

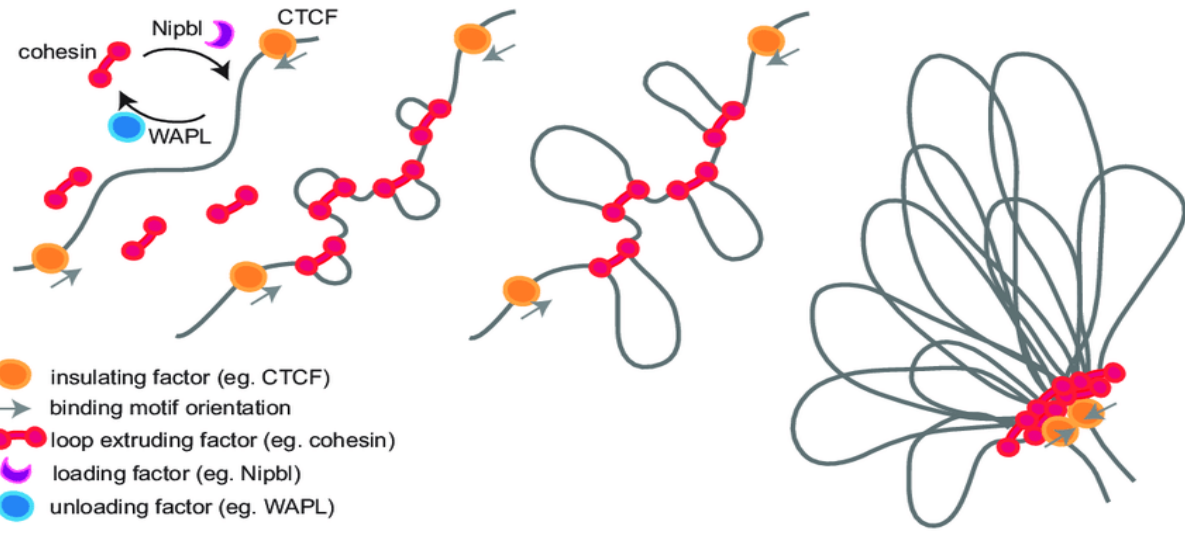
Роль ncRNA в структуре хроматина

Саша Галицына Данила Матвеев Павел Кузнецов

CTCF и EZH2 (PRC2) - это ключевые белки, которые участвуют в процессе преобразования хроматина и образовании TAD (Topologically associating domains). Мы предполагаем, что ncRNA (нкРНК) могут влиять на этот процесс, напрямую действуя на CTCF и EZH2. Мы планируем сопоставить данные различных клеточных линий Hi-C карт с данными ChIP-seq по CTCF и EZH2, после чего сравнить содержание ncRNA с данными ChIP-seq на масштабе всего генома.

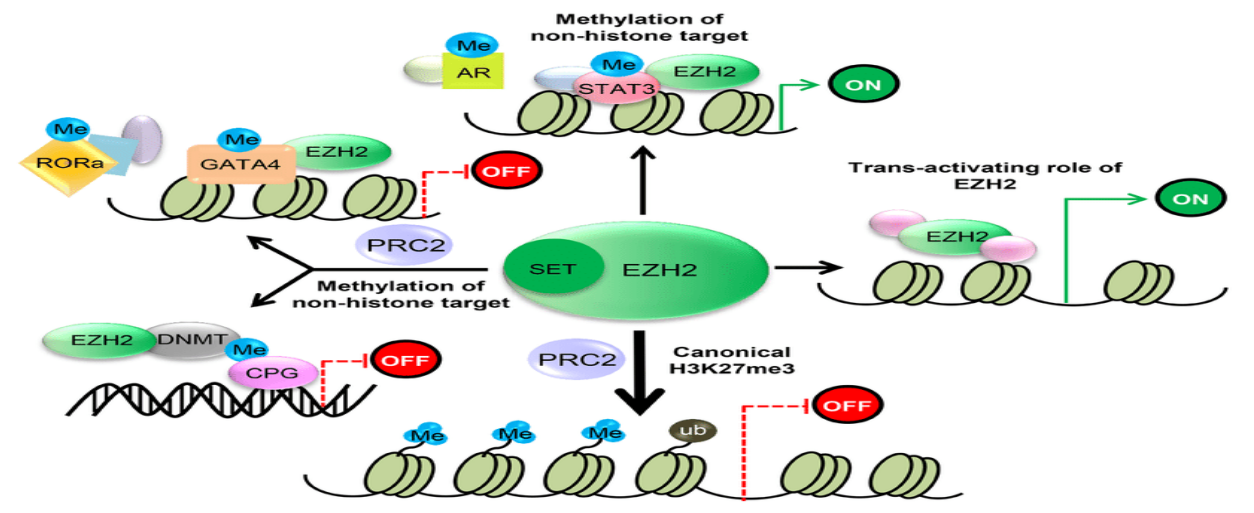
Используемые программы:

- Python • Pandas
- Numpy • Cooler / CoolTools
- Bioframe • HiGlass



Механизм работы CTCF

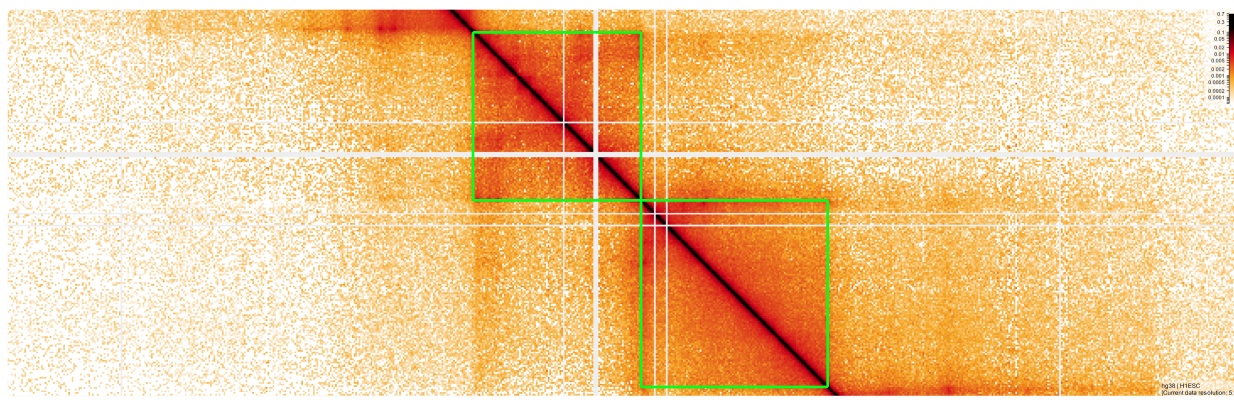
Defining Functionally Relevant Spatial Chromatin Domains: It is a TAD Complicated (Sikorska et al. 2019) - Scientific Figure on ResearchGate



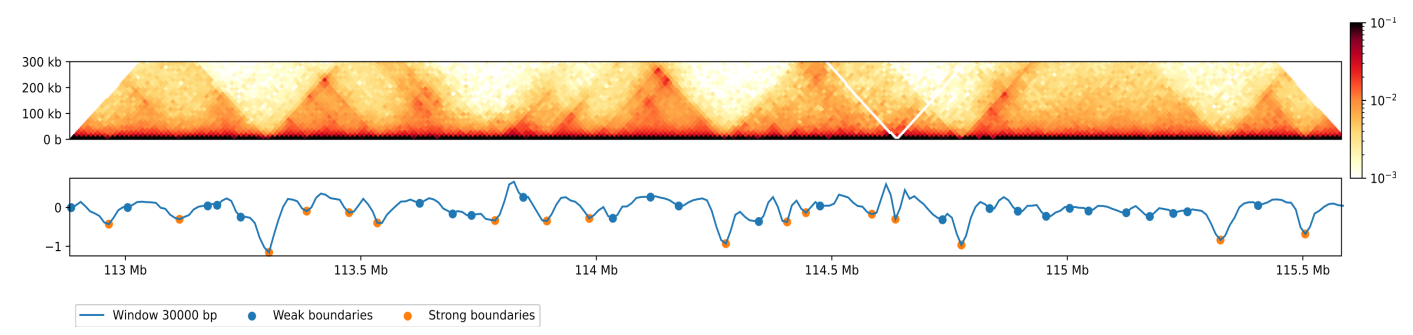
Механизмы работы EZH2

Role of EZH2 in cancer stem cells: From biological insight to a therapeutic target (Wen et al. 2015) - Scientific Figure on ResearchGate

Данные Hi-C по разным клеточным линиям (H1ESC, HUVEC, HeLa-S3, IMR-90, K562) были загружены при помощи cooler/cooltools, после чего границы доменов были найдены при помощи нахождения минимумов по таблице инсуляции при помощи NumPy и Pandas, границы, где показатели инсуляции являются локальными минимумами, разделены на слабые, внутри доменов которых мала вероятность найти большое количество ncRNA из-за их размера, и на сильные



Примеры доменов на Hi-C (H1ESC; chr11:112,884,917-116,554,270)



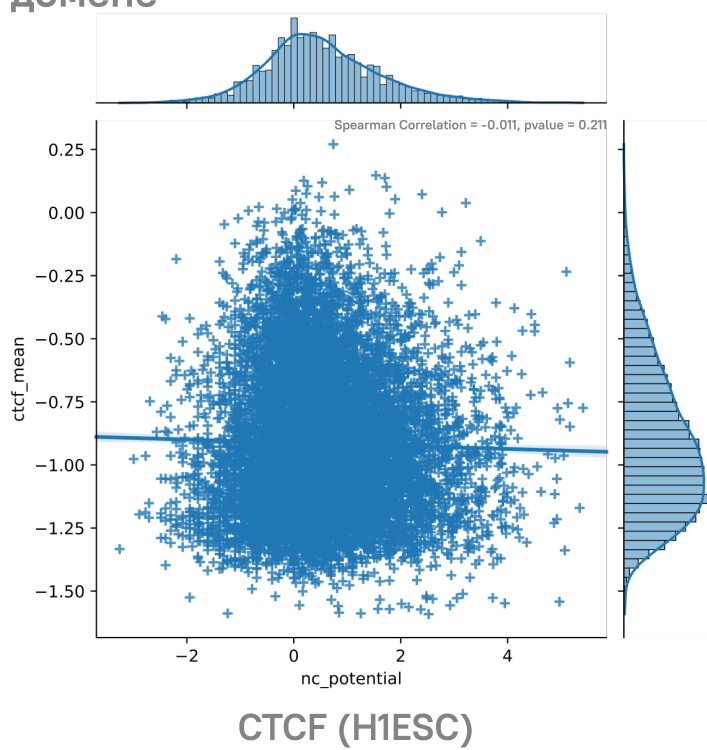
Просчет границ доменов на Hi-C (H1ESC; chr11:112,884,917-116,554,270)

После сохранения DataFrame-а с списком сильных границ доменов, мы обработали его при помощи BioFrame и получили итоговый список доменов, по которому далее мы провели поиск с использованием баз данных Ensembl (ncRNA/mRNA) и LNCipedia (ncRNA) и сформировали список ncRNA и mRNA для каждого домена в геноме

index_domain	chrom_domain	start_domain	end_domain	index_genemap	chrom_genemap	source_genemap	feature_genemap	start_genemap	end_genemap	score_genemap	strand_genemap	frame_genemap	attributes_genemap	l_distance_genemap	r_distance_genemap
0	8159	chr1	970000	1120000	1614	chr1	ensembl_havana	mRNA	966482	975865	.	+	ID=transcript:ENST00000379410;Parent=gene:ENSG...	3518	144135
1	8159	chr1	970000	1120000	1649	chr1	ensembl_havana	mRNA	966502	975008	.	+	ID=transcript:ENST00000379407;Parent=gene:ENSG...	3498	144992
2	8159	chr1	970000	1120000	1682	chr1	ensembl_havana	mRNA	966502	975008	.	+	ID=transcript:ENST00000379409;Parent=gene:ENSG...	3498	144992
3	8159	chr1	970000	1120000	1715	chr1	havana	lncRNA	970875	971523	.	+	ID=transcript:ENST00000480267;Parent=gene:ENSG...	875	148477
4	8159	chr1	970000	1120000	1719	chr1	havana	mRNA	973512	975865	.	+	ID=transcript:ENST00000491024;Parent=gene:ENSG...	3512	144135
...
449752	12251	chrY	21380000	22280000	3365084	chrY	havana	lncRNA	22101692	22147484	.	-	ID=transcript:ENST00000419158;Parent=gene:ENSG...	721692	132516
449753	12251	chrY	21380000	22280000	3365099	chrY	havana	lncRNA	22144966	22146831	.	+	ID=transcript:ENST00000253848;Parent=gene:ENSG...	764966	133169
449754	12251	chrY	21380000	22280000	3365104	chrY	ensembl_havana	mRNA	22168542	22182923	.	-	ID=transcript:ENST00000303766;Parent=gene:ENSG...	788542	97077
449755	12251	chrY	21380000	22280000	3365131	chrY	havana	lncRNA	22168542	22182957	.	-	ID=transcript:ENST00000481858;Parent=gene:ENSG...	788542	97043
449756	12251	chrY	21380000	22280000	3365143	chrY	ensembl	mRNA	22168542	22182982	.	-	ID=transcript:ENST00000454978;Parent=gene:ENSG...	788542	97018

DataFrame с списком mRNA и ncRNA для каждого домена в геноме (H1ESC)

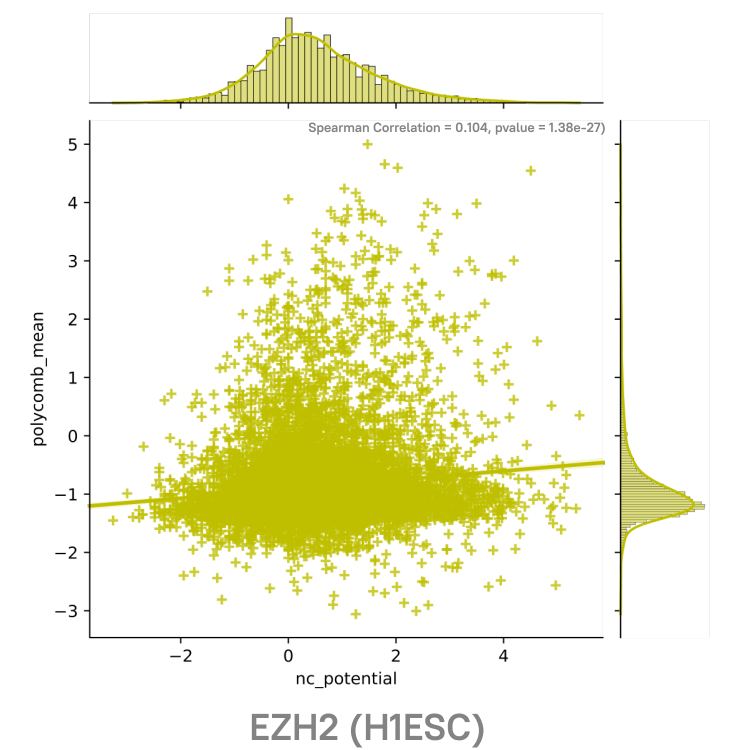
При помощи PyBigWig мы загрузили данные ChIP-seq по CTCF и EZH2, после чего просчитали для каждого домена средние показатели ChIP-seq и загрузили в отдельную таблицу, с которой построили графики отношения ChIP-seq (ctcf_mean/polycomb_mean) и плотности ncRNA (density_lncrna) в домене на плотности mRNA (density_gene) в домене



CTCF (H1ESC)

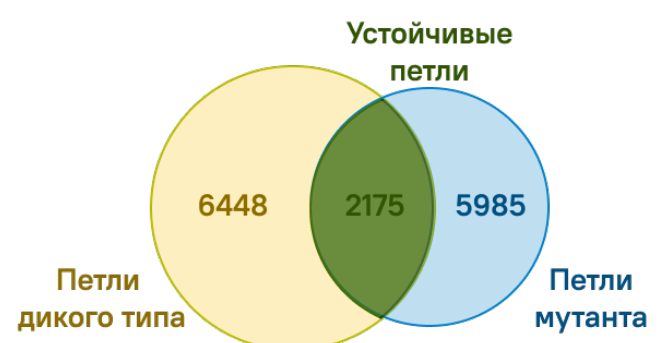
chrom	start	end	count_lncrna	count_genes	ctcf_mean	polycomb_mean	domain_len	density_lncrna	density_gene	nc_potential	
1	chr9	38720000	68590000	959	77	0.203663	0.081295	29870000	0.000032	2.577837e-06	12.454545
3	chr1	119730000	145740000	319	91	0.204197	0.046993	26010000	0.000012	3.498654e-06	3.505495
4	chr3	90560000	93990000	9	31	0.204218	0.060260	3430000	0.000003	9.037901e-06	0.290323
5	chr5	45880000	50550000	7	3	0.204865	0.064376	4670000	0.000001	6.423983e-07	2.333333
6	chr18	14080000	21190000	140	13	0.207475	0.049546	7110000	0.000020	1.828411e-06	10.769231
...
13527	chr8	144450000	144530000	68	40	1.126578	0.711886	80000	0.000850	5.000000e-04	1.700000
13528	chr5	177370000	177460000	35	29	1.134468	1.665969	90000	0.000389	3.222222e-04	1.206897
13529	chr16	68440000	68530000	11	2	1.146994	0.669946	90000	0.000122	2.222222e-05	5.500000
13530	chr16	4250000	4350000	79	17	1.158886	0.449386	100000	0.000790	1.700000e-04	4.647059
13531	chr17	45130000	45210000	44	21	1.311135	0.401478	80000	0.000550	2.625000e-04	2.095238

Plotting DataFrame (H1ESC)

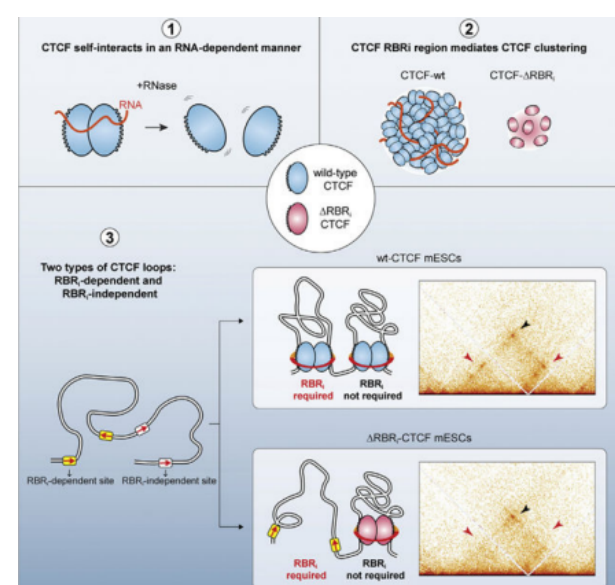


EZH2 (H1ESC)

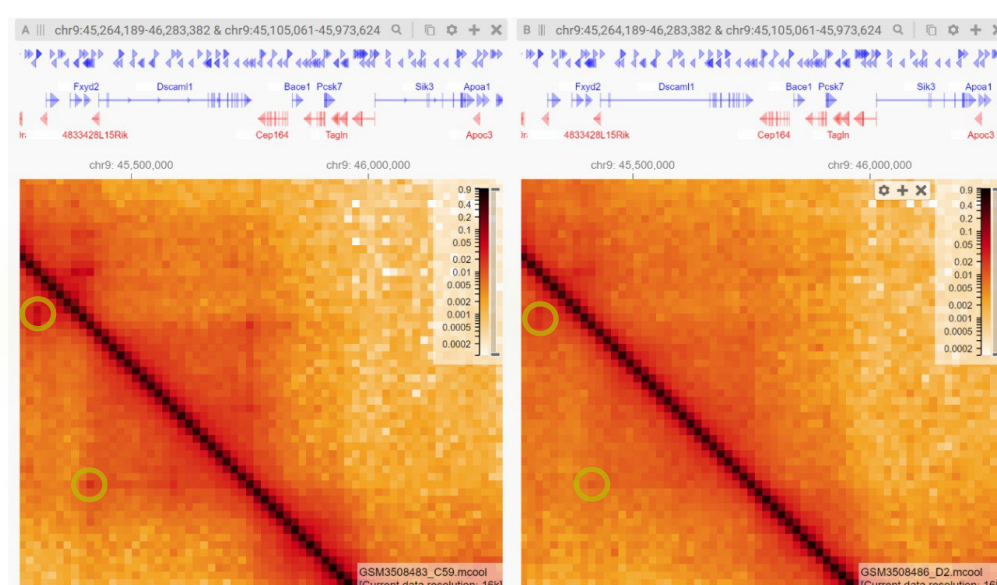
CTCF - это ДНК-связывающий белок, имеющий РНК-связывающие (RBR) сайты. Не так давно стало известно, что структура хроматина зависит от локализации CTCF в ДНК, а значит, и от связывания РНК. Взяв данные Hi-C для клеток мыши дикого типа и мутантных по RBR-CTCF, мы решили исследовать эту зависимость подробнее.



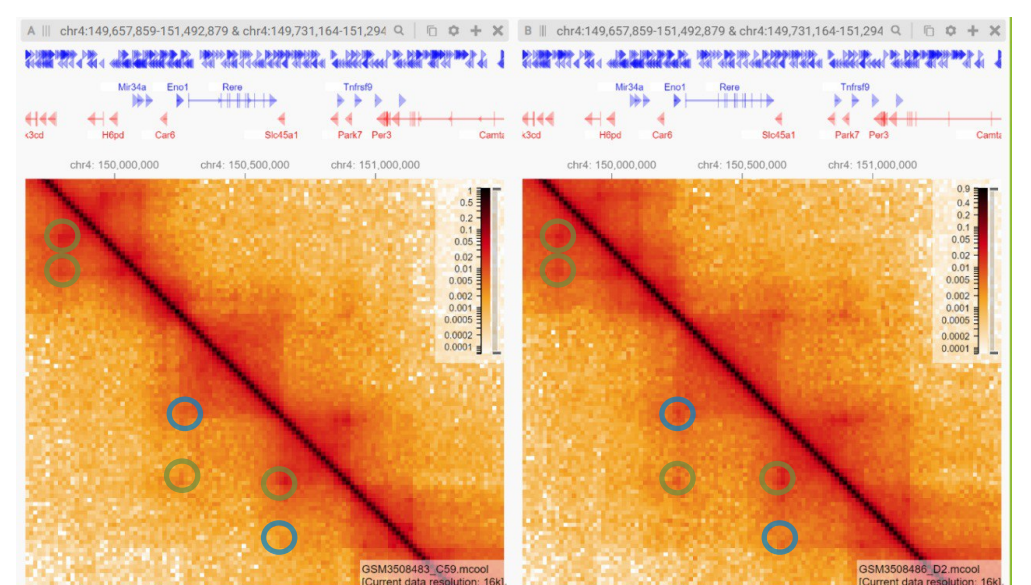
Классы петель Hi-C



Механизм работы CTCF



Примеры петель на регионе хромосомы 9



Примеры петель разных классов на регионе хромосомы 4

Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF (Hansen et al. 2019)

