

Создания, которых вы зовете мышами, вовсе не таковы, какими кажутся вам. Вы замечаете лишь, так сказать, отпечаток в нашем измерении огромных сверхразумных панпространственных существ.

Дуглас Адамс, Автостопом по галактике

Проект по майнору “Биоинформатика”

Поиск и изучение участков генома, где H3K27me3
присутствует в местах образования ZDNA

Выполнили студенты:

Плечова Е.П. (human, SK-N-SH)

Романченко П.М. (human, GM12878)

Гудиев М.А. (human, H1)

Грачёв Д.В. (human, H7)

Булатова Е.В. (human, MCF-7)

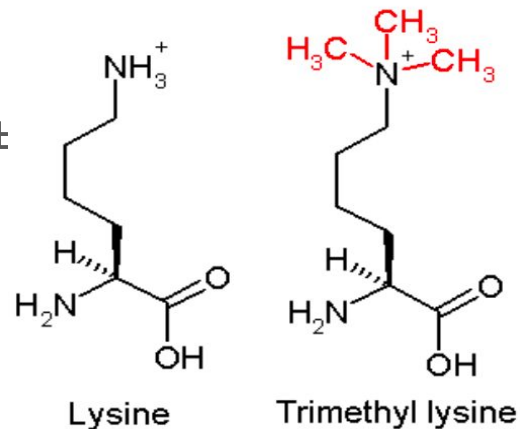
Аноприенко М.В. (human, H1)

Антонов Д.А. (mouse, C2C12)

Сибгатов С.И. (mouse, MEL)

Введение

H3K27me3 - гистоновая метка, отвечающая 3-метилированию 27-го лизина 3-го гистона. Регионы богатые метками H3K27me3 могут выполнять функцию сайленсеров при хроматиновых взаимодействиях.

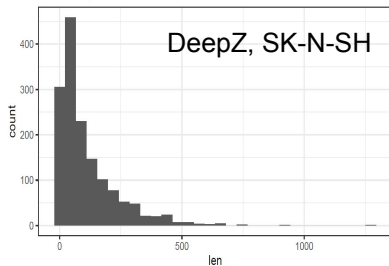


Обзор исходных файлов

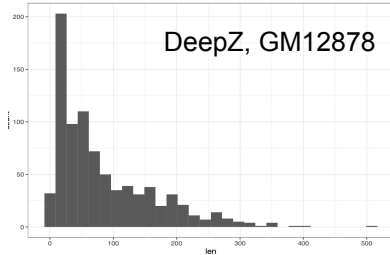
Фамилия	Организм	Тип клетки	Эксперимент вторичной структуры	Кол-во пиков метки	Кол-во пиков ZDNA	Кол-во пиков пересечения
Плечова	human	SK-N-SH	DeepZ	96017	19394	1515
Романченко	human	GM12878	DeepZ	41772	19394	837
Анопренко	human	H1	DeepZ	39125	19394	2404
Грачев	human	H7	DeepZ	83301	19394	2788
Гудиев	human	H1	ZHunt	53995	412600	13334
Булатова	human	MCF-7	ZHunt	79788	412600	9801
Сибгатов	mouse	MEL	mouse_1	72954	3258	42/19 (mm10/hg19)
Антонов	mouse	C2C12	mouse_2	82080	1651	12/6 (mm10/hg19)

Гистограммы длин пиков пересечений H3K27me3 и ZDNA

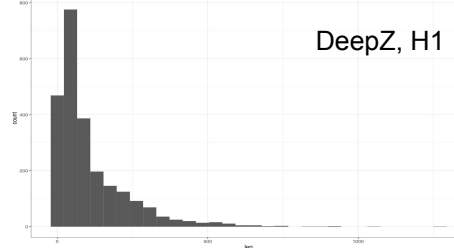
H3K27me3_SK-N-SH.intersect_with_DeepZ
Number of peaks = 1515



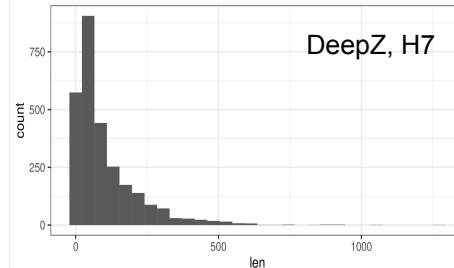
H3K27me3_ZDNA.intersect_with_DeepZ GM12878
Number of peaks = 837



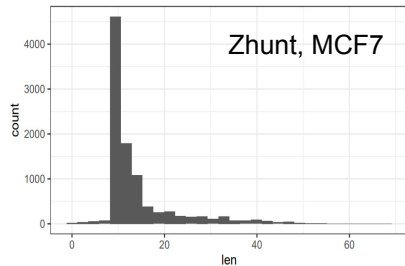
H3K27me3_H1.intersect_with_DeepZ
Number of peaks = 2404



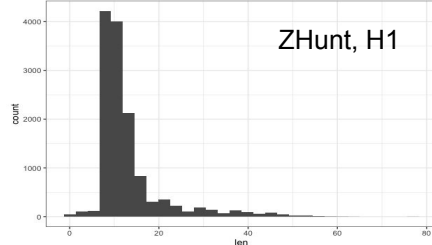
intersect_with_DeepZ
Number of peaks = 2788



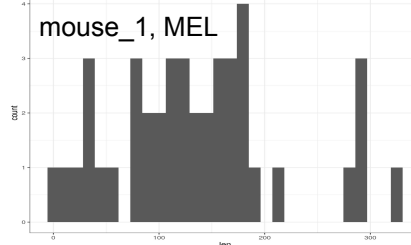
intersect_with_zhunt
Number of peaks = 9801



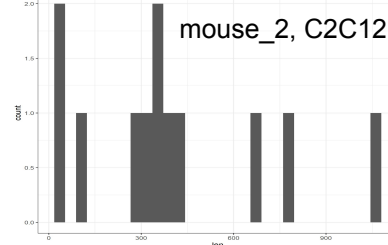
zhunt.intersect
Number of peaks = 13334



H3K27me3_MEL.intersect_with_mouseZ-DNA1
Number of peaks = 42

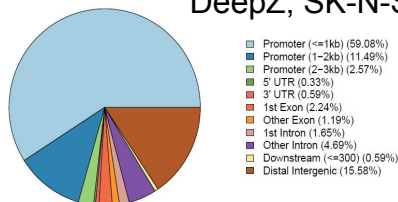


H3K27me3_C2C12.intersect_with_DeepZ
Number of peaks = 12

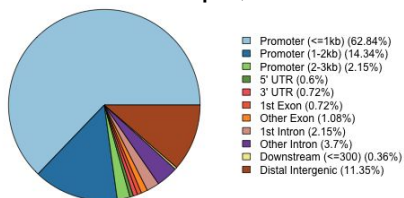


Расположение intersect-пиков относительно аннотированных генов

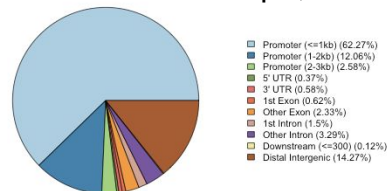
DeepZ, SK-N-SH



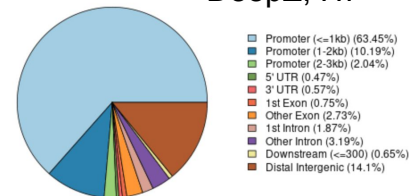
DeepZ, GM12878



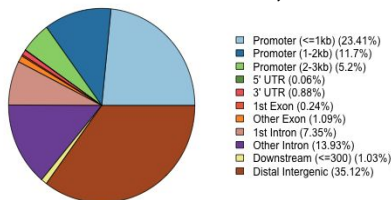
DeepZ, H1



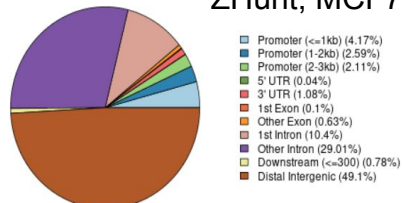
DeepZ, H7



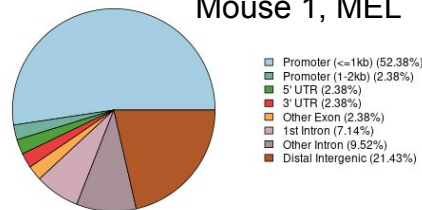
ZHunt, H1



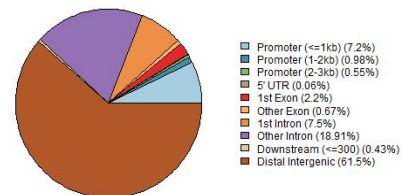
ZHunt, MCF7



Mouse 1, MEL



Mouse 2, C2C12



Техническая сторона того, что было сделано с исходными intersect-файлами

Объединили все файлы в один большой файл, после чего сгенерировали все возможные пересечения. Все самые интересные мы сохранили на гитхаб (в том числе всех, кроме мышей)

Пример конвертации мышей:

1. `wget`
<https://hgdownload.cse.ucsc.edu/goldenpath/mm10/liftOver/mm10ToHg19.over.chain.gz>
2. `liftOver H3K27me3_MEL.intersect_with_mouseZ-DNA1.bed mm10ToHg19.over.chain.gz H3K27me3_MEL_mm10tohg19.intersect_with_mouseZ-DNA1.bed _unmapped.txt`

Как пересекались интерсект файлы:

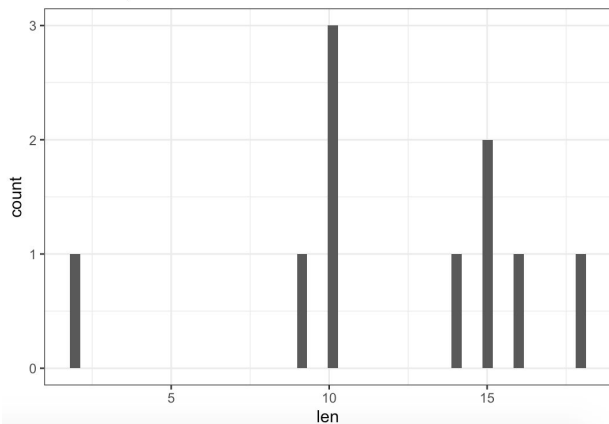
1. По всем возможным подмножествам -- нашли, что ZHunt и DeepZ лучше рассматривать отдельно
2. Все без мышей -- вручную с помощью `bedtools intersect`

Приводим статистику того, сколько пиков получилось после тотально пересечения

После пересечения всего, кроме мышей – 10 пиков

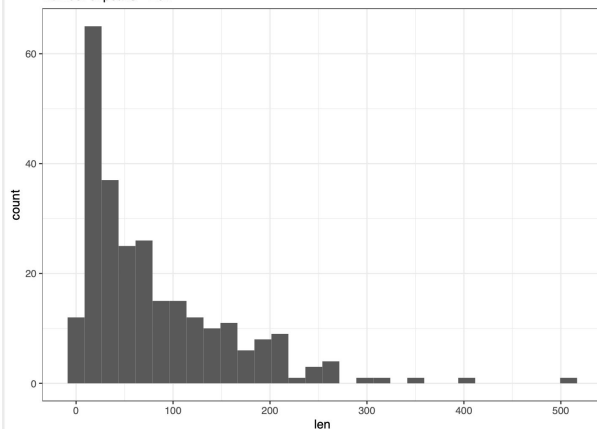
All 6 human intersections

Number of peaks = 10



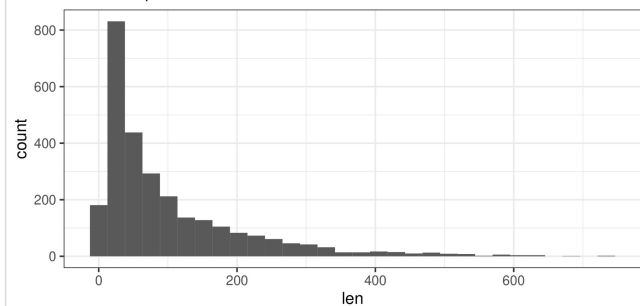
DeepZ intersections

Number of peaks = 264



ZHunt.intersection

Number of peaks = 2781

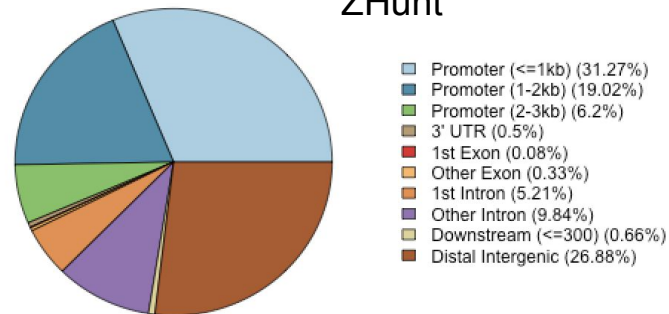


Круговые диаграммы расположения относительно аннотированных генов

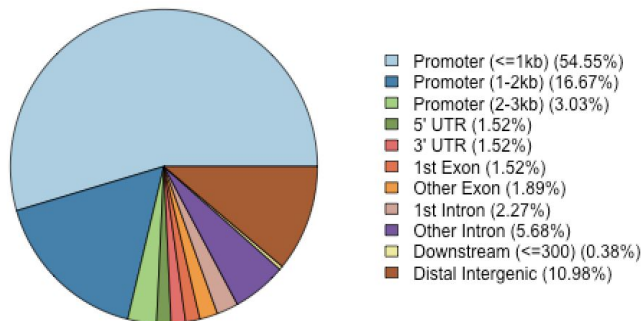
6 типов клеток для человека



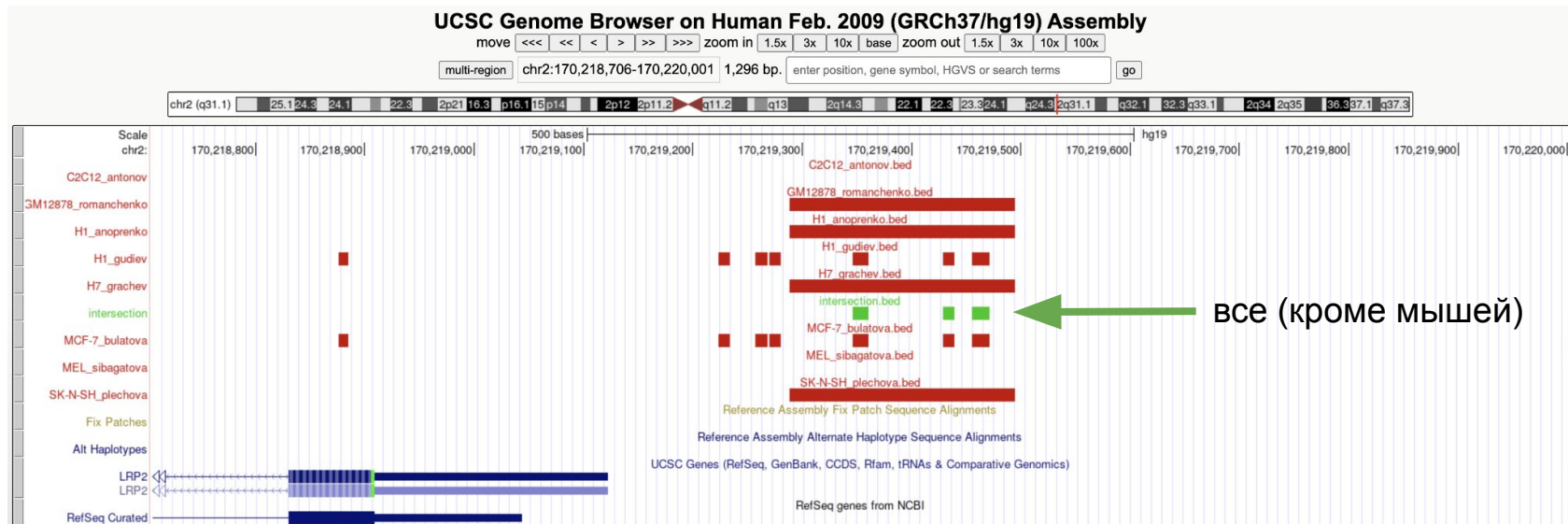
ZHunt



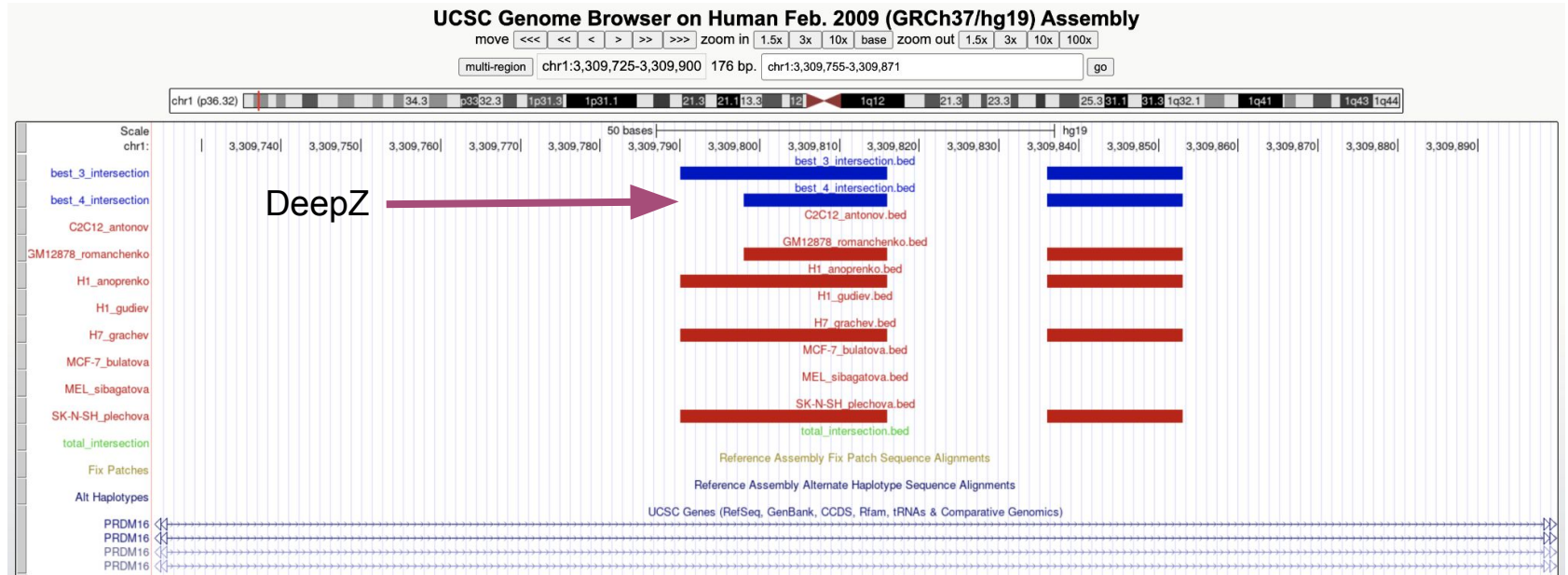
Только DeepZ



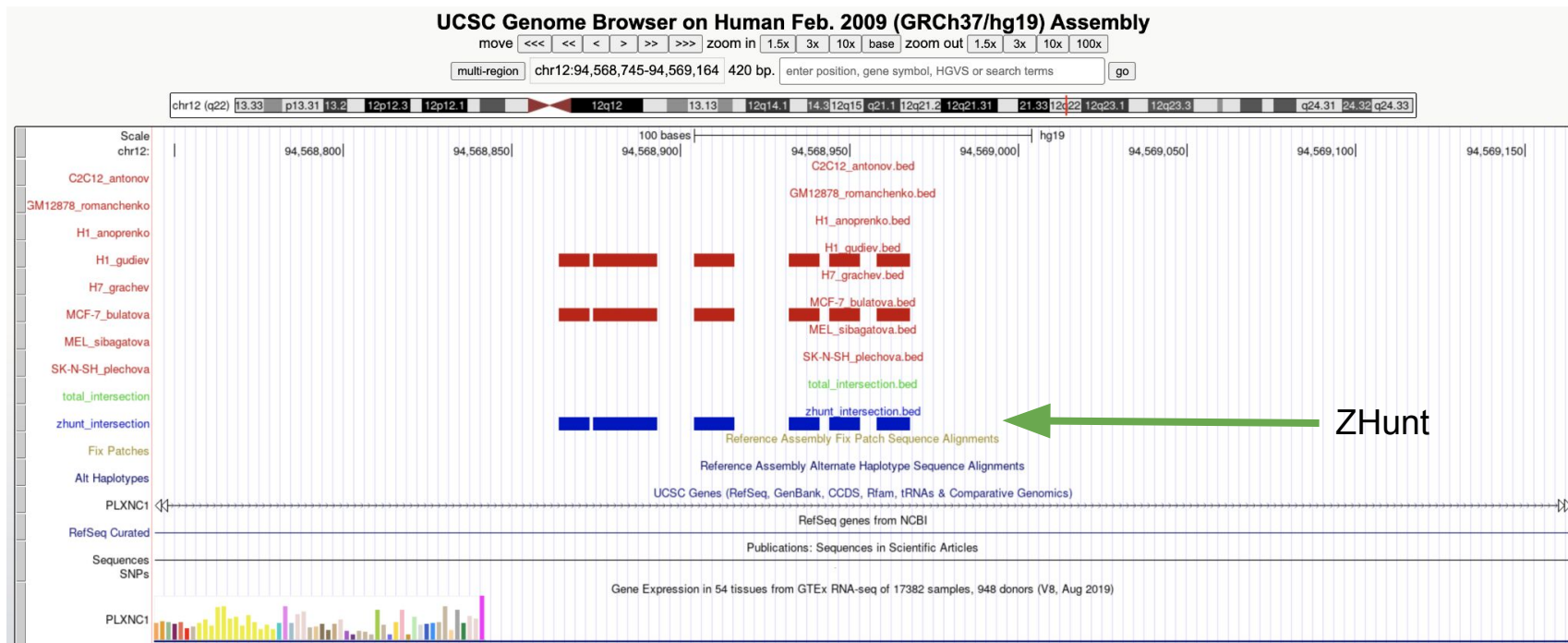
Визуализация в геномном браузере




Визуализация в геномном браузере



Визуализация в геномном браузере



Количество пиков, которые удалось ассоциировать с генами

Эксперимент вторичной структуры	Исходное кол-во пиков	Кол-во проассоциированных генов	Кол-во уникальных генов
DeepZ	264	98	74
ZHunt	2781	271	158
Мыши	0	0	0 

Наиболее значимые GO-категории для нашего списка генов

	Homo sapiens (REF)	Client Text Box Input (▼ Hierarchy_ NEW! ?)					
GO biological process complete	#	#	expected	Fold Enrichment	+/-	raw P value	FDR
kidney development	282	8	.93	8.59	+	4.87E-06	7.70E-02
↳ renal system development	291	8	.96	8.33	+	6.09E-06	4.82E-02

(на DeepZ)

	Homo sapiens (REF)	Client Text Box Input (Hierarchy_) NEW! (?)					
GO biological process complete	#	#	expected	Fold Enrichment	+/-	raw P value	▲ FDR
anatomical structure development	5330	88	38.04	2.31	+	7.67E-18	6.06E-14
multicellular organism development	4941	84	35.27	2.38	+	1.40E-17	7.38E-14
system development	4330	79	30.91	2.56	+	4.73E-18	7.48E-14
developmental process	5797	91	41.38	2.20	+	3.67E-17	1.45E-13
multicellular organismal process	6926	98	49.44	1.98	+	4.87E-16	1.54E-12

(на ZHunt)

Выводы

1. Методы ZHunt и DeerZ дают разные положения пиков, которые имеют очень маленькое пересечение. Методы слабо согласуются друг с другом. Пики в hg19 и mm10 также почти не имеют пересечений, согласования нет.
2. Для разных типов клеток в методе DeerZ преимущественно пики в пересечении со вторичной структурой находятся на промоторах. В ZHunt преобладает ассоциация с межгенным пространством.
3. В случае ZHunt наблюдается более сильная ассоциация вторичной структуры ДНК с пиками гистоновых меток, чем в случае DeerZ. Средняя доля пиков пересечения к общему числу пиков составляет 0.0325 для DeerZ и 0.185 для ZHunt.
4. Данный эффект консервативен по виду клеток.