ИНСТИТУТ
БИОИНФОРМАТИКИ

Поиск молекулярных маркеров дифференциальной окраски хромосом

Руководитель: Барбитов Юрий Александрович,
Институт биоинформатики

Студент: Климина Александра

G(R)-рисунок в кариограмме мужчины при окраске по Гимза

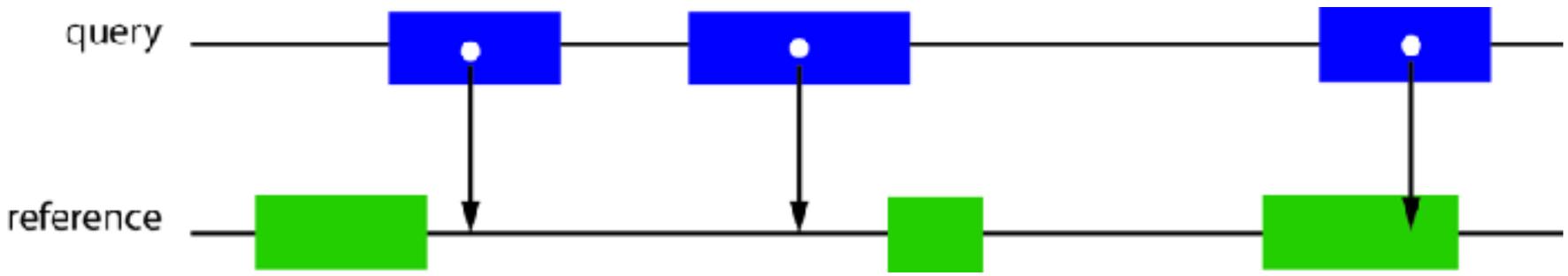
Основные закономерности, связанные с окраской хромосом

Chromosome band	G/Q band	R band	T band (R subgroup)
Replication timing	Late	Early	Very early
GC%	AT-rich	Medium	GC-rich
Gene density	Low	High	Very high
Chromatin appearance	Compact	Loose	Loose

Цель проекта:
изучить связь различных черт последовательности генома и геномных элементов с дифференциальной окраской хромосом.

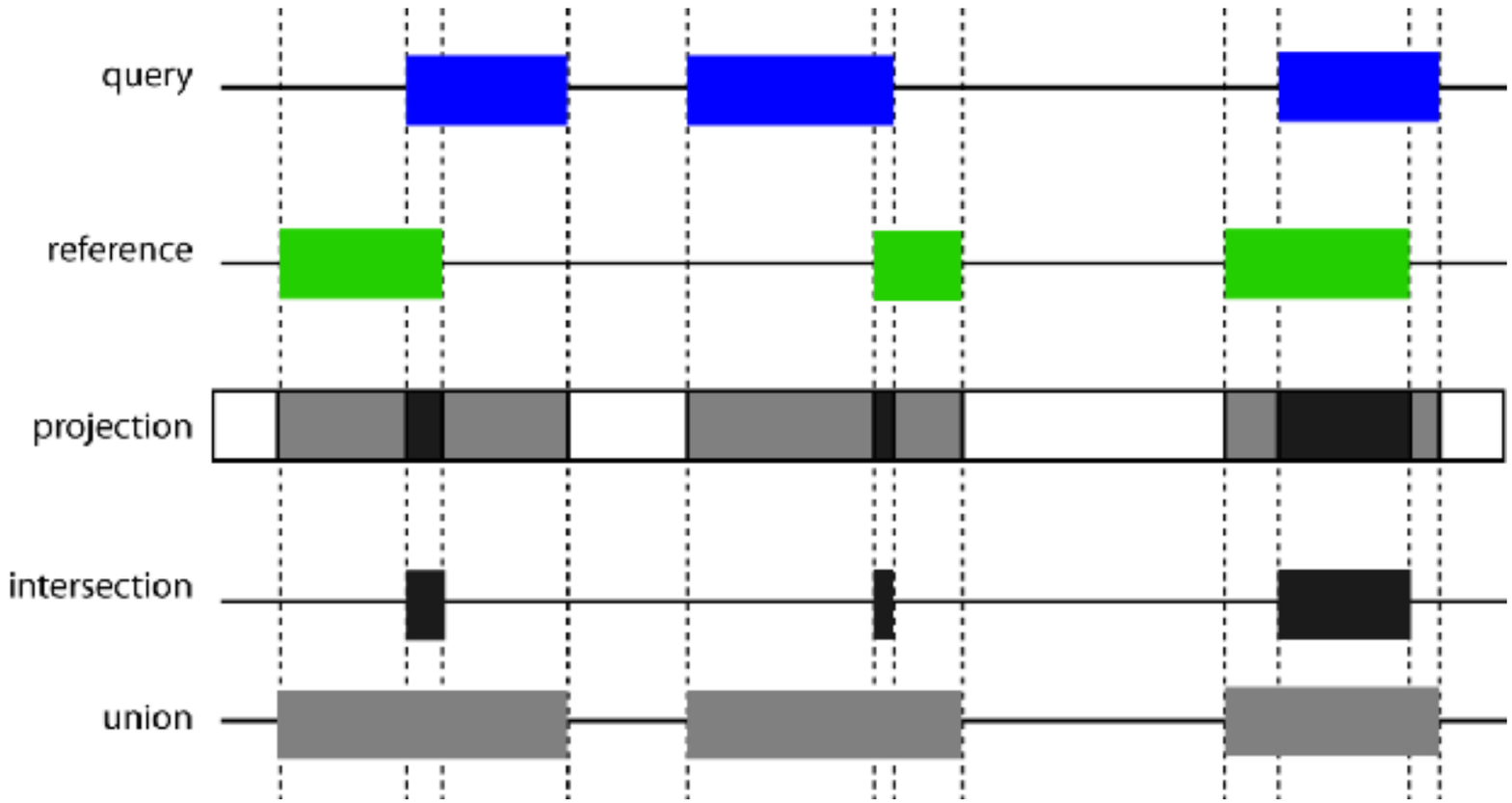
Задачи:
- Реализовать вычисление разных метрик пересечения интервалов (ProjectionTest, JaccardTest)
- Получить распределение значений метрик пересечения для случайных интервалов чтобы проверить эффективность тестов
- Протестировать корреляции данных геномных экспериментов с паттерном дифференциального окрашивания

Projection test



Число интервалов тестируемой последовательности, не менее, чем на половину перекрывающихся с референсной последовательностью

Jaccard test

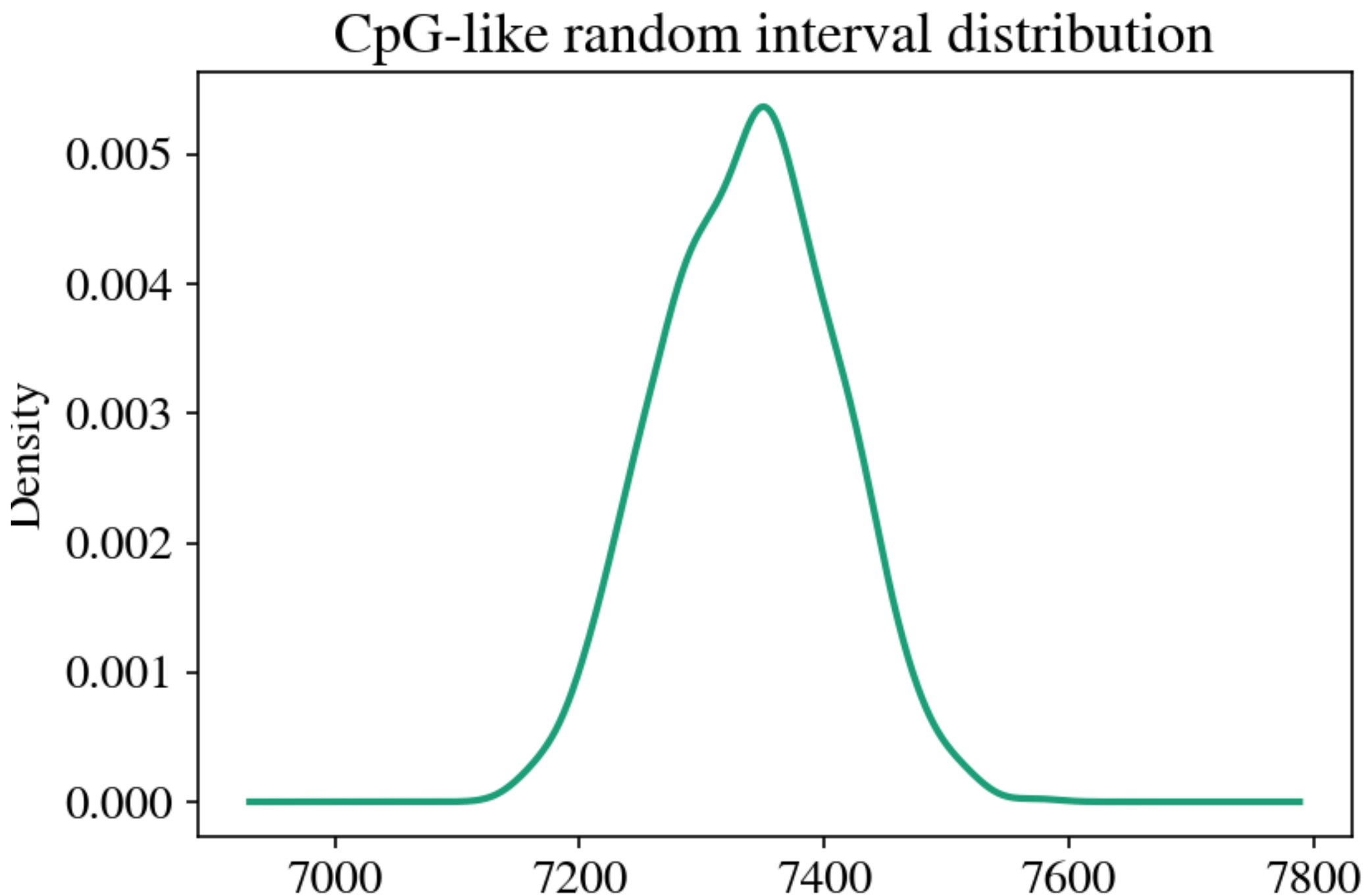


Отношение длин областей пересечения интервалов к длине областей их объединения

$$\text{Jaccard measure (index):} \quad J(A,B) = \frac{A \cap B}{A \cup B}$$

Написан инструмент для анализа последовательностей:

- реализованы ProjectionTest и JaccardTest для отдельных треков и для целого генома;
- модуль для генерации случайных треков по заданному запросу (используется исходные длины интервалов, но случайное их распределение по хромосоме);
- проверка гипотезы о том, что исследуемый трек можно рассматривать как случайный.


CpG-like random interval distribution
Density
0.005
0.004
0.003
0.002
0.001
0.000
7000
7200
7400
7600
7800

ProjectionTest

Microsatellit
For query: 14067
For query-like tracks:
mean = 12473.695
sd = 89.820053301

CpG
For query: 2777
For query-like tracks:
mean = 7337.836
sd = 71.0640915231

DNAse HS
For query: 31530
For query-like tracks:
mean = 55229.485
sd = 179.21827411

JaccardTest

Microsatellit
For query: 6.293760668613641E-4
For query-like tracks:
mean = 0.000575267523199
sd = 4.42819753615e-06

CpG
For query: 0.00238119918564809 94
For query-like tracks:
mean = 0.00628882916869
sd = 8.82474644317e-05

DNAse HS
For query: 0.00513456950623885 3
For query-like tracks:
mean = ?
sd = ?

Спасибо за внимание!