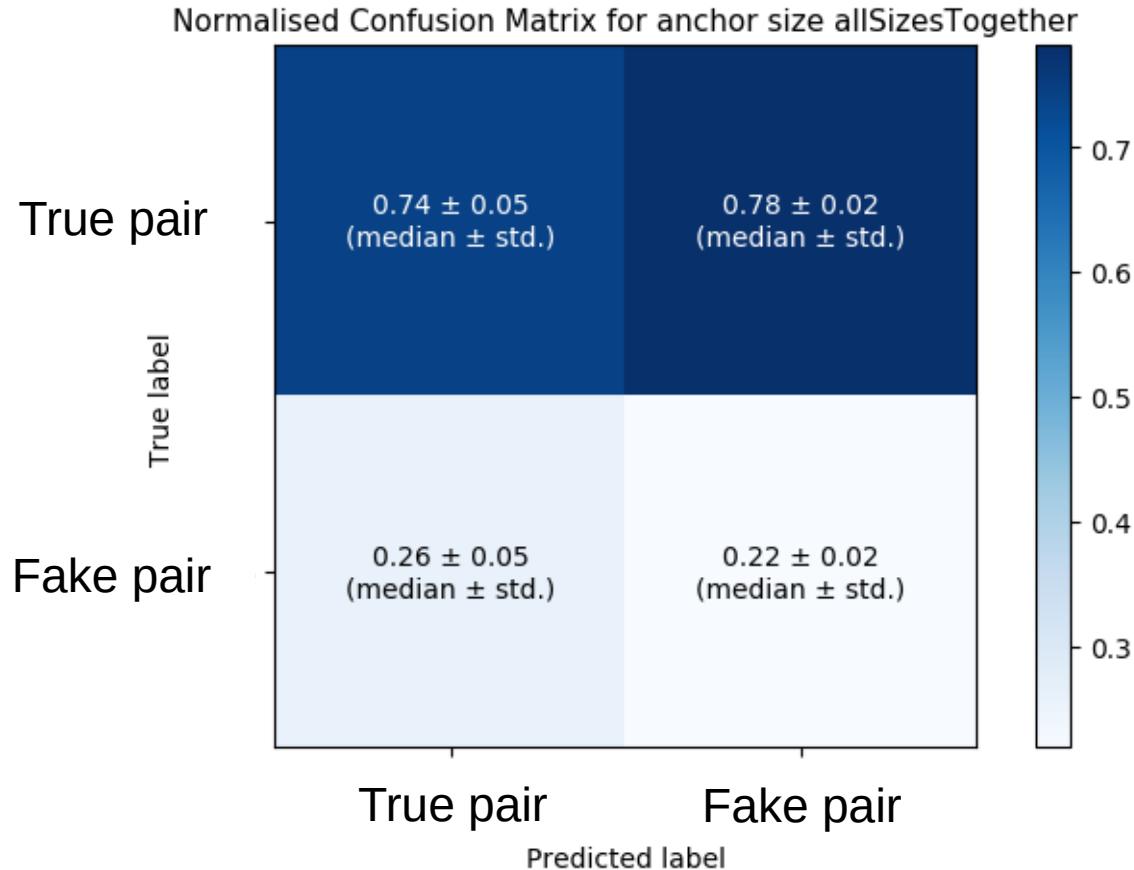
- What marks make you a loop anchor?
- What marks make you part of a *specific* loop?
 - Are there signals that cause specific anchors to interact?

Questions

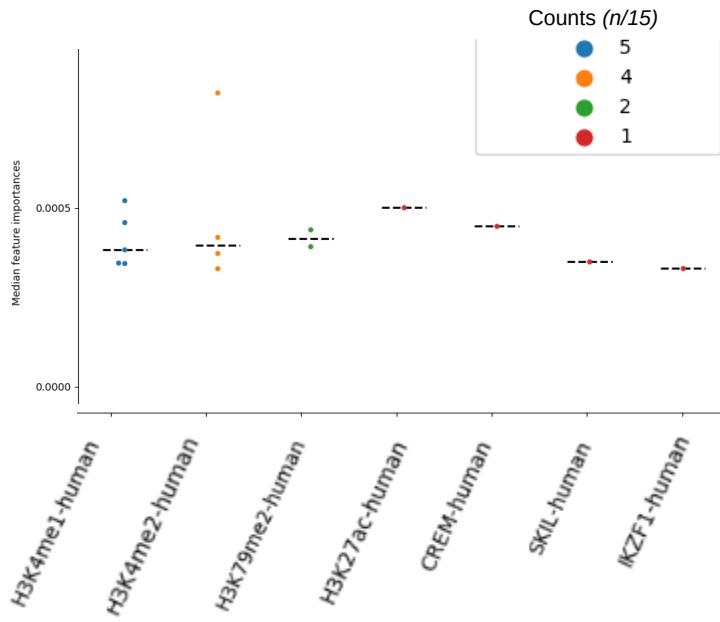




What error is made?



What about features?



What marks you as part of a specific loop?

- No combination of features can discern real interacting loop anchors from fake ones.
- Possible explanations:

What marks you as part of a specific loop?

- No combination of features can discern real interacting loop anchors from fake ones.
- Possible explanations:
 - There is simply no anchor-based predestination of who will loop with whom

What marks you as part of a specific loop?

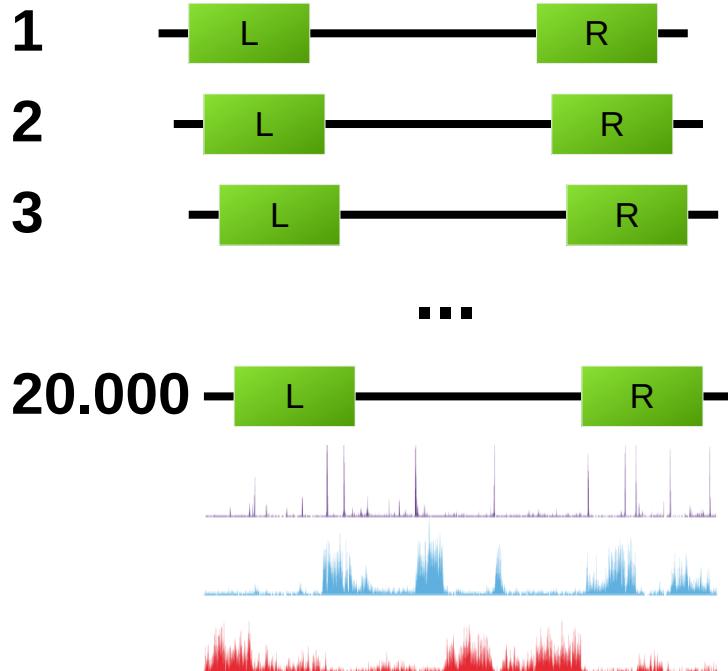
- No combination of features can discern real interacting loop anchors from fake ones.
- Possible explanations:
 - There is simply no anchor-based predestination of who will loop with whom
 - Many alternative (fake) loops actually are formed under certain conditions,
so there is no real division to train on

Questions

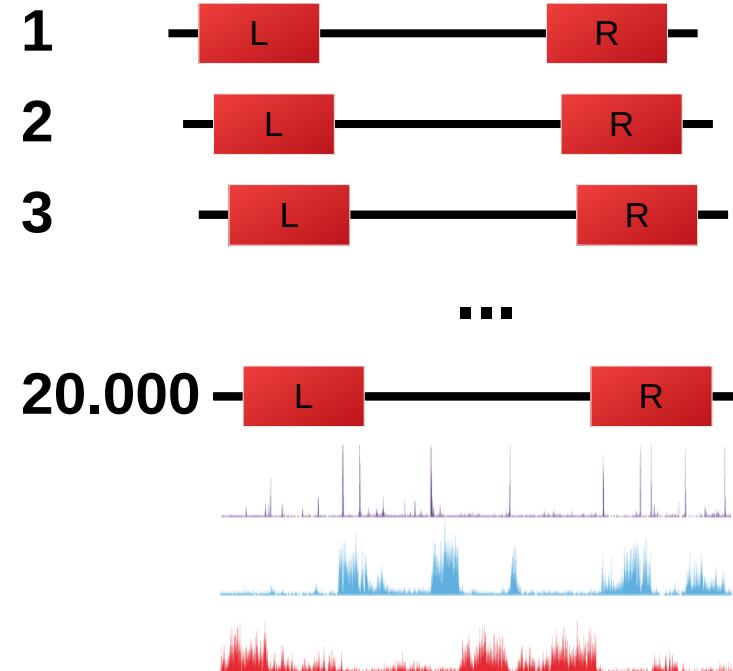
- What marks make you a loop anchor?
- What marks make you part of a *specific* loop?
 - Are there signals that cause specific anchors to interact?
- What marks distinguish present loops from possible loops ?

Questions

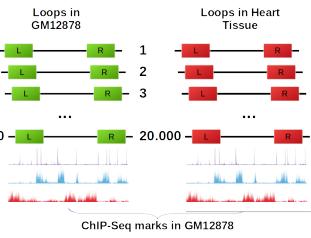
Loops in
GM12878



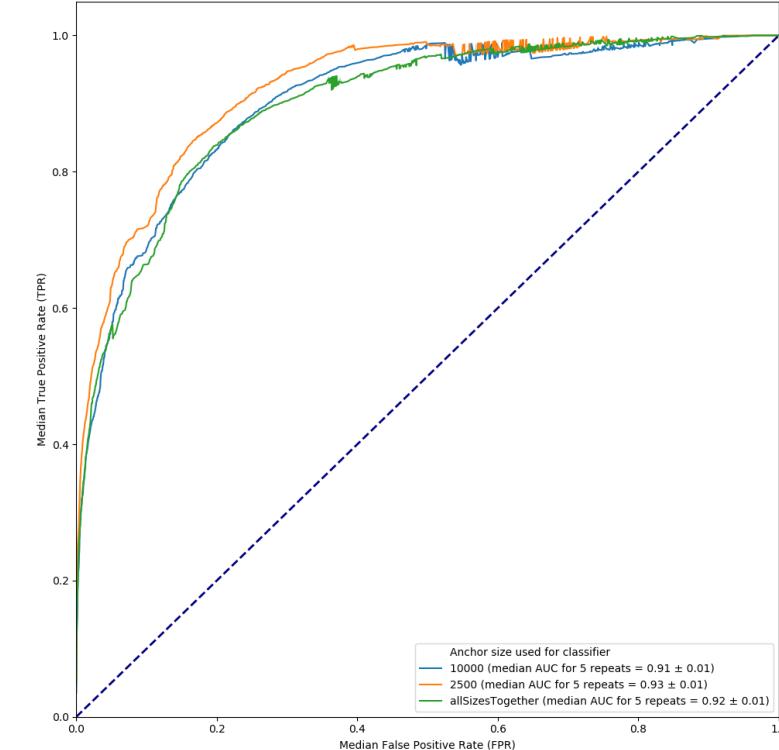
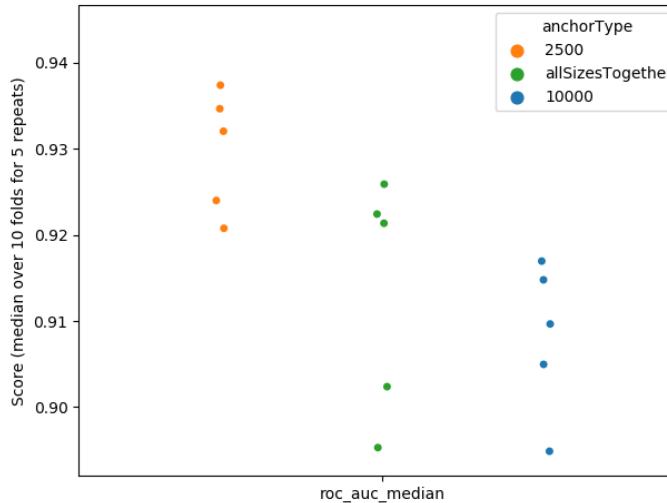
Loops in Heart
Tissue



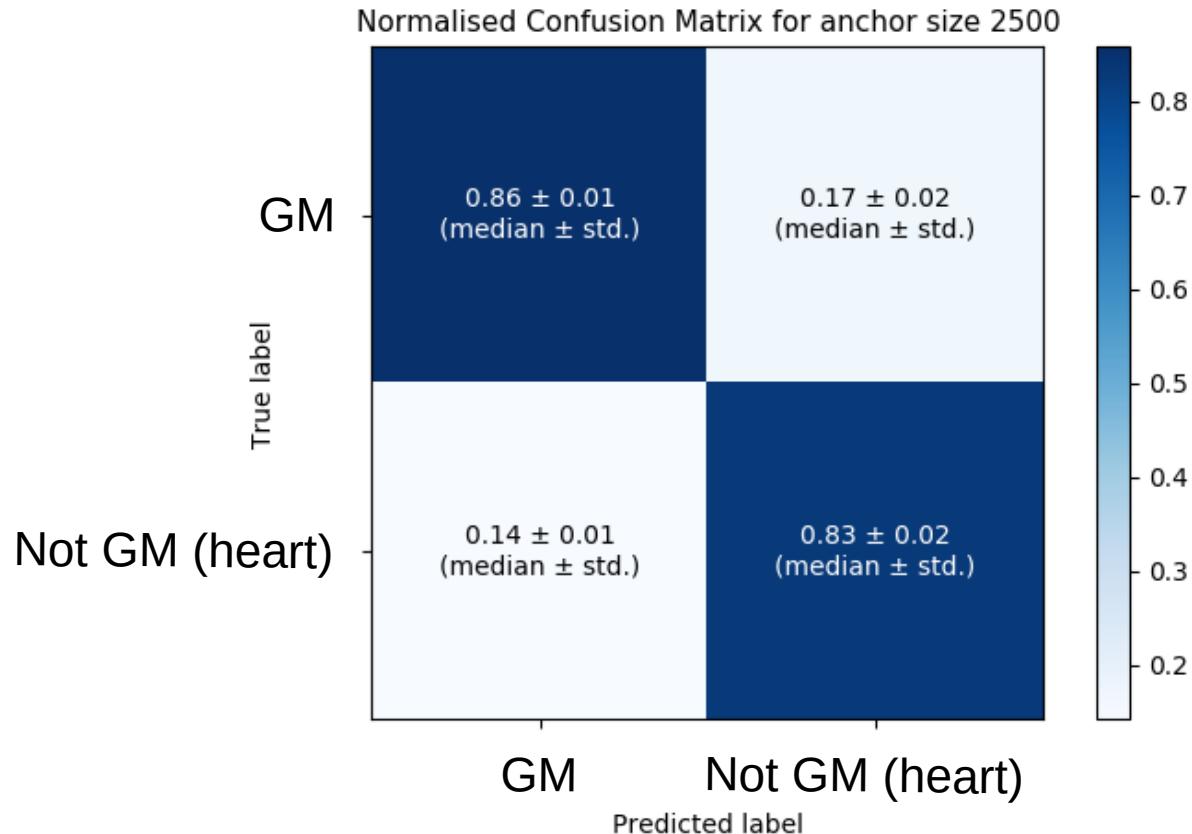
ChIP-Seq marks in GM12878

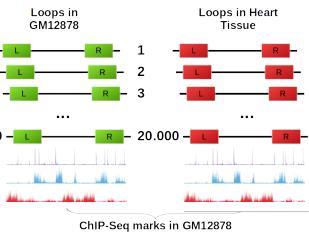


Good classification result

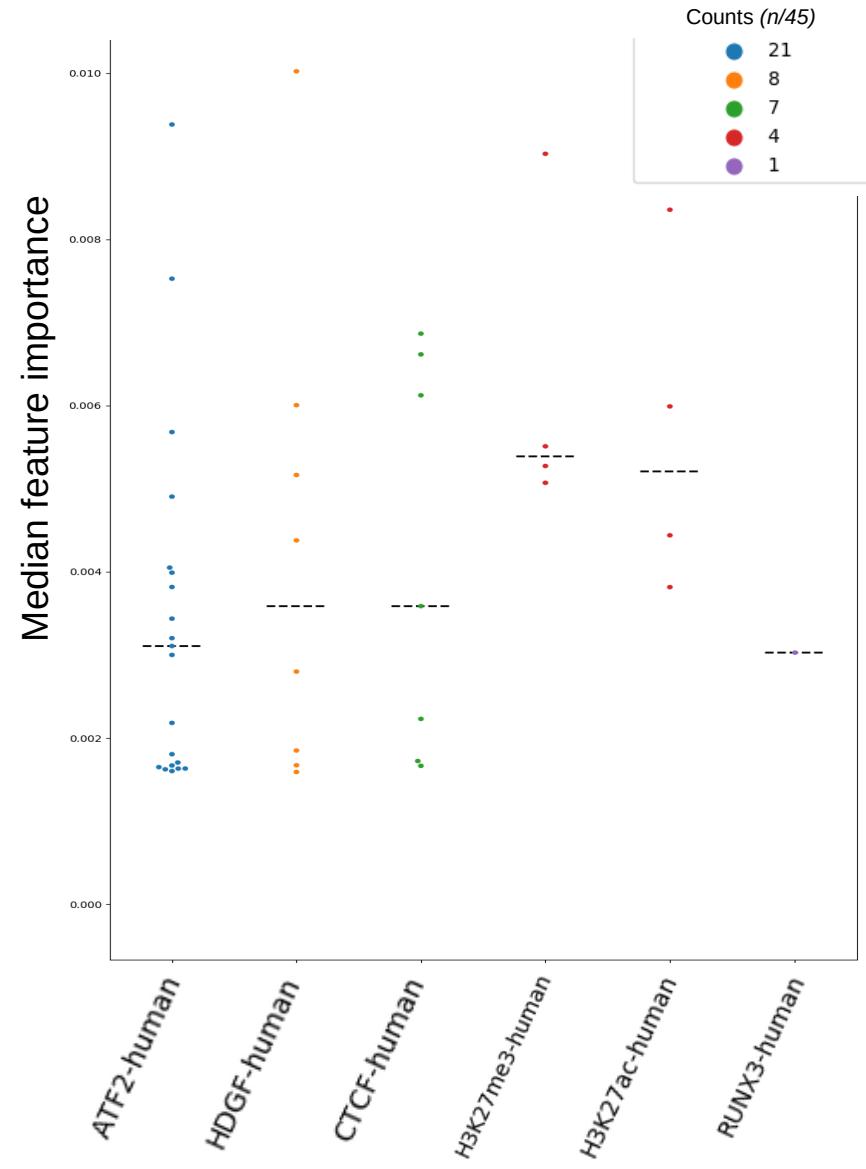


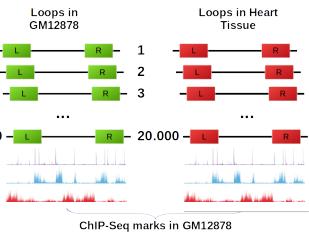
Balanced in what goes wrong





What marks are involved?



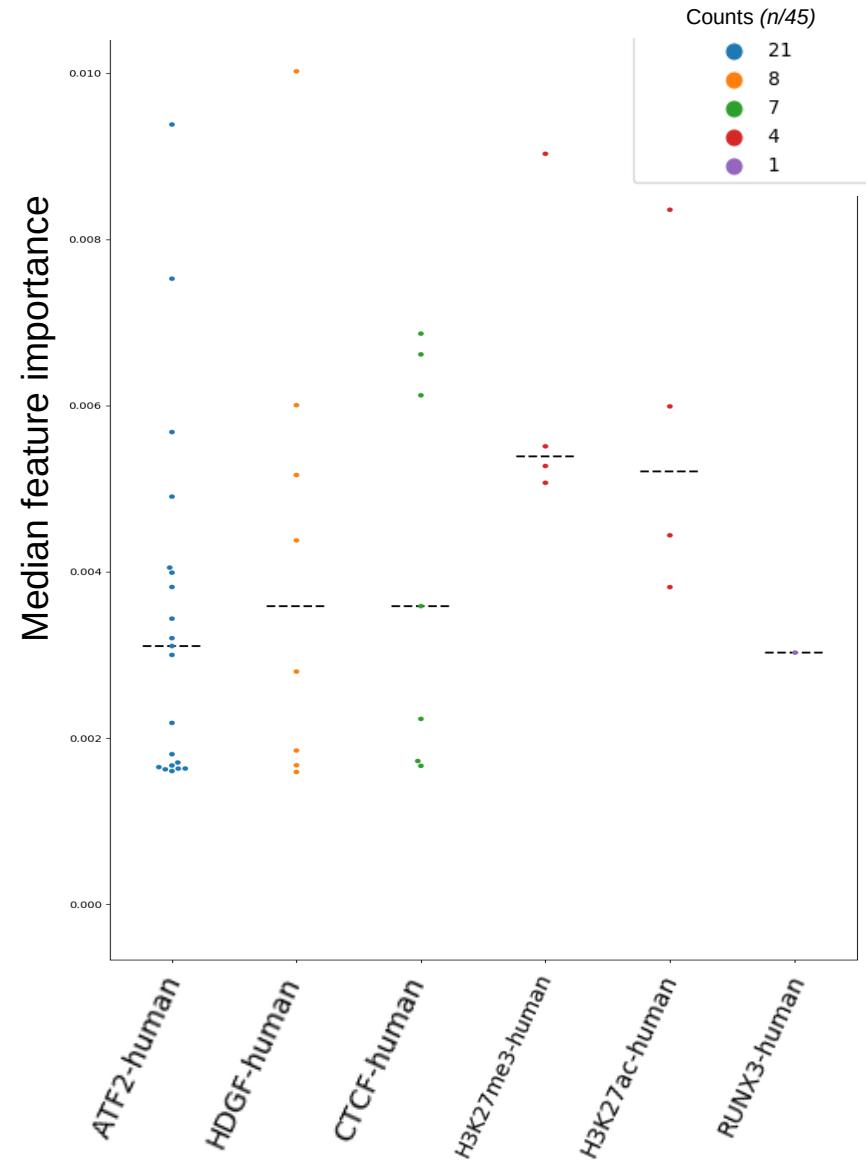


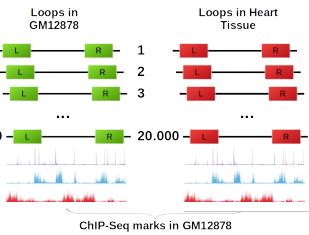
What marks are involved?

- ATF2

- Acetylates histones H2B and H4 *in vitro*. Yeast homologues involved with histone deacetylation and correct heterochromatin formation.

Lau, E., & Ze'ev, A. R. (2012); Kawasaki, H., et al. (2000).



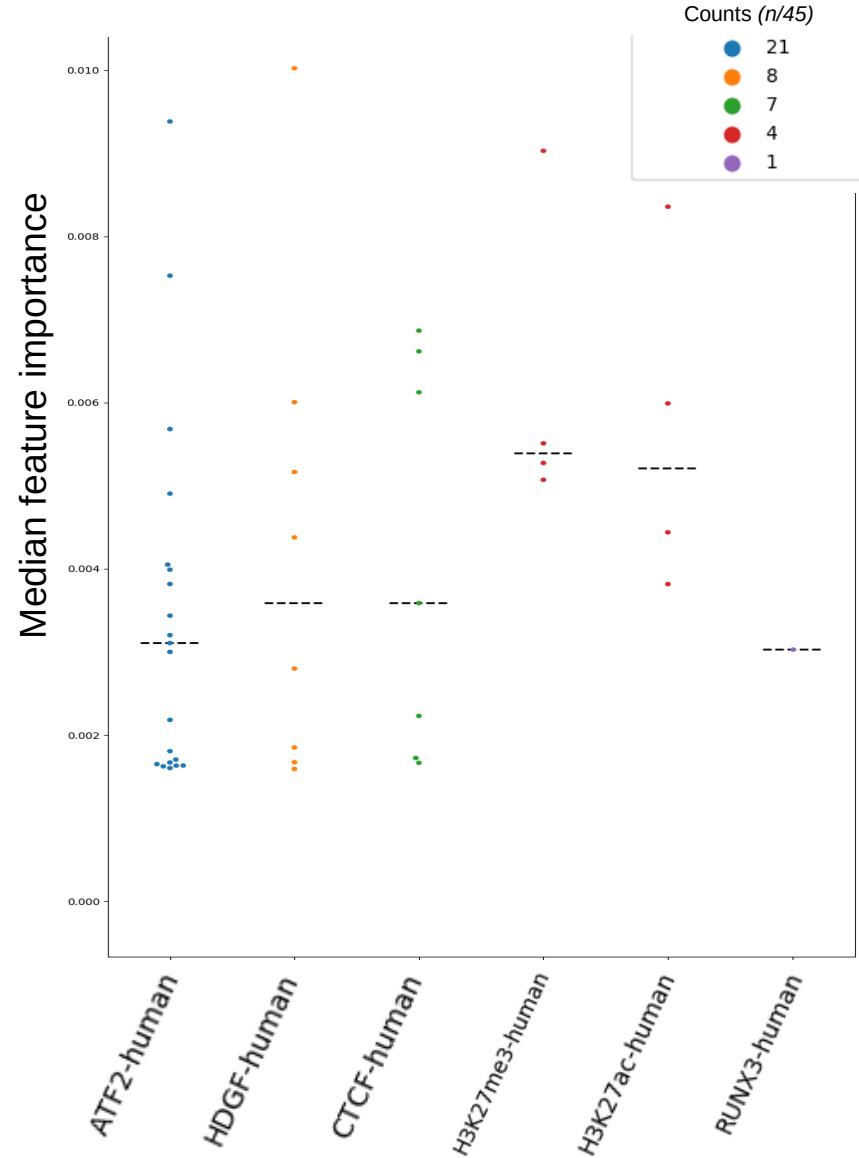


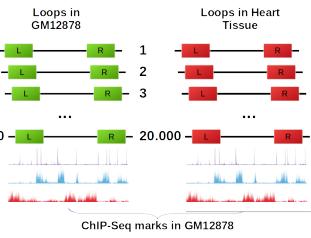
What marks are involved?

- ATF2
 - Acetylates histones H2B and H4 *in vitro*. Yeast homologues involved with histone deacetylation and correct heterochromatin formation.

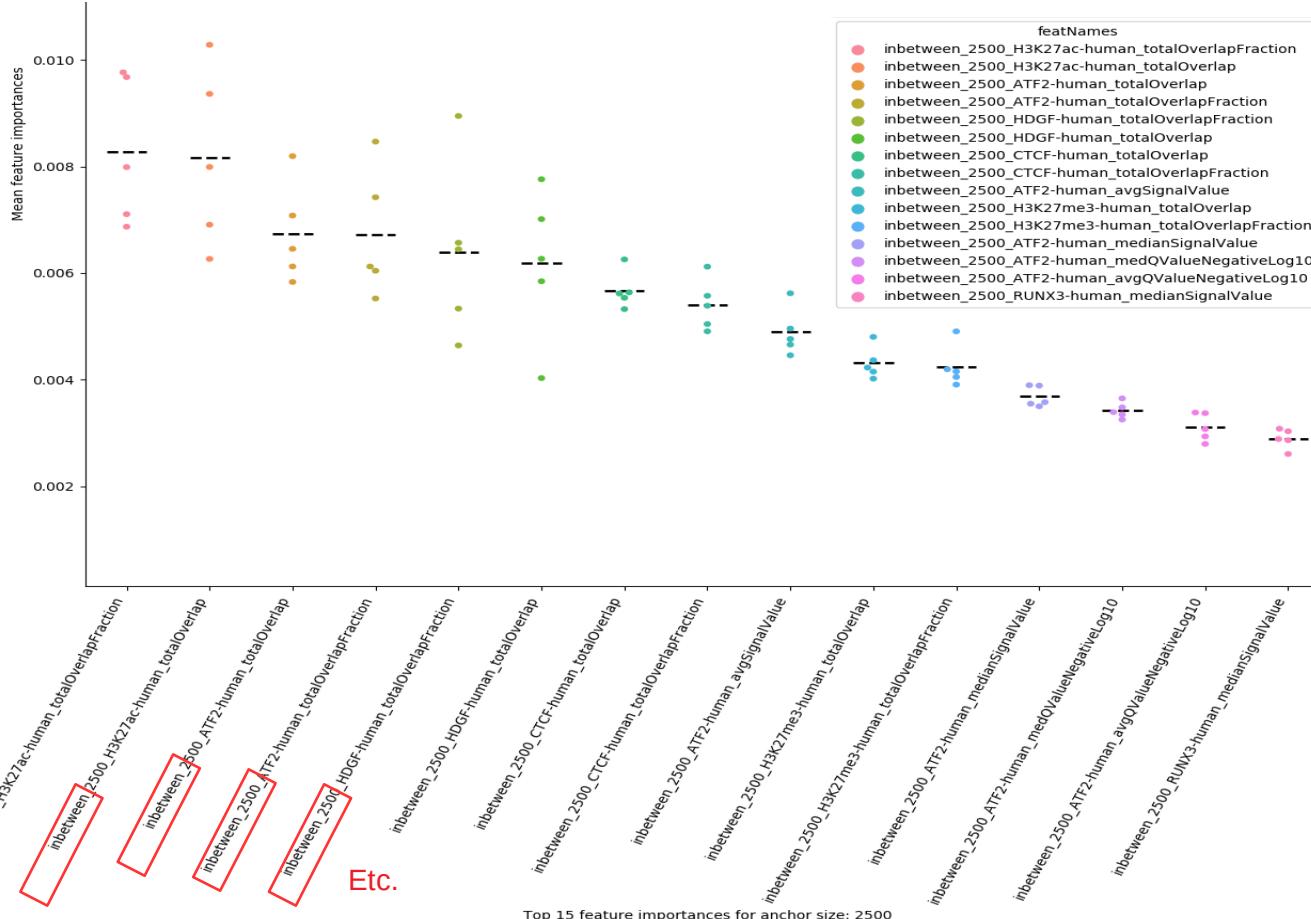
Lau, E., & Ze'ev, A. R. (2012); Kawasaki, H., et al. (2000).
- HDGF
 - Nuclear-targeted mitogen, involved in many cancers. No idea why it's here.
- CTCF

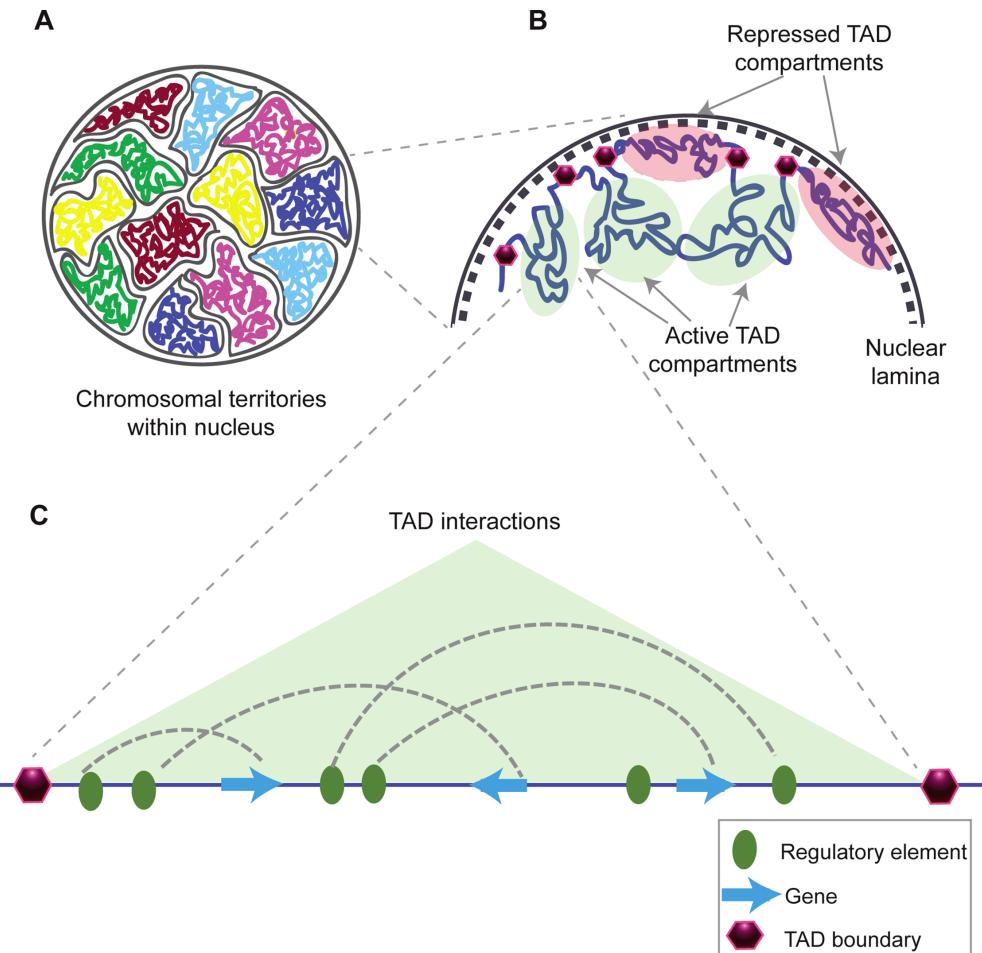
Thakar, K., et al. (2008); Everett, A. D., & Bushweller, J. (2003).





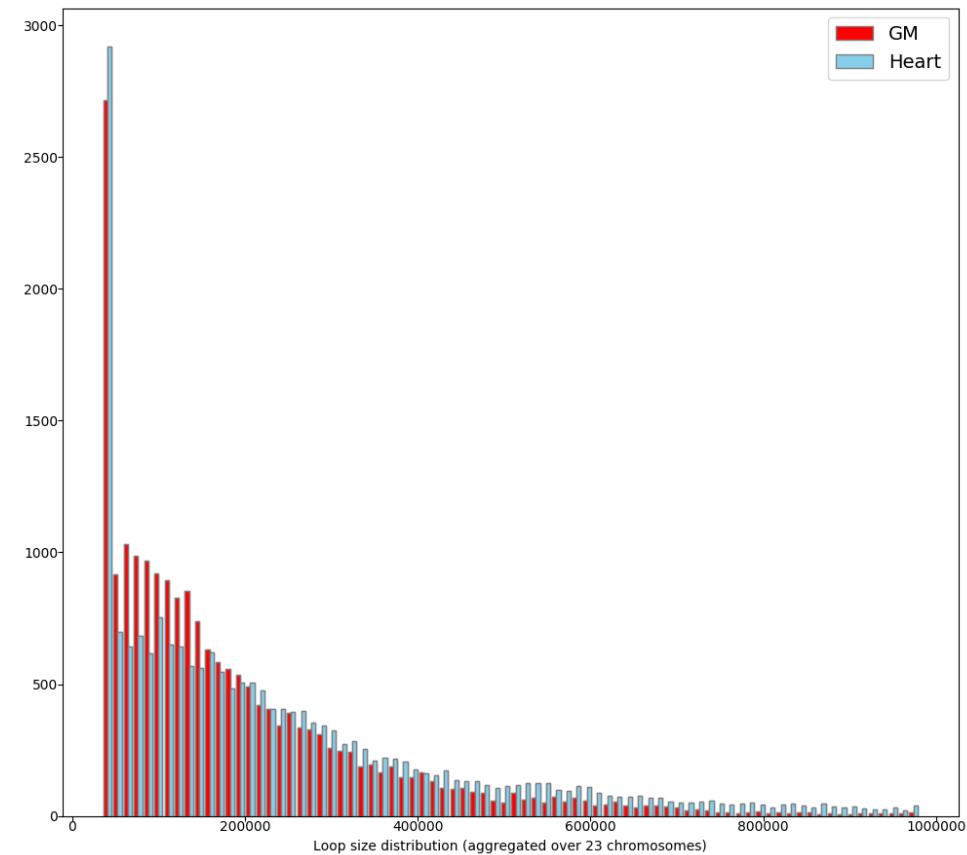
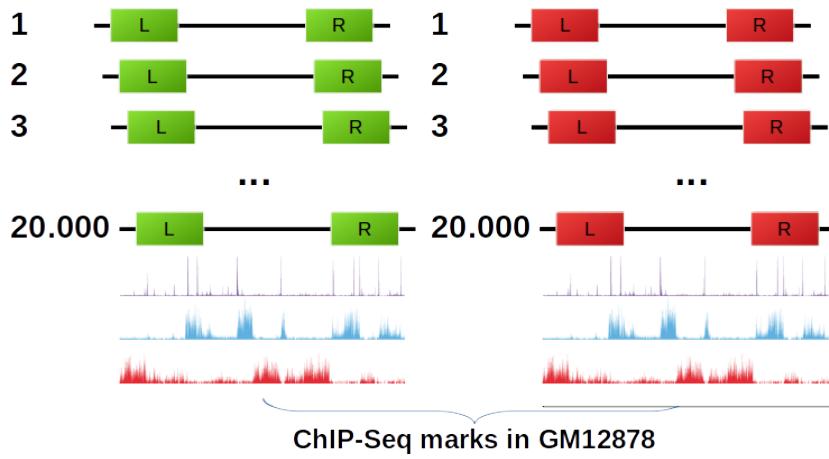
What location is used?





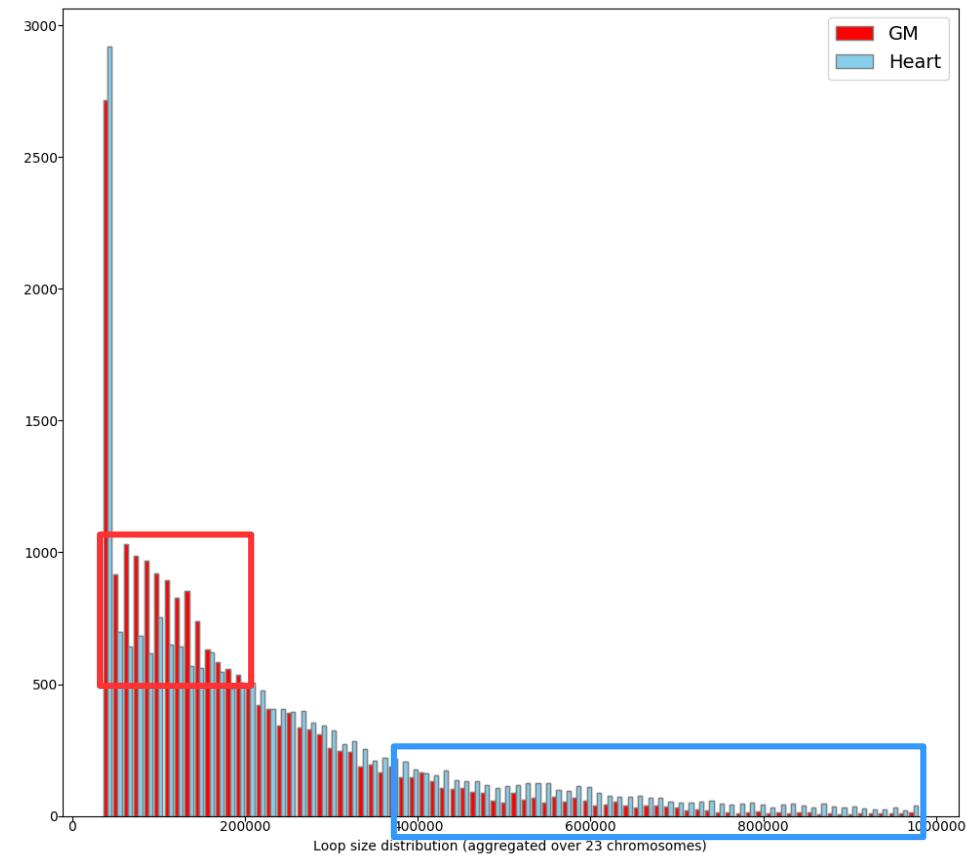
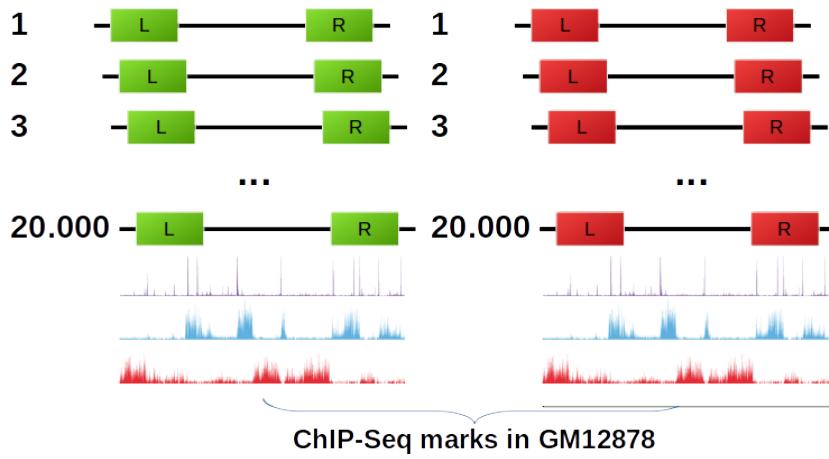
Loops in
GM12878

Loops in Heart
Tissue



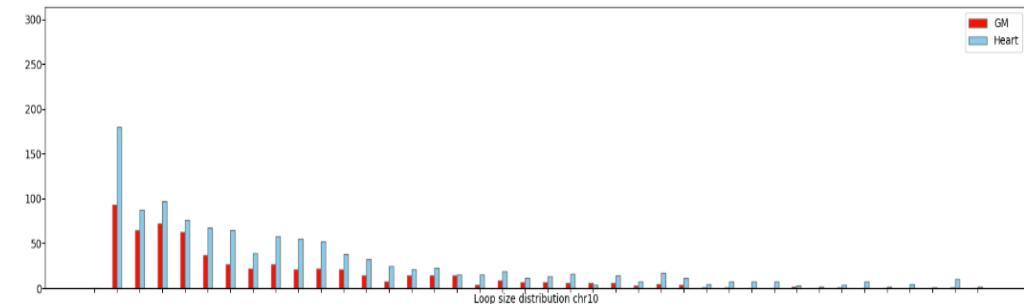
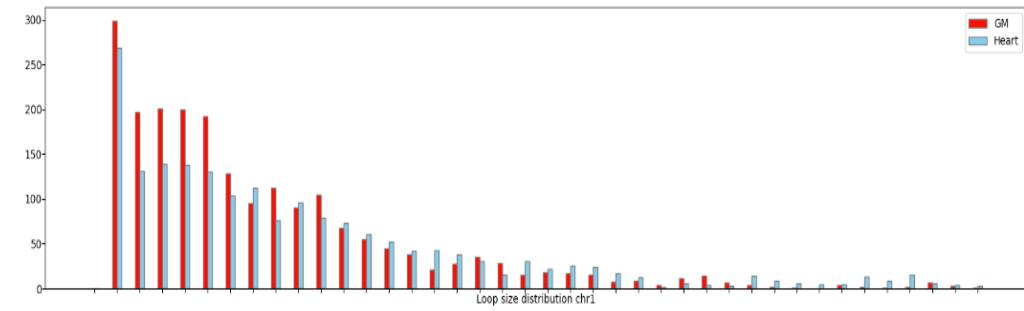
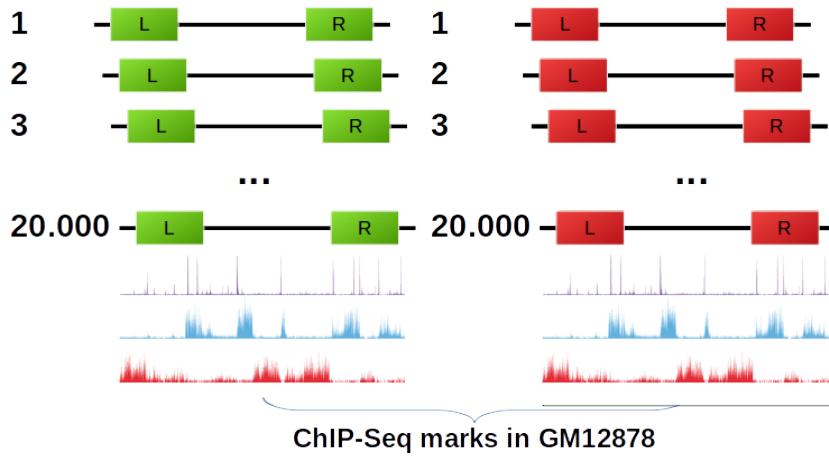
Loops in
GM12878

Loops in Heart
Tissue

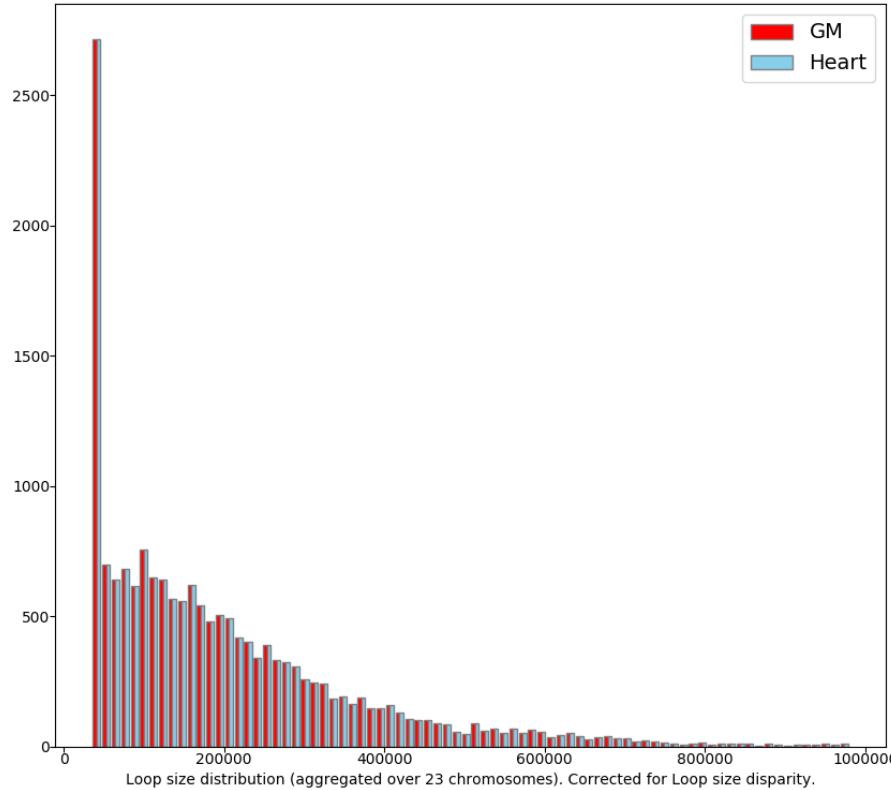


Loops in
GM12878

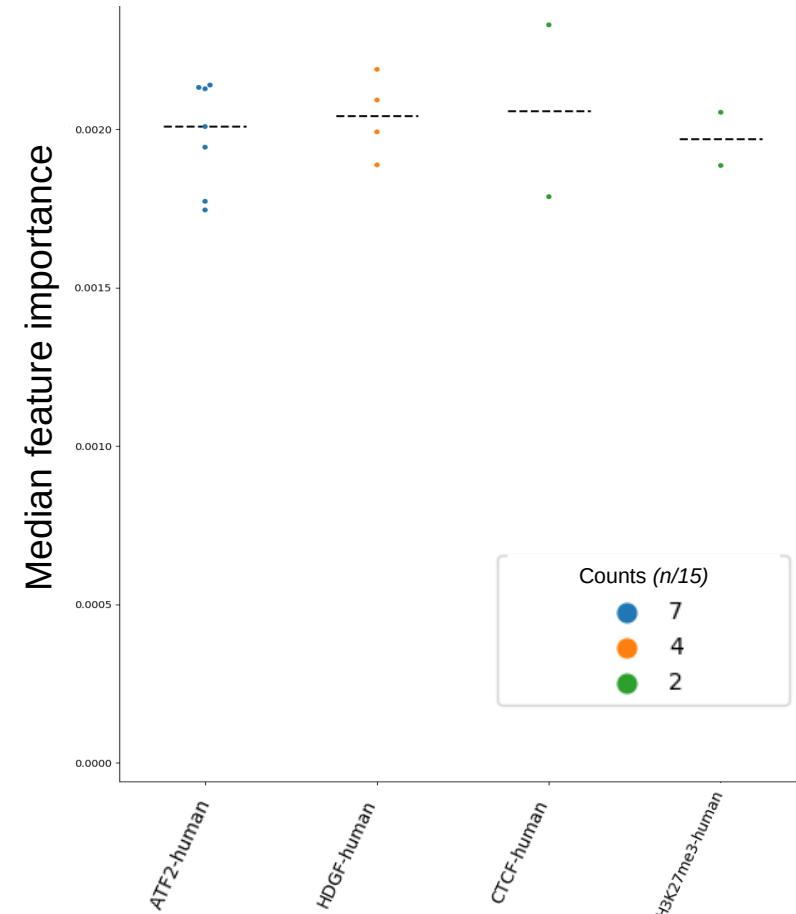
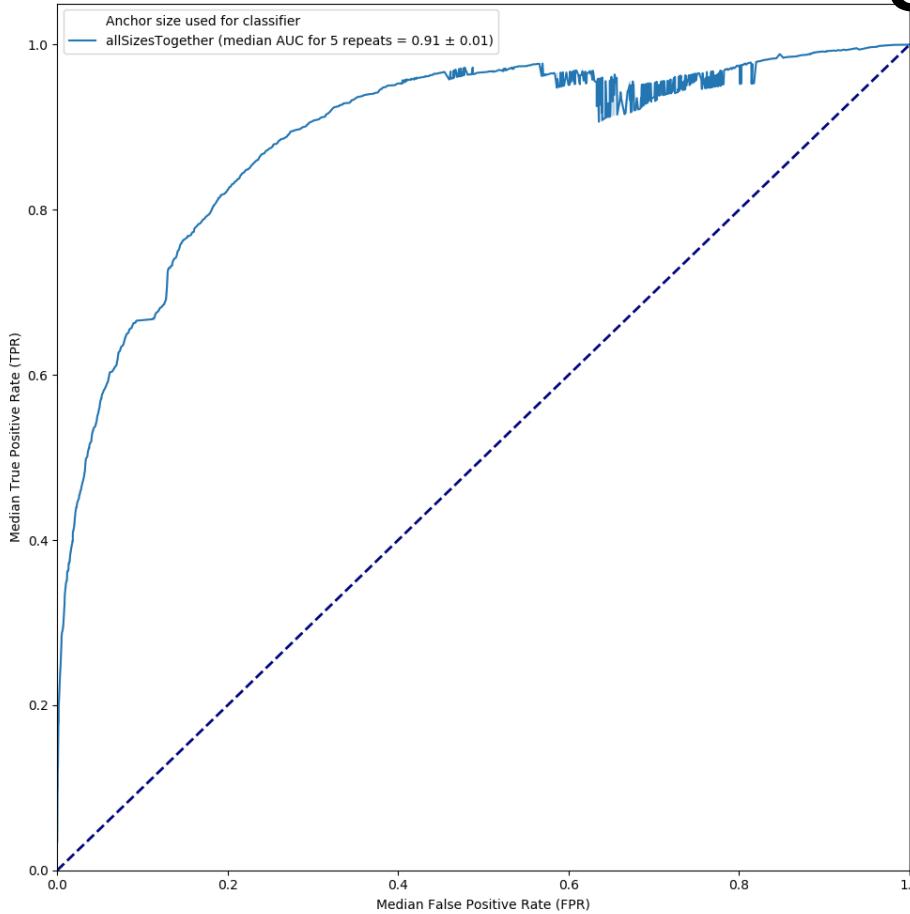
Loops in Heart
Tissue



Correct this by selective downsampling



Changed loop distributions: unchanged result

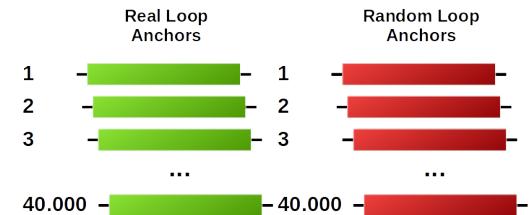


What marks distinguish present from possible loops?

- Combination of in-between marks focussing on the broader genomic area
 - Don't serve as proxies for loop size
- ATF2, HDGF and CTCF are important

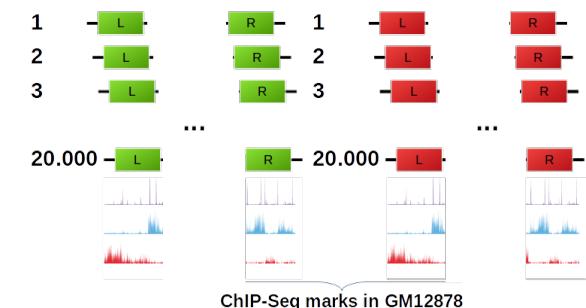
Conclusions

- Anchor locations have a (somewhat) distinct chromatin mark footprint



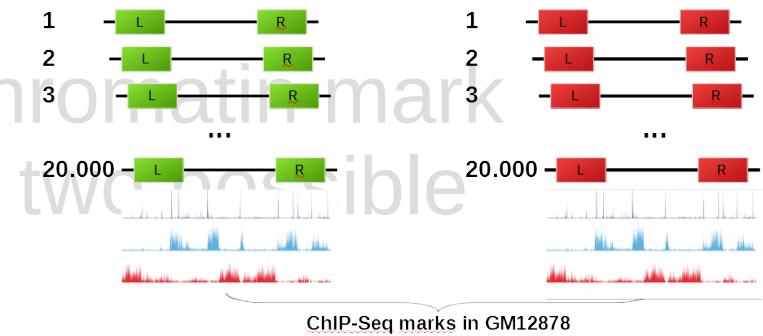
Conclusions

- Anchor locations have a (somewhat) distinct chromatin mark footprint
- There are no anchor-based chromatin mark combinations that predict that two possible anchors will interact



Conclusions

- Anchor locations have a (somewhat) distinct chromatin mark footprint
- There are no anchor-based chromatin mark combinations that predict that two possible anchors will interact
- Predicting loops with high certainty requires broad knowledge of surrounding (in-between) marks

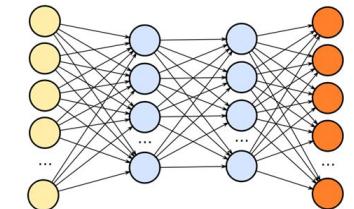


Improvements

- Divide the in-between area up into segments
 - See whether certain subsections are important
- Test the difference between split and non-split anchors
- Remove the top n features and see how performance reacts

Future work

- Deep learning:
 - Training samples cannot be solely focused on anchors and immediate surrounding area.
- Separate analysis of ATF-2 and HDGF presence near loops





That's all Folks!



Extra slides I: all feature names

intersectsPerFactor, totalOverlap,
totalOverlapFraction, avgSignalValue, medianSignalValue,
AvgQValueNegativeLog10, medQValueNegativeLog10,
sumPointPeaksInInterval

