



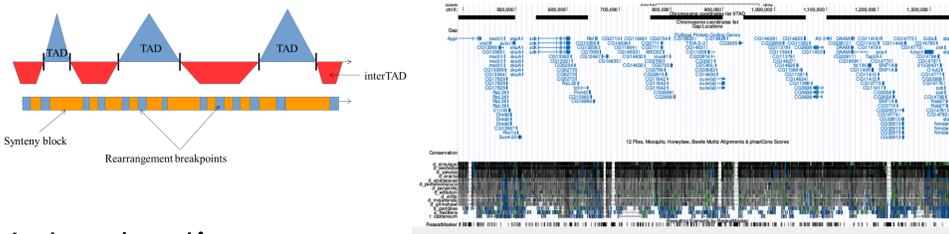
# Перестройки в 3D

## Rearrangements in 3D

Kristina Perevoshchicova, Ekaterina Khrameeva

### Abstract

Topologically associating domains (TADs) are evolutionarily conserved units of chromatin topology. What happens to the integrity of TADs when genomes are challenged by structural rearrangements? If TADs act as modular units (e.g., to regulate gene expression), then large-scale rearrangements would be expected to occur at domain borders, so as to maintain the integrity of these structures. We have combined the genome-rearrangement and Hi-C data in *Drosophila* and have demonstrated that chromosome translocations occur more frequently between spatially close genome fragments. Moreover, rearrangements are observed more frequently between interTAD regions. This result suggests that TADs function as stable units that are selected against breakage during genome rearrangements.



### Introduction

ДНК в ядре, связанная с белками и называемая хроматином, имеет определенную пространственную организацию. Опыты с хроматином млекопитающих показывают, что его пространственная структура может коррелировать с функциями генов и перестройками в геноме. Мы выявили зависимость между локализацией перестроек в геноме дрозофил и структурой хроматина в этом месте. Перестройки чаще встречаются вне участков компактизованного хроматина - ТАДов. Но локализация участков, подверженных перестройкам может зависеть не только от структуры хроматина, но и от пространственного сближения двух участков, между которыми происходят перестройка, что также было подтверждено в ходе работы.

Пример исходных данных:

Хромосома у D. melanogaster	Старт гена в D. melanogaster	Конечн гена в D. melanogaster	Старт гена в D. simulans	Конечн гена в D. simulans	Нумерация генов в D. melanogaster	Нумерация генов в D. simulans	Идентификатор генов в D. melanogaster
X	353293	365902	288061	290906	26	30	CG4122
X	366240	384859	297708	316568	27	31	CG18104
X	374043	380042	305873	307309	28	32	CG4262
X	1,3E+07	1,3E+07	305873	306934	29	1289	CG4396
X	384763	387113	316743	318184	30	33	CG4293
X	388441	434960	340689	353112	31	34	CG7727
X	447863	454643	368799	374587	32	35	CG6172

### Methods

- Извлекли из статьи Bhutkar et al. при помощи программирования на Perl данные о конце и начале синтенных блоков на хромосомах дрозофилы.
  - Просчитали координаты середин синтенных блоков
  - Загрузили в онлайн сервис GALAXY в формате bed имеющиеся у нас координаты ТАДов и полученные нами координаты середин промежутков между синтенными блоками, а затем при помощи функции нахождения пересечений там же получили данные о попадании/непопадании координат середин промежутков между синтенными блоками (условно - середины перестроек) в границы ТАДов. Затем мы собрали данные, полученные таким способом, в таблицу.
- Выписываем координаты перестроек.
  - При помощи сервиса GALAXY находим участки пересечения ТАДов и перестроек, суммируем длину участков пересечения.
  - Имея величины длин всех ТАДов и всех перестроек находим длину пересечения перестроек с интерТАДами.
- При помощи скрипта разбиваем каждый ТАД на блоки по 10000 нуклеотидов
  - Нумеруем участки по их удаленности от границ ТАДА (самые крайние участки получают самые маленькие номера)
  - В GALAXY находим пересечение между блоками ТАДов и серединами перестроек.
  - Считаем количество пересекающихся с серединами перестроек блоков ТАДов, имеющих разные номера, считаем общее количество блоков ТАДов, имеющих конкретный номер.
- Нормируем количество пересекающихся блоков с определенным номером на общее число блоков с таким номером и строим графики.
  - Находим пересечение блоков и перестроек в GALAXY
  - Считаем все количество блоков с определенным номером и количество всех блоков, пересекшихся с границами ТАДов.
  - Нормируем количество пересечений с определенным номером блока на все число пересечений.
- Из базы Flybase мы взяли кодификаторы генов D. melanogaster, вместе с их локализацией на хромосоме и координатами
  - Из файла-приложения к статье, в которой определялись синтенные блоки рода *Drosophila*, извлекли информацию о координатах и расположении генов D. simulans, записанных под кодификаторами и порядковыми номерами D. melanogaster
  - При помощи скрипта мы по кодификаторам определили, какие из генов D. melanogaster (пронумерованные и названные по кодификатору), соответствуют генам D. simulans и какие у них координаты и расположение в каждом из двух геномов.
  - Сортируем полученную шагом выше таблицу по координатам генов в геноме D. simulans и нумеруем гены по их локализации.
  - Сортированная таблица содержит кодификаторы генов D. melanogaster сопоставленные с их номером в D. melanogaster и D. simulans. Т.к. таблица сортирована по расположению генов в D. simulans, то любая перестройка в D. Melanogaster – это резкое изменение порядка гена.
  - При помощи написанного скрипта ищем такие события в геноме и выписываем "старые" и "новые" координаты гена из генома D. melanogaster
  - При помощи скрипта сопоставляем старые и новые координаты с вероятностью контакта этих двух точек в пространстве
  - Рандомизировали точки перестроек и посчитали вероятность контакта для случайно выбранных точек в пространстве
  - Сравнили данные графики для перестроек и для рандомного списка по Т-тесту

### Conclusions

- Перестройки чаще происходят рядом с интерТАДами, чем ТАДами, возможно, потому, что ферментам, осуществляющим перестройки, проще работать с некомпактизованным хроматином интерТАДов.
- Средины перестроек ассоциированы в большей степени с границами ТАДов, где наблюдается промежуточная степень конденсации хроматина, и в меньшей степени с серединами ТАДов, где хроматин очень плотно компактизован.
- Подобная корреляция наблюдается и при пересечении целых перестроек с блоками ТАДов.
- Доли перестроек, перекрывающихся с ТАДами и с интерТАДами, равны. Нуклеотиды синтенных блоков имеют большую вероятность попасть в ТАД, чем нуклеотиды промежутком между синтенными блоками.
- Перестройки чаще происходят между участками, сближенными в пространстве.

Лаборатория бактериальной и функциональной геномики

Laboratory of Bacterial and Functional Genomics

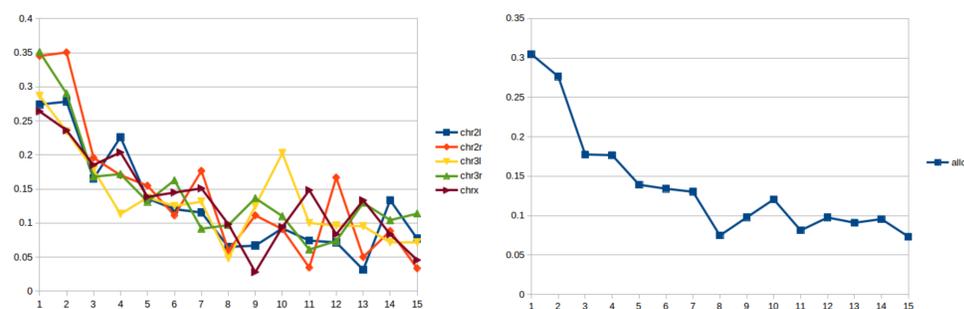
### Results

Table 1. Intersections between rearrangement breakpoints and TADs.

	chr2L	chr2R	chr3L	chr3R	chrX	sum
matches	321	323	312	407	303	1666
mismatches	66	104	103	142	70	485
TADs	19960000	17760000	20280000	23800000	18680000	100480000
interTADs	2420000	2980000	3460000	3760000	3520000	16140000
matches/TAD	0,00160822	0,00181869	0,00153846	0,00171008	0,00162206	0,00001658
mismatches/intertads	0,00272727	0,00348993	0,00297688	0,0037766	0,00198864	0,00003005

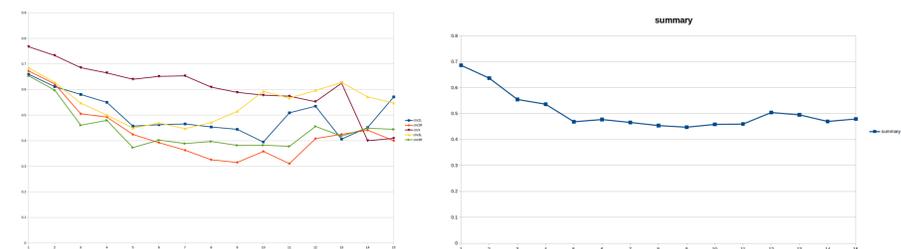
Средины разрывов между синтенными блоками чаще попадают в интерТАДы, чем в ТАДы, значит перестройки чаще происходят между интерТАДами, чем между ТАДами (см. Методы, 1).

Figure 1. Distribution of rearrangement breakpoints within TADs. Left – different chromosomes, right – all chromosomes. X axis – distance from TAD border, in 20Kb bins. Y axis – frequency of rearrangements.



При удалении от границ ТАДов к их центру вероятность пересечения с серединой перестройки падает (см. Методы, 2).

Figure 2. Distribution of regions between synteny blocks within TADs. Left – different chromosomes, right – all chromosomes. X axis – distance from TAD border, in 20Kb bins. Y axis – frequency of rearrangements.



Результат и вывод аналогичны картинке 1 (см. Методы, 3).

Table 2. Intersections between inter-synteny regions and TADs.

#intersections borders with TADs	3312
#intersections borders with interTADs	990
length of TADs	100480000
length of interTADs	16140000
#TADs/LTADs	0,000033
#interTADs/LinterTADs	0,000061
difference	1,860891

Rearrangements intersect with TADs	38198632
Rearrangements intersect with interTADs	6241830
Synteny blocks intersect with TADs	62281368
Synteny blocks intersect with interTADs	7681229
Full length of rearrangements	44440462
Full length of synteny blocks	69962597
Full length of TADs	100480000
Full length of interTADs	16140000

Frequency nucl. Of rearrangement being in TAD	0,380	0,983
Frequency nucl. Of rearrangement being in interTAD	0,386	
Frequency nucl. Of synteny block being in TAD	0,619	1,302
Frequency nucl. Of synteny block being in interTAD	0,475	

Доли разрывов между синтенными блоками, перекрывающихся с ТАДами и с интерТАДами, одинаковы, но доли синтенных блоков, перекрывающихся с ТАДами, значительно больше, чем с интерТАДами. Границы разрывов между синтенными блоками (как и середины этих разрывов) пересекаются с интерТАДами чаще, чем с ТАДами (см. Методы, 4).

Figure 3. Breakpoints of chromosomal translocations in *D. melanogaster* vs. *D. simulans* tend to form contacts.

