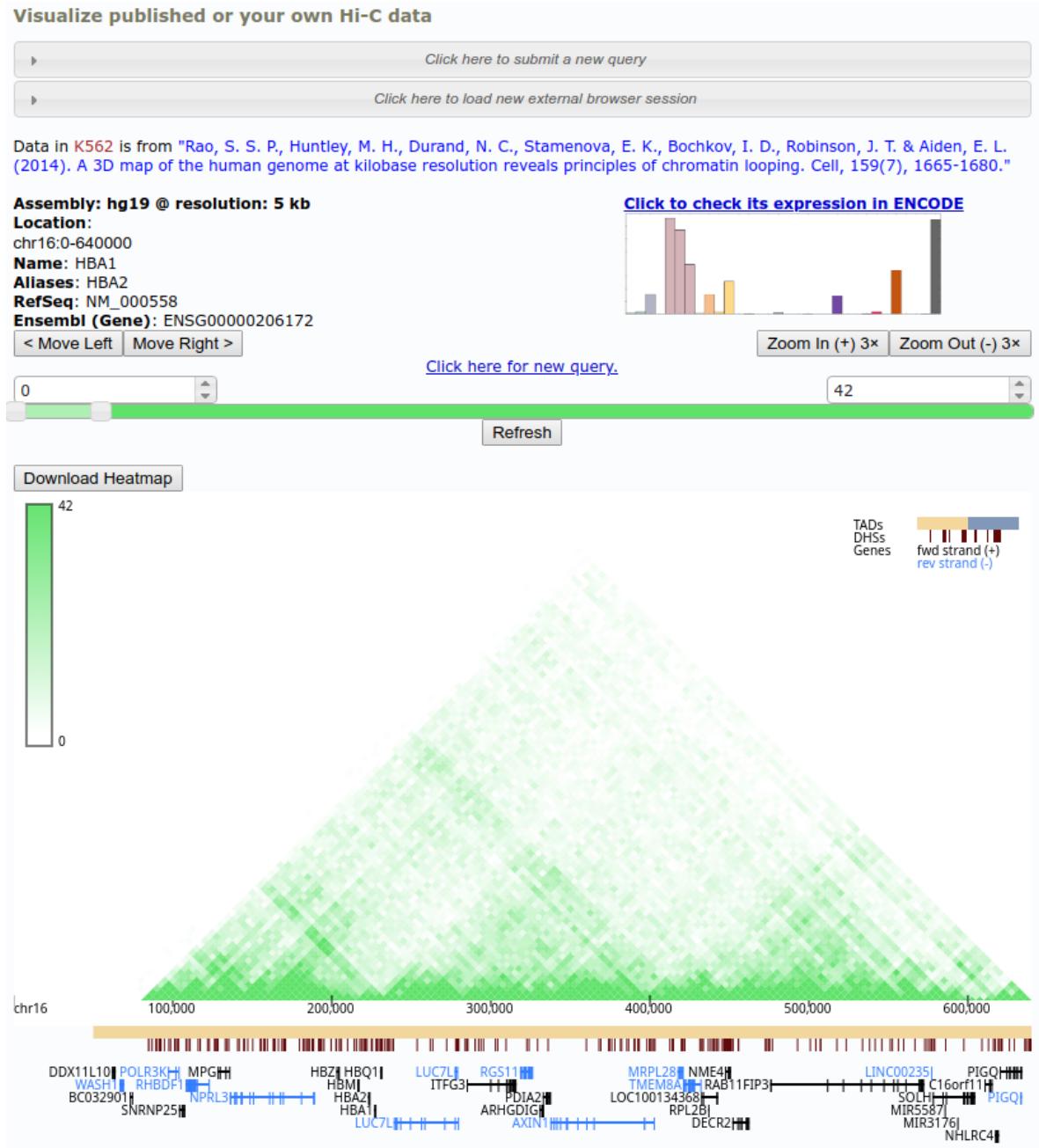


Chromatin Structure Analysis

Лев Мазаев | мАДБМ18

Задание 1. Yue lab browser

а:



Visualize published or your own Hi-C data

▶

[Click here to submit a new query](#)

▶

[Click here to load new external browser session](#)

Data in K562 is from "Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T. & Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7), 1665-1680."

Assembly: hg19 @ resolution: 5 kb

Location:

chr16:135000-480000

< Move Left | Move Right >

Zoom In (+) 3x | Zoom Out (-) 3x

[Click here for new query.](#)

2 ▲ ▼

100 ▲ ▼

Refresh

[Download Heatmap](#)

100
2

TADs
DHSS
Genes
fwd strand (+)
rev strand (-)



▼ Load External Browser

[Click here to load UCSC or WashU Epigenome Browser](#)

▼ Optional: Color Options

Color 1 Refresh

The browser is developed with Google Chrome 42*. While we test our browser on a variety of web browsers and operating systems, we recommend our users to use Chrome for the most optimal experience.

HBA1 находится близко к границе двух визуально выделяющихся треугольников, это могут быть неаннотированные ТАД-ы меньшего размера. Теперь загрузим данные тех же клеток, но с использованием VC-коррекции (b):

Visualize published or your own Hi-C data

▶

[Click here to submit a new query](#)

▶

[Click here to load new external browser session](#)

Data in K562 is from "Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S. & Aiden-Lieberman, E. 2014.. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell.* 2014;159(7):1665-1680."

Assembly: hg19 @ resolution: 5 kb

Location:

chr16:0-1925000

< Move Left | Move Right >

[Zoom In \(+\) 3x](#) | [Zoom Out \(-\) 3x](#)

[Click here for new query.](#)

0

10

Refresh

[Download Heatmap](#)

10
0

TADs
DHSs
Genes
fwd strand (+)
rev strand (-)



ТАДы стало видно лучше, но пропала их аннотация. Увеличим участок с геном HBA1:

Visualize published or your own Hi-C data

[Click here to submit a new query](#)

[Click here to load new external browser session](#)

Data in K562 is from "Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S. & Aiden-Lieberman, E. 2014,. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell.* 2014;159(7):1665-1680."

Assembly: hg19 @ resolution: 5 kb

Location:

chr16:90000-560000

< Move Left | Move Right >

Zoom In (+) 3x | Zoom Out (-) 3x

1

[Click here for new query.](#)

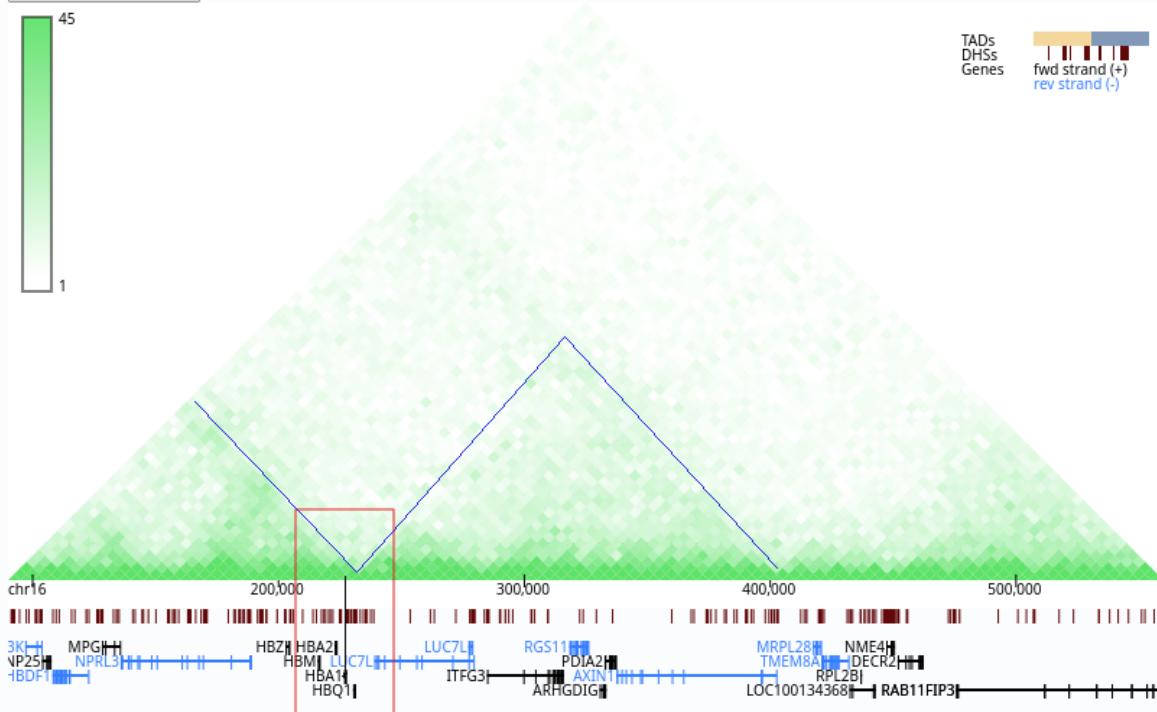
45

Refresh

[Download Heatmap](#)

TADs
DHSs
Genes

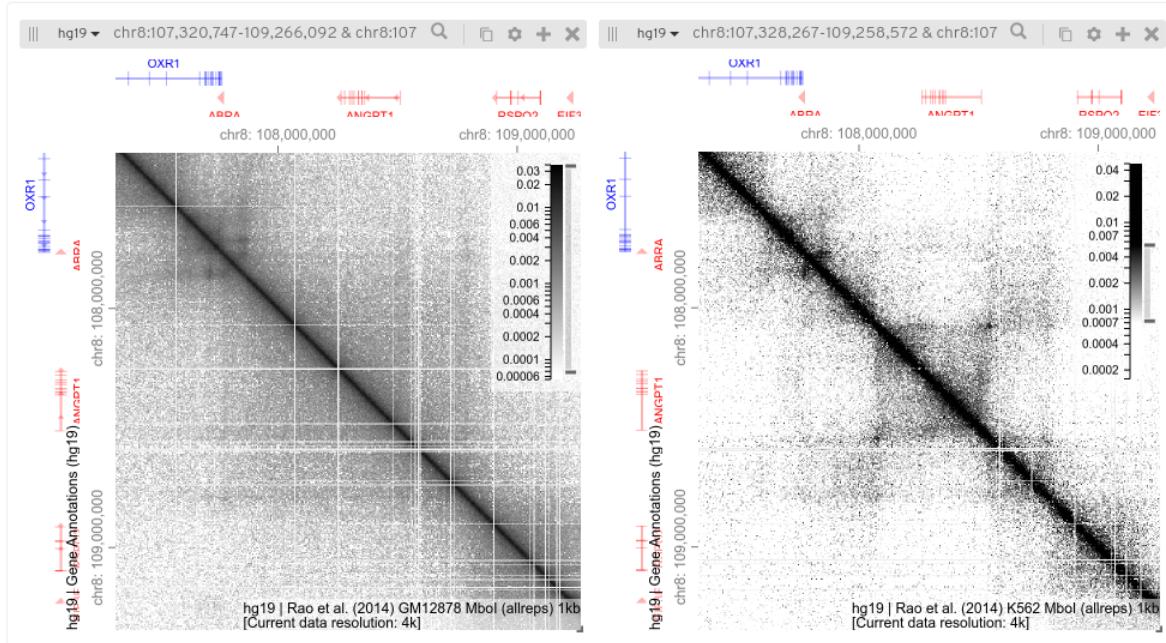
fwd strand (+)
rev strand (-)



Видим, что он расположен на границе двух визуально выделяющихся ТАД-ов. (с) Судя по разметке DHS, рядом с этим геном очень много сайтов, чувствительных к ДНК-азе, при этом сам ген находится между такими сайтами, значит он в области закрытого хроматина, а слева и справа много открытых участков. (д) Учитывая то, что клеточная линия K562 получена из эритролейкемии (при этой болезни снижается уровень гемоглобина в крови), а HBA1 - ген альфа-глобина, то вполне ожидаемо, что активность этого гена будет снижена при этой болезни.

Задание 2: Сравнение данных Hi-C в HiGlass

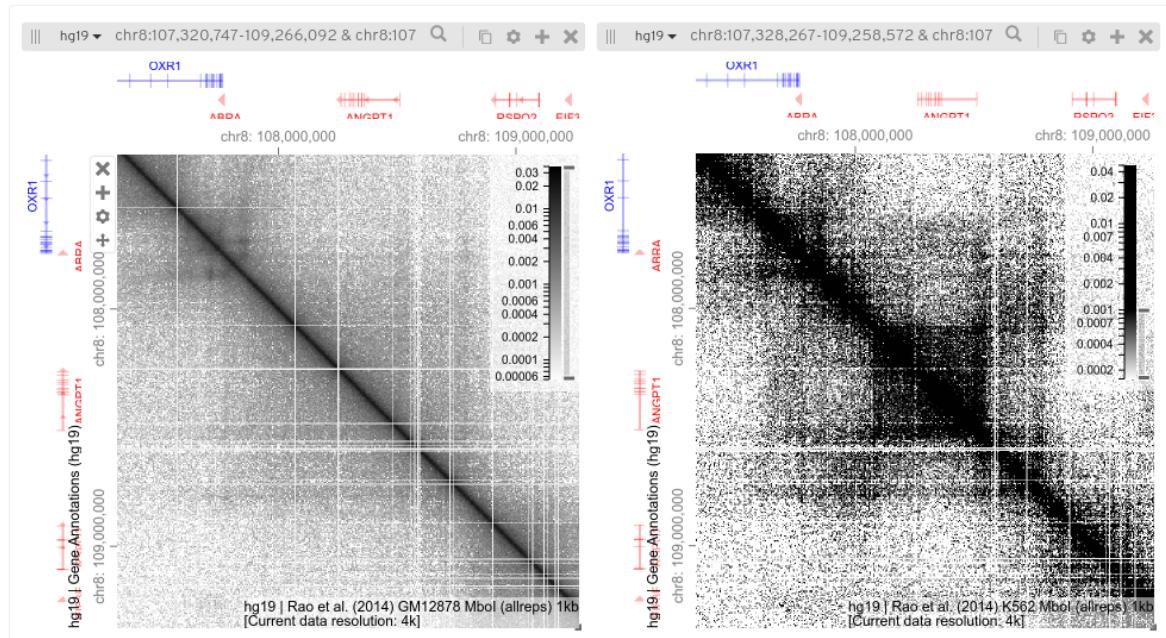
а: Первое бросающееся в глаза отличие между клеточными линиями GM12878 и K562 (шкала примерно одинаковая) - у первой затемненные участки (контакты в хроматине) значительно больше, чем у второй: визуально хорошо различается несколько крупных квадратов (ТАД-ов) слева (GM12878), которые плохо видны справа (K562), в то время как справа хорошо различимы более мелкие ТАД-ы. То есть, можно предположить, что у GM12878 несколько более плотный хроматин, обеспечивающий большее число удаленных контактов. Также на левой матрице слева вверху выделяются точки увеличенной плотности контактов в одном из ТАД-ов, они могут соответствовать петлям. У K562 такие точки в той области менее заметны. Зато у K562 выделяются петли между доменами в середине:



б: Скриншот изменения параметров цветовой шкалы с целью добиться схожести:

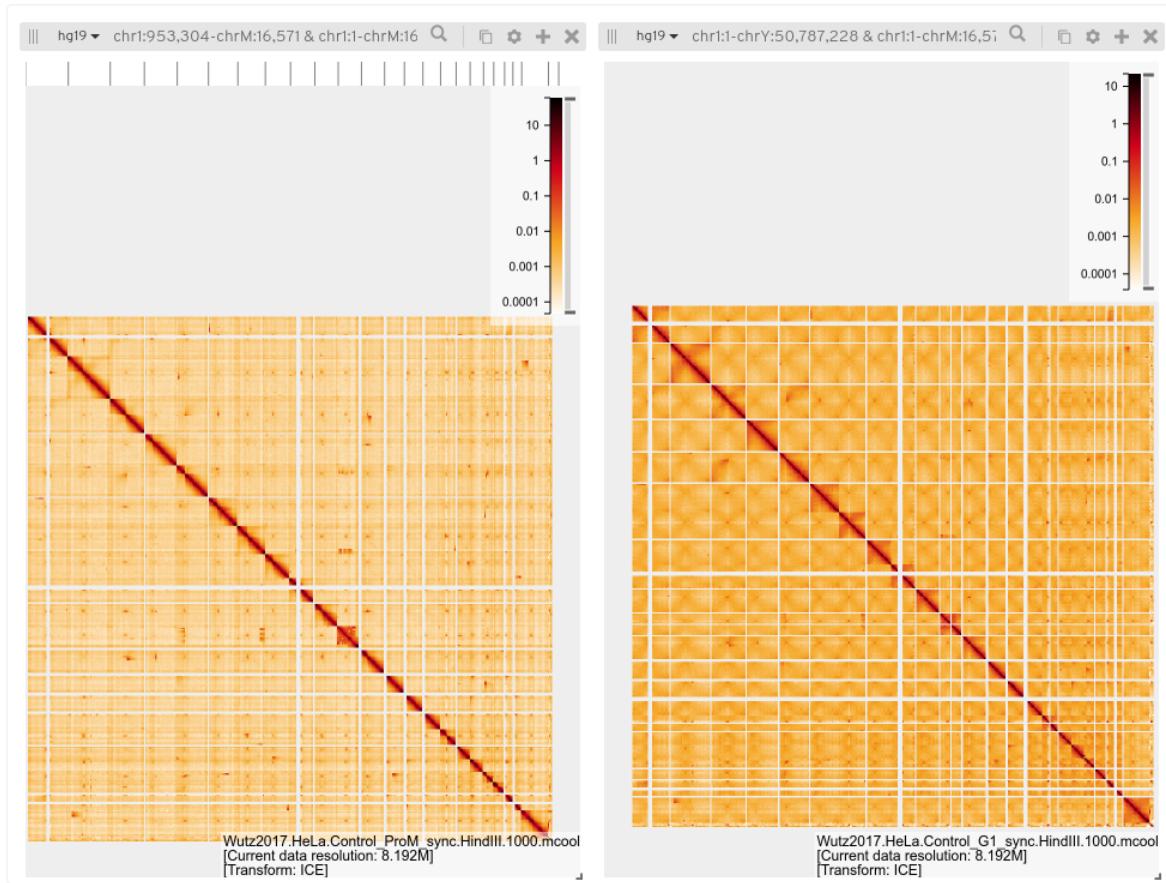
Two Linked Views

HiGlass can also be configured to show two or more views. These views can be synchronized to always show the same location. For more information, please see the documentation about [replacing tracks](#), [adding new views](#), and [synchronizing the locations of different views](#).

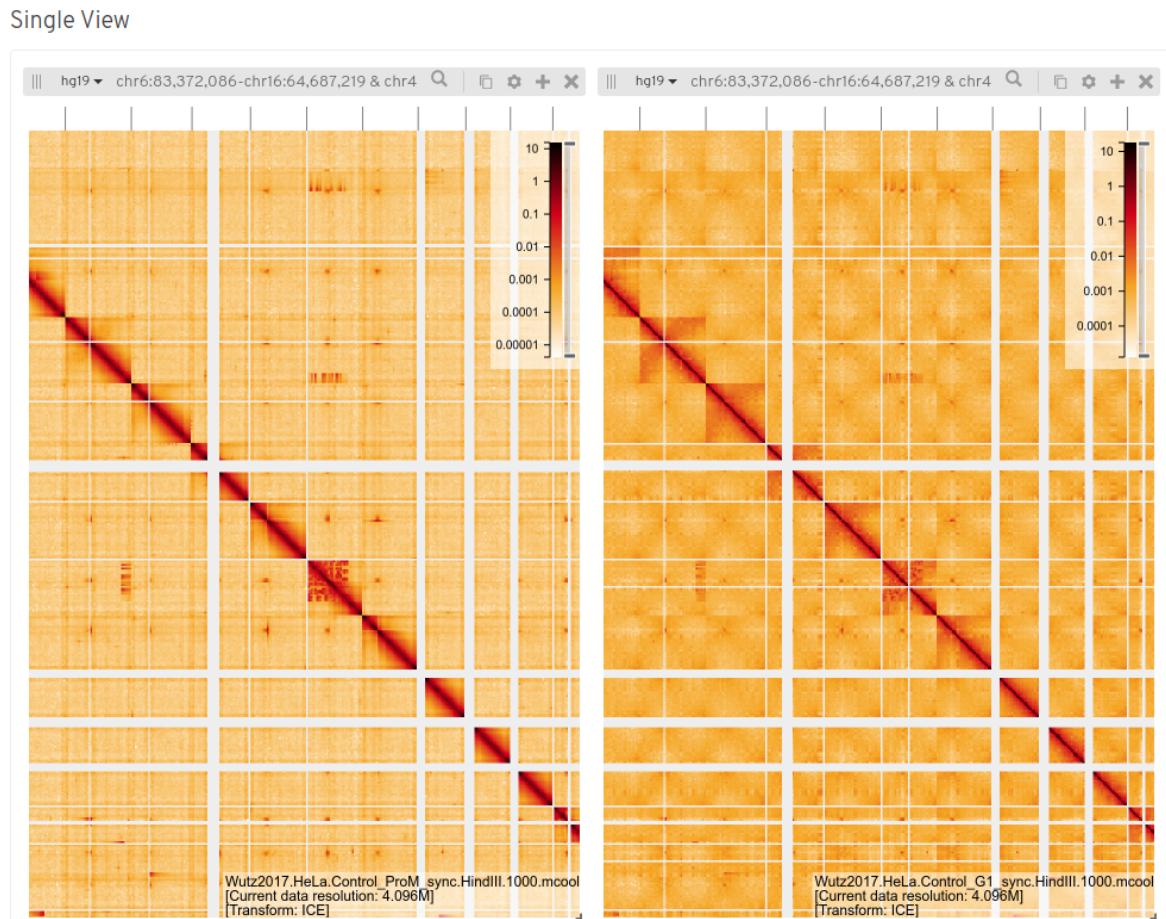


с: Не сказать, что структуры становятся похожи. В целом качество экспериментов неравноценно, матрица справа кажется куда более зашумленной даже при полном ползунке. Но при этом число различимых доменов на правой картинке также выше, будто бы контрастность больше, непонятно ввиду данных ли или качества эксперимента. В целом, всё это несколько сбивает с толку при сравнении.

д, е: Открыты обе матрицы + настроено отражение.



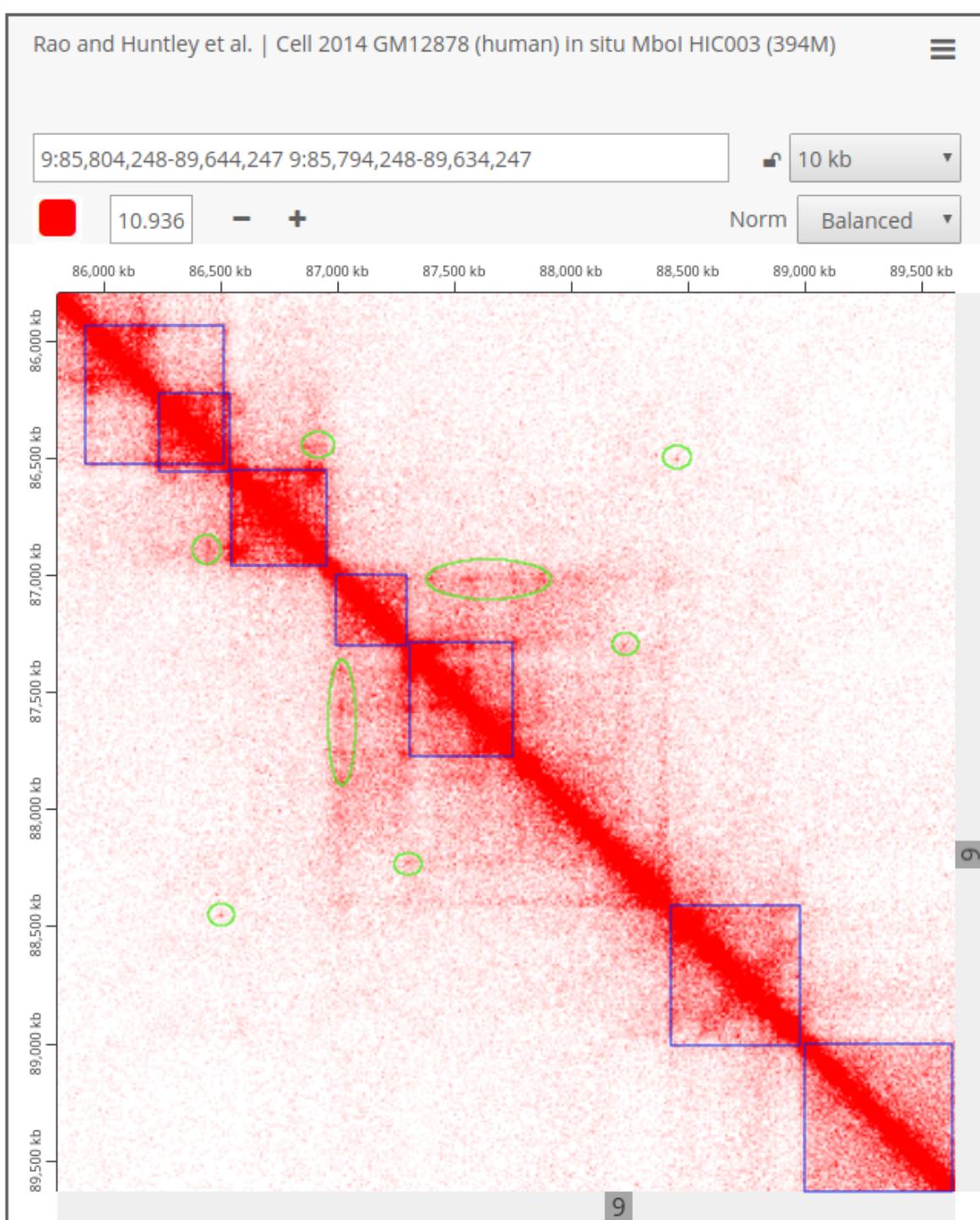
f, g: В целом уже по картинкам выше видно, что и внутрихромосомных, так и межхромосомных контактах на матрице справа больше, особенно заметно в правом нижнем углу. Но приблизим в центре:



Какая разница заметна: матрица справа теплее, значит больше контактов в целом согласно шкале. Кроме того, на ней чётче выделяются домены, относящиеся к хромосомам, обозначая большее число дальних внутрихромосомных контактов. Насчёт межхромосомных контактов - все те же петли одинаковы и слева, и справа, но общий "контактный фон" (тёплость матрицы вне хромосомных доменов) справа выше. Объяснение наблюдаемых различий следующее - слева клетки находятся в прометафазе, это одна из фаз деления клетки. В ходе неё разрушается ядерная оболочка, а хромосомы оказываются в экваториальной плоскости веретена деления. Предыдущая фаза - профаза, в ходе неё хромосомы конденсируются (становятся различимы), что очевидно приводит к уменьшению числа как дальних внутрихромосомных, так и межхромосомных контактов. Данные справа - фаза G1, период роста клетки после митоза, в это время клетка активно синтезирует мРНК и белки, при этом ДНК представляет собой плотный хроматин.

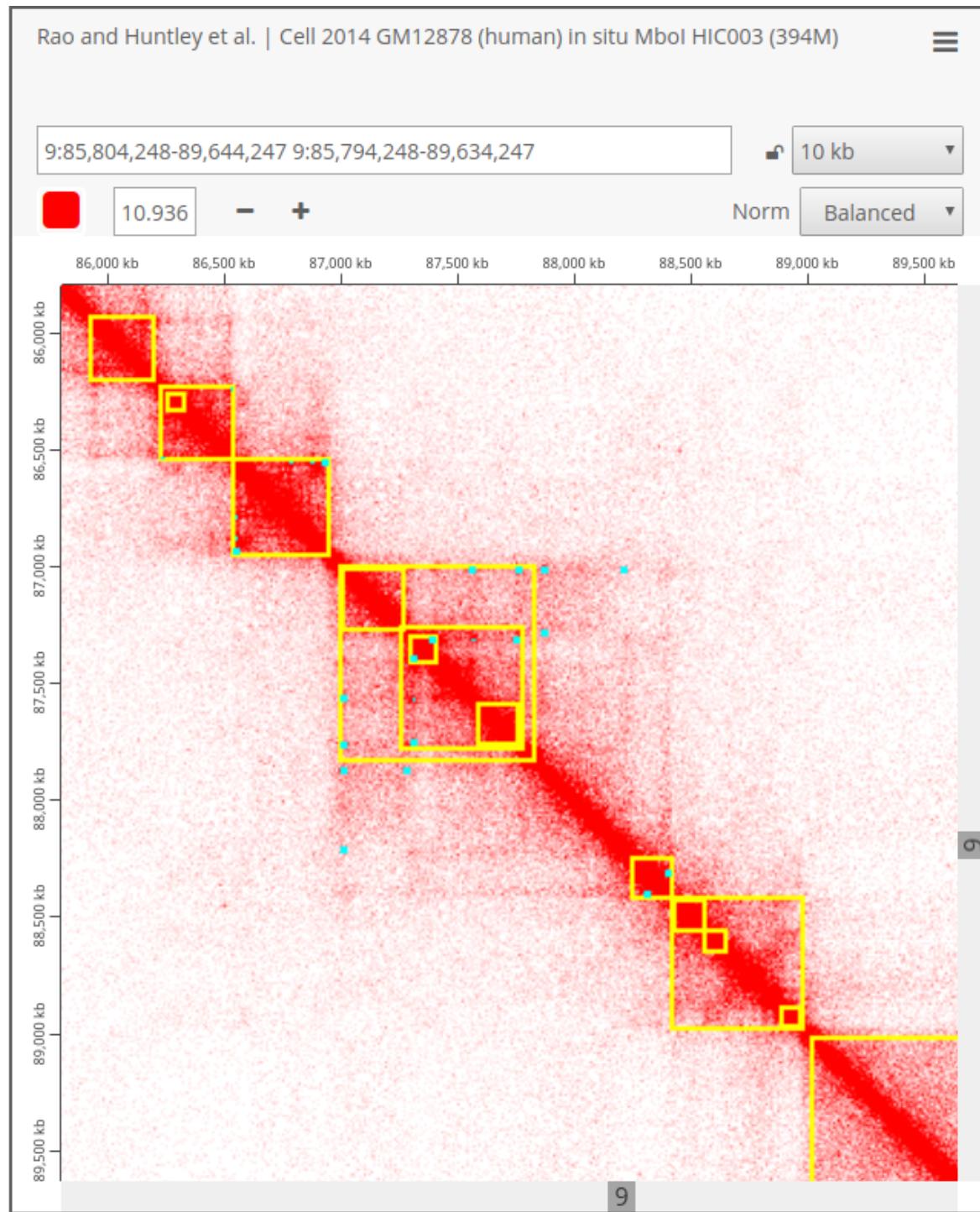
Задание 3: Петли, ТАДы и СТСФ

а: ТАДы выделены синим, петли - зелёным.

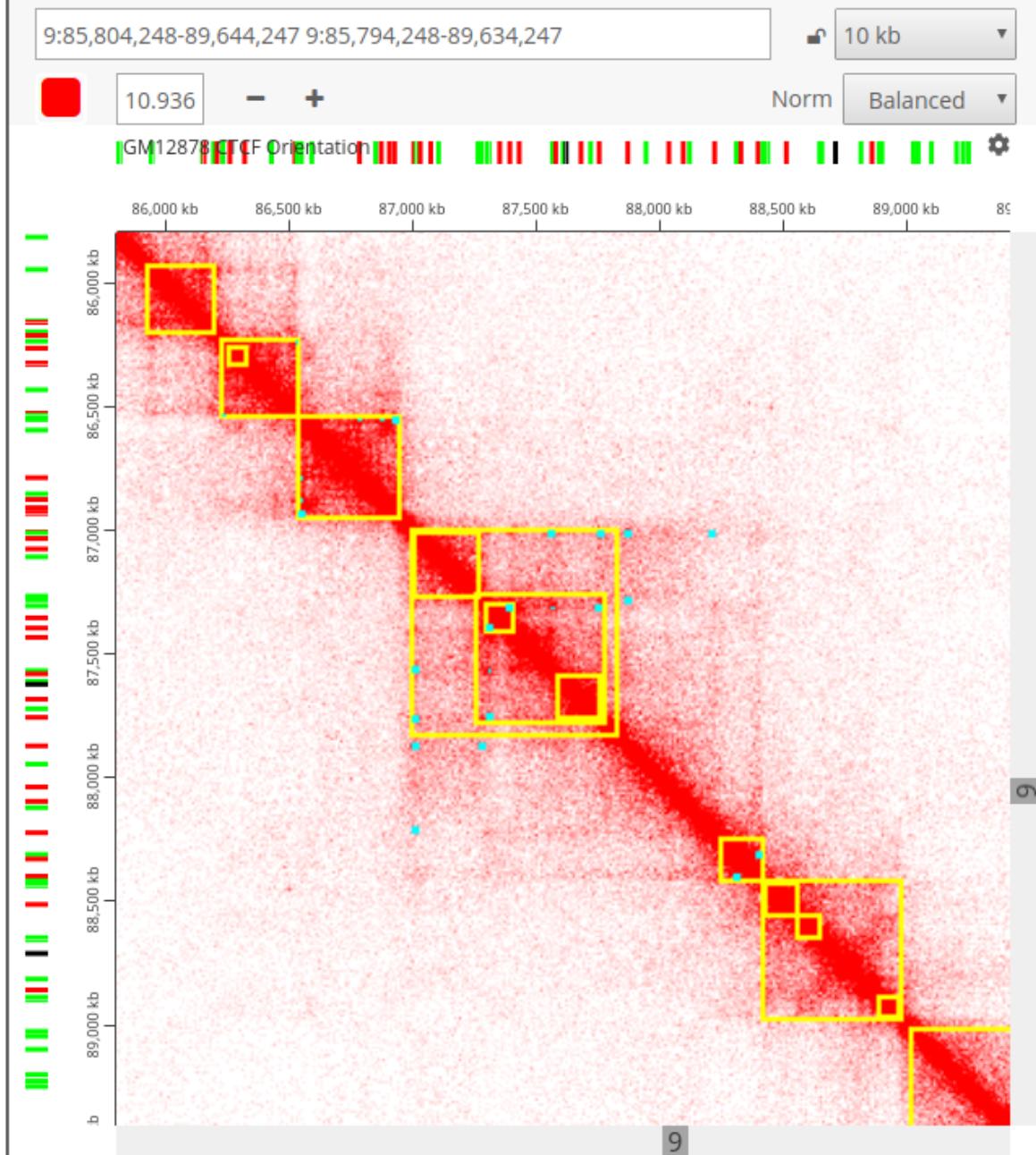


Наблюдаю по крайней мере 7 визуально хорошо различимых ТАД-ов, также 3 отдельных петли и одну группу петель, которая выделена овалом.

б: Сравнивая с моим выделением выше, могу сказать, что с некоторым крупными ТАД-ами было угадано, но маленькие разглядеть не представляется возможным. Группы петель также удалось выделить, но не все.



с:



Глядя на картинку, создаётся такое ощущение, что CTCF любит концентрироваться на границах ТАД-ов. Там же находятся и петли. Внутри ТАД-ов или вне их CTCF встречается более разреженно.

Задание 4: Настройка системы

```
conda create -n chr python=3.6
conda activate chr
conda install -c bioconda deeptools
conda install -c bioconda -c conda-forge cooler
conda install -c bioconda -c conda-forge hicexplorer
pip install hic2cool
```

a:

```
(chr) leo@MS7922:~$ cooler -h | head
Usage: cooler [OPTIONS] COMMAND [ARGS]...

Type -h or --help after any subcommand for more information.

Options:
-v, --verbose  Verbose logging.
-d, --debug    On error, drop into the post-mortem debugger shell.
-V, --version   Show the version and exit.
-h, --help     Show this message and exit.

(chr) leo@MS7922:~$ deeptools --help | head
usage: deeptools [-h] [--version]

deeptools is a suite of python tools particularly developed for the efficient analysis of
high-throughput sequencing data, such as ChIP-seq, RNA-seq or MNase-seq.

Each tool should be called by its own name as in the following example:

$ bamCoverage -b reads.bam -o coverage.bw

If you find deepTools useful for your research please cite as:
(chr) leo@MS7922:~$ _
```

b:

```
(base) leo@MS7922:~/ME/HSE/NGS/8/NGS_HW_Hi-C_data$ ls
Kc167.10000.cool  Kc167-CTCF.bigWig
```

Задание 5: Анализ данных Hi-C

а: Разрешение - 10 Kb, геномная сборка - dm3 (дрозофилы), количество контактов - 28196414.

```
(chr) leo@MS7922:~/ME/HSE/NGS/8/NGS_HW_Hi-C_data$ cooler info Kc167.10000.cool
{
  "bin-size": 10000,
  "bin-type": "fixed",
  "creation-date": "2018-04-19T19:13:35.039403",
  "format": "HDF5::Cooler",
  "format-url": "https://github.com/mirnylab/cooler",
  "format-version": 2,
  "generated-by": "cooler-0.7.5",
  "genome-assembly": "dm3",
  "nbins": 12044,
  "nchroms": 7,
  "nnz": 2963436,
  "sum": 28196414
}
```

Изменим разрешение до 20 Kb:

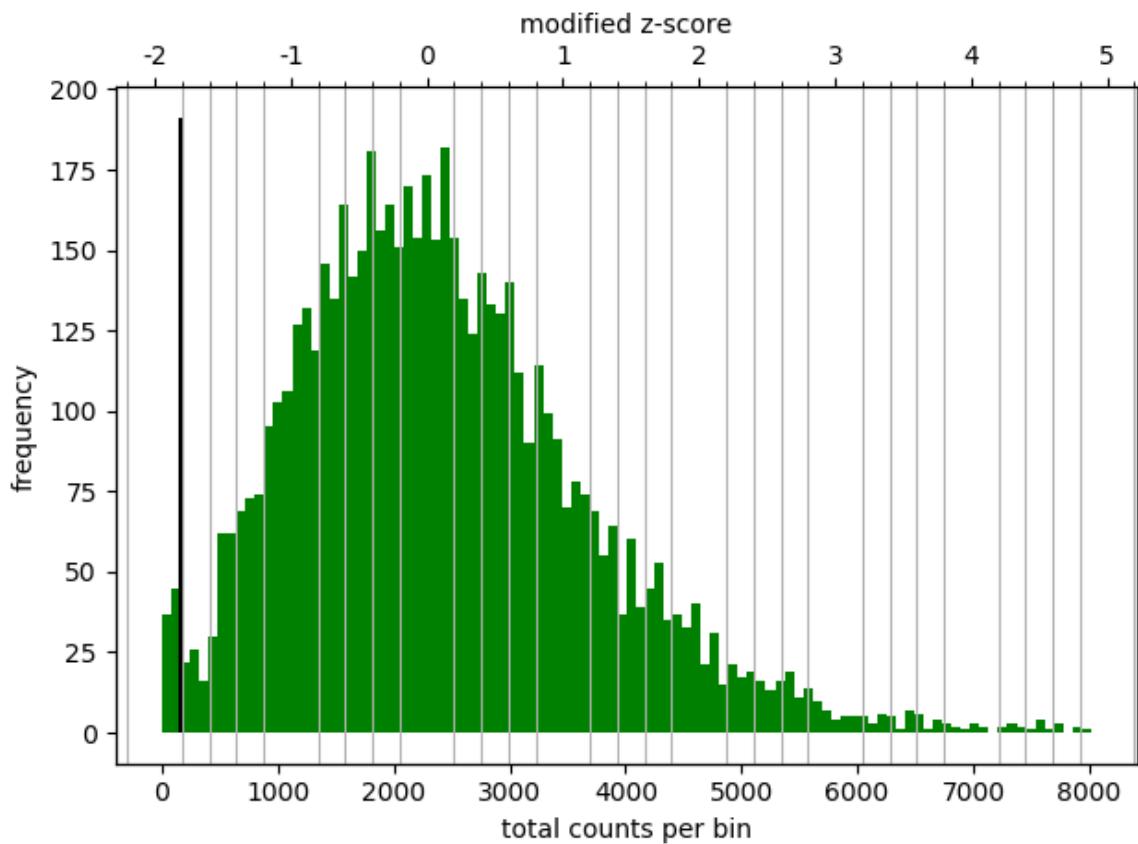
```
cooler coarsen -k 2 -o Kc167.20000.cool Kc167.10000.cool
```

Пересохраним в h5:

```
hicConvertFormat --matrices Kc167.20000.cool -o Kc167.20000.h5 --inputFormat
cool --outputFormat h5
```

б, с: Проведём итеративную коррекцию, а именно для начала построим диагностический график:

```
hicCorrectMatrix diagnostic_plot -m Kc167.20000.h5 -o hic_corrected.png
```

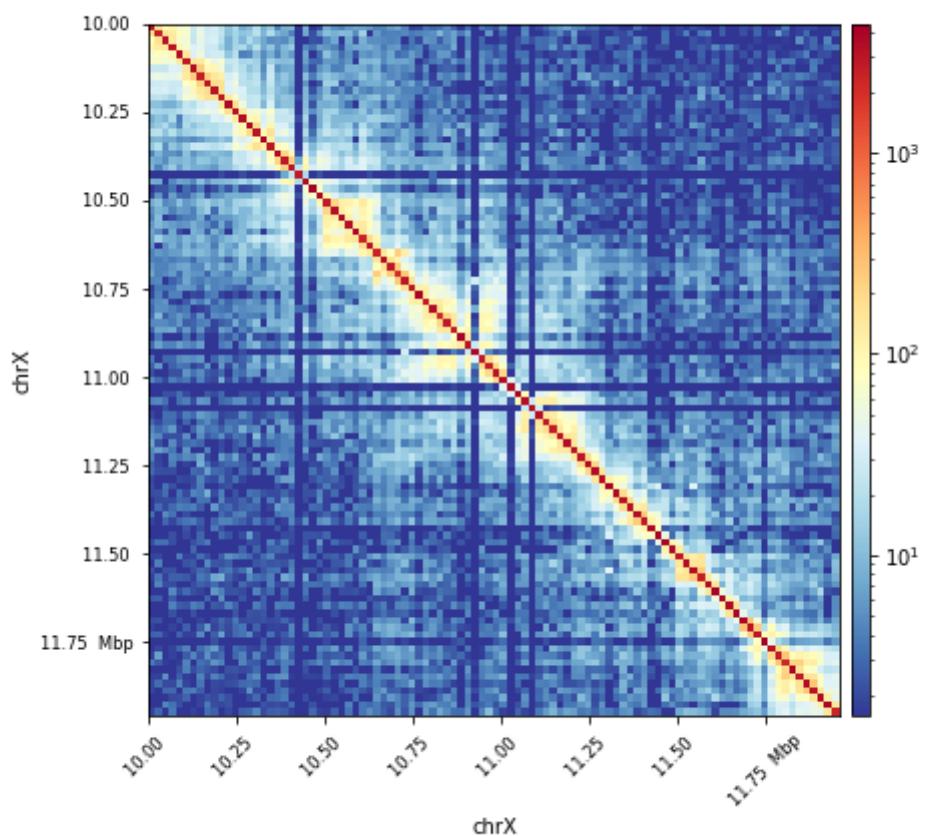
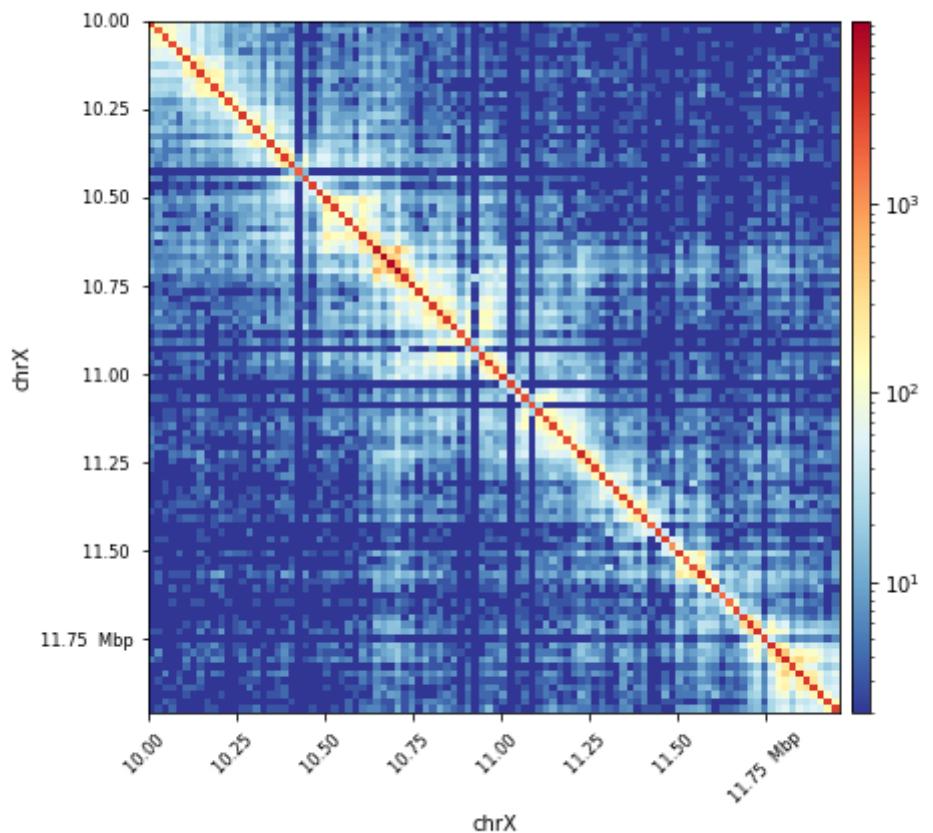


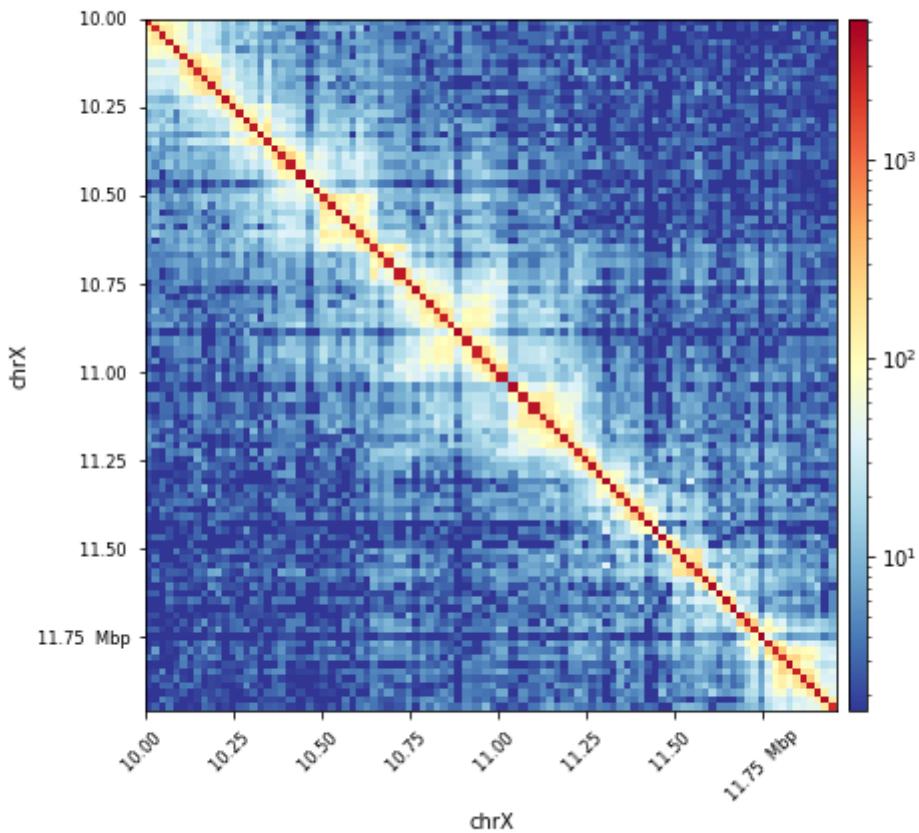
Видим, что пороги надо выбирать следующие: -1.6 и 4, чтобы убрать выбросы. Теперь проведём коррекцию без подбора параметров и с подбором, и построим карты по порядку (исходная, наивная коррекция, коррекция с подобранными параметрами):

```

hicCorrectMatrix correct -m Kc167.20000.h5 --filterThreshold -10 10 -n 10 --
out Kc167.corr.20000.h5
hicCorrectMatrix correct -m Kc167.20000.h5 --filterThreshold -1.6 4 -n 500 --
out Kc167.corr2.20000.h5
hicPlotMatrix -m Kc167.20000.h5 -o Kc167.raw mtx.png --log1p --clearMaskedBins
--region chrX:10000000-12000000
hicPlotMatrix -m Kc167.corr.20000.h5 -o Kc167.corr mtx.png --log1p --
clearMaskedBins --region chrX:10000000-12000000
hicPlotMatrix -m Kc167.corr2.20000.h5 -o Kc167.corr2 mtx.png --log1p --
clearMaskedBins --region chrX:10000000-12000000

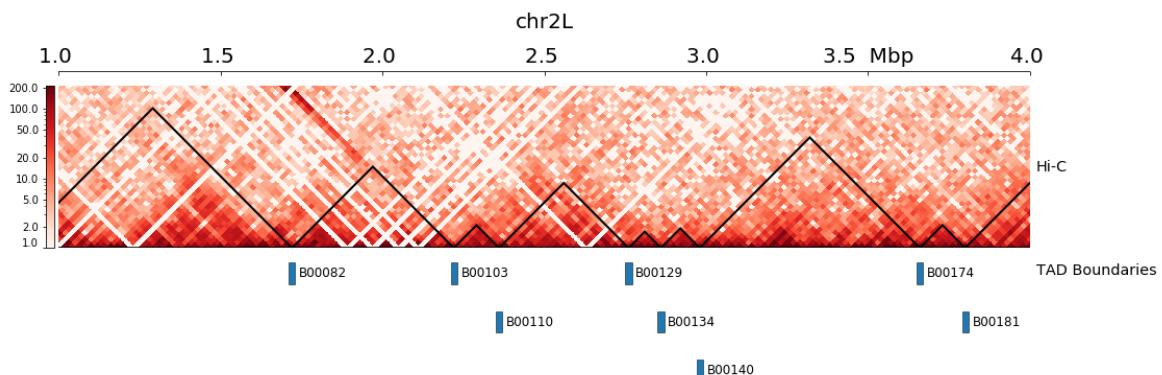
```





Как видим, при проведении наивной коррекции часть шума с карты ушла, ТАДы стали чуть более заметными. Но подбор параметров коррекции ощутимо улучшил результат: шума стало сильно меньше, квадраты ТАД-ов теперь визуально прослеживаются. Собственно, теперь поищем ТАДы в скорректированной версии карты и посмотрим, что получилось:

```
hicFindTADs -m Kc167.corr2.20000.h5 --outPrefix tad1 --minDepth 60000 --
maxDepth 1000000 --step 20000 --thresholdComparisons 0.05 --delta 0.01 --
correctForMultipleTesting fdr
hicPlotTADs --tracks tracks1.ini --region chr2L:1000000-4000000 -o tads1.png
```



д: ТАДы получились размерами от 10 Кб до порядка одной мегабазы. Параметры графика в следующем файле:

```
[x-axis]
where = top
fontsize = 20
```

```

[hic_matrix]
file = Kc167.corr2.20000.h5
title = Hi-C
colormap = Reds
depth = 1000000
transform = log1p
x_labels = True
file_type = hic_matrix
show_masked_bins = True
scale_factor = 1

[tads]
file = tad1_domains.bed
display = triangles
border_color = black
color = none
line_width = 2
overlay_previous = share-y

[spacer]

[bed_boundaries]
file = tad1_boundaries.bed
file_type = bed
height = 4
title = TAD Boundaries
min_value = 0
max_value = 2.5

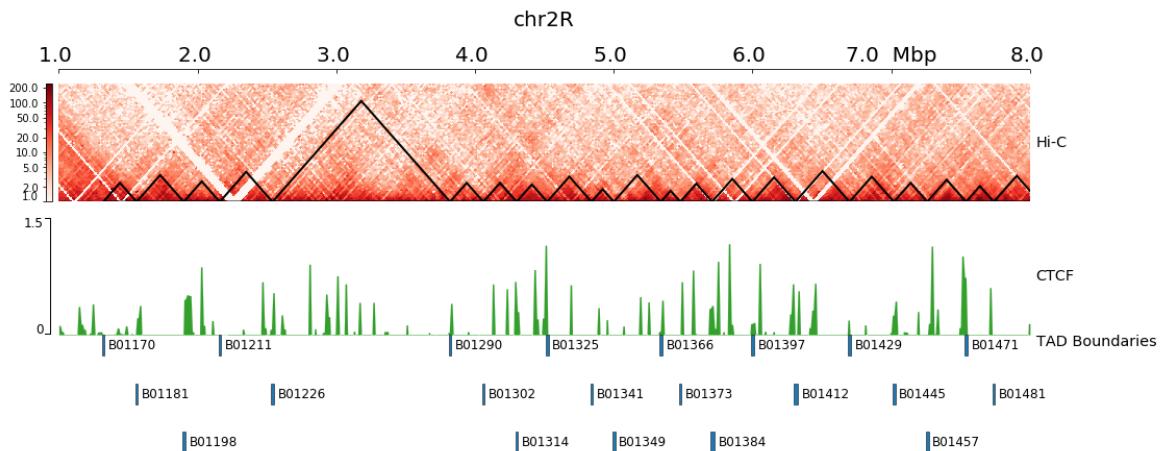
```

Попробуем сделать по-другому (и немножко другие параметры в ini, добавим трек CTCF):

```

hicFindTADs -m Kc167.corr2.20000.h5 --outPrefix tad2 --minDepth 60000 --
maxDepth 200000 --step 20000 --thresholdComparisons 0.05 --delta 0.01 --
correctForMultipleTesting fdr
hicPlotTADs --tracks tracks2.ini --region chr2R:1000000-8000000 -o tads2.png

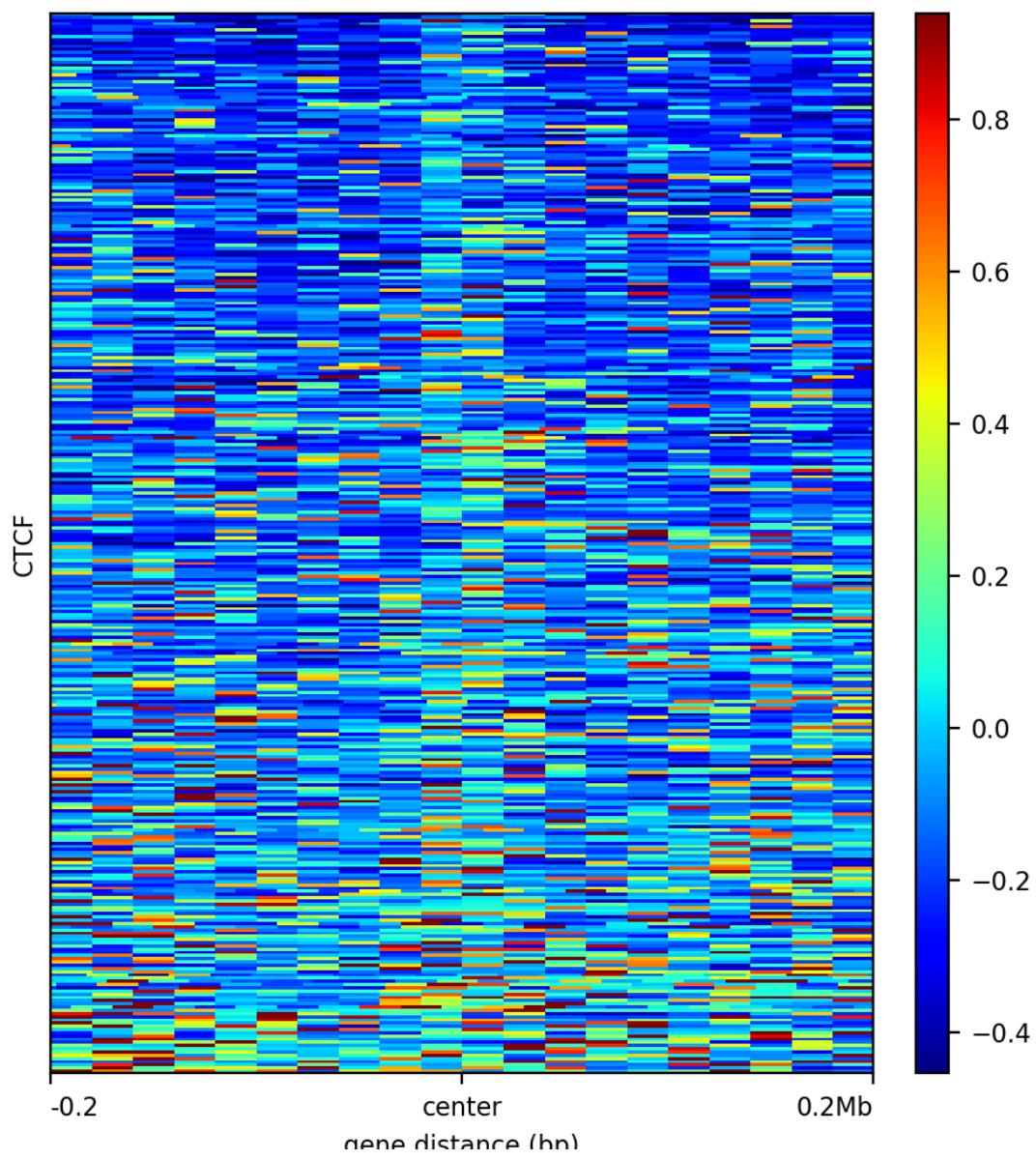
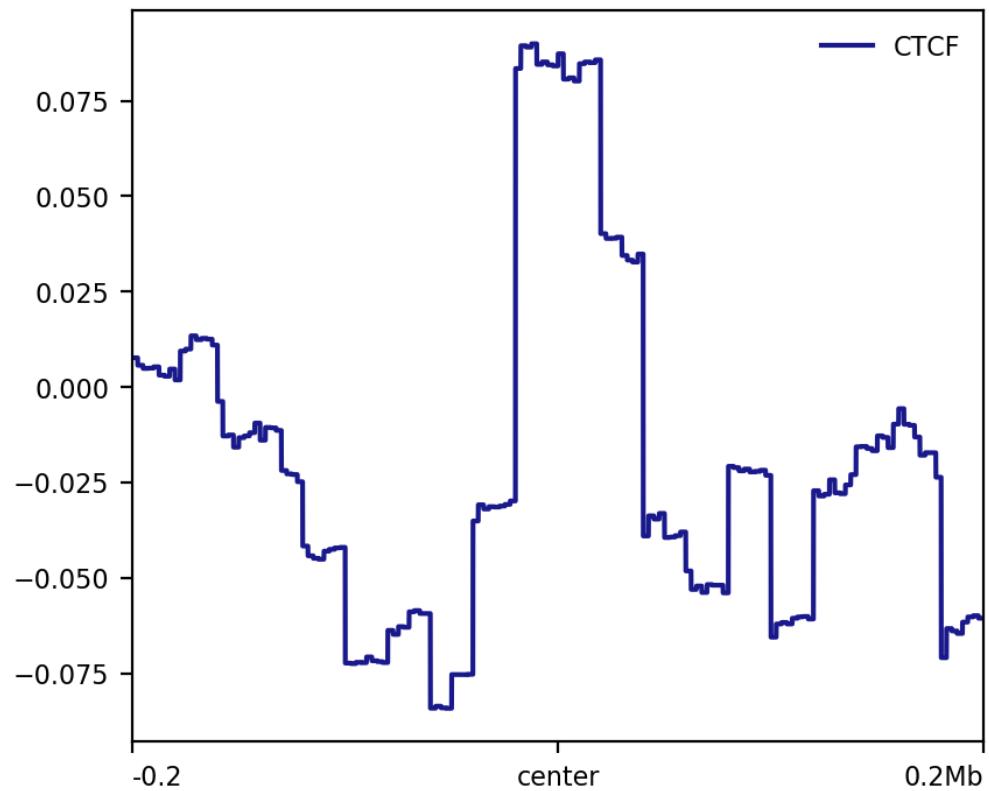
```



Здесь ТАДы размером от нескольких десятков килобаз, до 1.3 Мб. Также можно увидеть профиль ChIP-Seq для CTCF - многие пики приходятся на границы ТАД-ов. Теперь сравним ТАДы с этим профилем с помощью deeptools (e):

```
computeMatrix reference-point -S Kc167-CTCF.bigWig -R tad2_boundaries.bed --  
referencePoint center -a 200000 -b 200000 -out me.tab.gz -p 8  
plotHeatmap -m me.tab.gz -out me.png --heatmapHeight 15 --heatmapWidth 10 --  
colorMap jet --sortRegions ascend --regionsLabel 'CTCF'
```

Kc167-CTCF



По оси X здесь расстояние от границы ТАД-ов в мегабазах. По оси Y - регионы границ ТАД-ов, сортированные по среднему значению (снизу), сигнал CTCF (сверху). Паттерн распределения пиков CTCF вокруг границы ТАД-ов действительно такой, что наибольший сигнал приходится именно на саму границу, потом ослабляется при отходе от границы, а затем снова усиливается, но не достигает такого значения как на границы.

То есть, основной вывод такой - на границе ТАД-ов плотность контактов уменьшается, что может свидетельствовать о обширных участках открытого хроматина, и это косвенно подтверждается данными ChIP-Seq для CTCF.