

## Руководство OMTools v1.4

Олден Леунг, Тинг-Фунг Чан  
Китайский университет Гонконга

10 марта 2018 г.

Авторские права 2018, Все права защищены

## Содержание

1 Введение	1
2 Достоинство и реализация	1
3 Быстрый старт	2
4 Основные процедуры анализа	2
Якаторграф	4
5 OMBlastMapper	4
6 OMHAMapper	8
7 OMFMMapper	12
8 Парное выравнивание	16
II Моделирование	19
9 OptMapDataGenerator	19
10 Генератор случайных ссылок	21
III SV-обнаружение	22
11 SVDetection	22
IV Fasta Инструменты	24
12 FastaToOM	24
V Инструменты данных	25
13 DataTools	25
14 Проверка качества данных	27
15 Данные Статистика	28
16 Дублированных Молекул Обнаружение	29
17 Дублированных Молекул Удаление	30
18 FrequentKmerВыделение	31
Инструменты выравнивания VI	32

19 РезультатИнс трументы	32
20 РезультатОбъединение	36
21 РезультатСтатис тика	37
22 PrecisionRecallGraphDataGenerator	38
23 Запрос Обратный	40
24 ВыравниваниеВыделе ние	41
<b>VII Множес твенное выравнивание</b>	<b>42</b>
25 Множес твенное выравнивание	42
<b>VIII Инс трументы для множес твенног о выравнивания</b>	<b>49</b>
26 CBLTools	49
27 Анализ производительнос ти множес твенног о выравнивания	50
28 БлокСоединениеГрафГенерац ия	51
<b>IX Ф илог енетика</b>	<b>52</b>
29 UPGMATreeConstruction	52
<b>X Визу ализац ия</b>	<b>53</b>
30 OMВид	53
<b>XI Друг ие с ц енарии</b>	<b>69</b>
31 TWINResultRepeatRemover	69
32 ОтдельныйBNXScan	70

## 1 Введение

Оптическое картирование — это метод захвата определенных участков фермента на длинной молекуле ДНК. Выходными данными этой технологии являются оптическая карта, представленная кортежем (рис. 1).



Рисунок 1: Пример оптической карты. Зеленый прямоугольник — это отцов ДНК, а черные столбцы — сайт фермента. Оптическая карта содержит 6 сигналов и 7 сегментов и может быть представлена в виде кортежа: [1518; 15487; 8455; 1350; 25188; 17845; 4948]. Первый и последний сегменты обычно игнорируются при анализе, поскольку их размеры неточны.

OMTools — программный пакет, обеспечивающий эффективную и интуитивно понятную обработку и визуализацию данных. Модули для обработки данных оптического картирования.

## 2 Достоинства и реализация

### 2.1 Достоинства OMTools

можно получить по адресу <https://github.com/aldenleung/OMTools> и выпустить под лицензией GPL (подробности см. в дис трибутиве программного обеспечения).

### 2.2 Минимальные требования

OMTools реализован на Java 1.8. Программное обеспечение было протестировано на Ubuntu 14.10 и Microsoft Windows 7, 10.

### 2.3 Установка

Все необходимые библиотеки размещены в папке lib/.

1. Скомпилируйте пакет OMTools в папке OMTools:  
`javac -d bin -sourcepath src -cp "lib/*" @classes`
2. Создайте запускаемый jar-файл для OMTools:  
`jar cvfm OMTools.jar manifest -C bin .`
3. Запустите OMTools:  
`java -jar OMTools.jar Имя Модуля`

### 3 Быстрый старт

```
java -jar OMTools.jar FastaToOM --fastain Ecoli.fa --refmapout Ecoli.ref --enzyme BspQI
```

```
java -jar OMTools.jar Статистика данных --optmapin Ecoli.ref --statout ReferenceStat.txt
```

```
java -jar OMTools.jar OptMapDataGenerator --refmapin Ecoli.ref --optmapout EcoliExample.sdata --cov 100
```

```
java -jar OMTools.jar Статистика данных --refmapin Ecoli.ref --optmapin EcoliExample.sdata --statout Статистика данных .txt
```

```
java -jar OMTools.jar OMBlastMapper --refmapin Ecoli.ref --optmapin EcoliExample.sdata --optresout EcoliExample.umd --thread 2
```

```
java -jar OMTools.jar ResultStatistics --refmapin Ecoli.ref --optmapin EcoliExample.sdata --optresin EcoliExample.umd --statout ResultStat.txt
```

```
java -jar OMTools.jar OMView --viewrefin Ecoli.ref --viewresin EcoliExample.umd --viewregion 1:1-1000000
```

### 4 Основные процедуры анализа

Перед генерацией данных оптического картирования пользователям рекомендуется провести моделирование и определить лучший фермент для использования. Если у пользователей есть файлы с борки последовательностей (например, короткие контиги и последовательности, длинные каркасы последовательностей или существующие референтные последовательности), они могут выполнить *in silico*-переваривание. Такое преобразование из fasta в данные оптического картирования можно выполнить с помощью модуля FastaToOM. Пользователи получают две части информации. Во-первых, с помощью модуля DataStatistics пользователи могут генерировать некоторые основные статистические данные переваренной последовательности. Одним из наиболее важных чисел, на которые следует обратить внимание, является плотность участков фермента (плотность сигнала). Анализ данных в оптическом картировании становится очень сложным, если плотность сигнала слишком высока или слишком низка (диапазон «хорошей» плотности сигнала зависит от платформы, используемой для генерации данных оптического картирования, и пользователям рекомендуется проконсультироваться с поставщиками для получения более подробной информации).

Во-вторых, если конструкция платформы основана на маркировке ферментом никеля, разрыв сайта никеля может стать серьезной проблемой для непрерывности борки оптического картирования. Для каждого сайта фермента модуль FastaToOM предоставляет расстояние до ближайшего сайта фермента. Пользователи могут предсказать количество разрывов сайта никеля, установив отсечку расстояния (опять же, расстояние между двумя сайтами никеля, которое может привести к разрыву, зависит от платформы, используемой для генерации данных оптического картирования, и пользователям рекомендуется проконсультироваться с поставщиками для получения более подробной информации). Кроме того, если пользователи хотят завершить борки генома с использованием данных оптического картирования, важно посмотреть на местоположение прогностических разрывов сайта никеля в интересующих контигах. Лучше избегать разрыва сайта никеля вблизи концов контига.

После определения используемого фермента пользователи могут перейти к эксперименту по оптическому картированию. Когда пользователи получают необработанные данные оптического картирования с высокой пропускной способностью, рекомендуется генерировать статистику по данным оптического картирования с помощью модуля DataStatistics. Обычно большой интерес представляют несколько критериев: пропускная способность данных, плотность сигнала и средняя длина молекулы. Затем пользователи могут использовать модуль DataTools для обработки и фильтрации данных. Иногда выходные данные дублируются, когда несколько необработанных файлов оптического картирования объединяются из-за некоторых неожиданных человеческих ошибок. В таком случае пользователи могут захотеть выполнить быстрое сканирование с помощью модуля DuplicatedMoleculesDetection и удалить любые дублированные молекулы с помощью модуля DuplicatedMoleculesRemover.

Далее, одним из анализов данных входного потока будет выравнивание. OMTools предоставляет несколько модулей выравнивания, включая модули OMBlastMapper, OMHAMapper и OMFMMapper. Модуль PairwiseAlignment также может помочь выполнить попарное выравнивание по нескольким наборам данных. Затем пользователи могут обрабатывать и фильтровать результаты выравнивания с помощью модуля ResultTools и генерировать

статистика с использованием модуля ResultStatistics. Поскольку несколько методов недавно были опубликованы другими исследователями группами, которые могут выполнять выравнивание данных оптического картирования, пользователи могут рассмотреть возможность получения объединения или пересечения результатов выравнивания из нескольких методов с использованием модуля ResultMerger.

Когда дело доходит до разработки новых алгоритмов, связанных с данными оптического картирования, пользователи могут быть заинтересованы в инструментах моделирования. К ним относятся модуль OptMapDataGenerator, который генерирует данные оптического картирования с учетом опорной оптической карты, и модуль RandomReferenceGenerator, который генерирует случайную опорную оптическую карту. Если пользователи разрабатывают алгоритм выравнивания, модуль PrecisionRecallGraphDataGenerator генерирует таблицу, которую можно использовать для генерации графика точности-воспоминания.

И последнее, но не менее важное: пользователи могут захотеть дополнительно исследовать результаты и продемонстрировать некоторые примеры, визуализируя данные оптического картирования. Модуль OMView служит для этого многоцелевым визуализатором данных оптического картирования.

## Часть 1

# Картограф

## 5 OMBlastMapper

Выполняет выравнивание данных оптического картирования. Алгоритм OMBlast использует подход «seed-and-extend» для выравнивания оптических карт.

5.1 Общие параметры Mapper --minsig Минимальный сигнал

запроса для выравнивания. [По умолчанию: 5]

--minsize Минимальный размер запроса для выравнивания. [По умолчанию: 50000]

--exactmatch Включить точное соответствие запроса ссылке. Отключить эту опцию при выполнении самовыравнивания. [По умолчанию: true]

5.1.1 Параметры модуля слияния перекрывающихся выравниваний

--overlapmerge mode Режим: 0: Отключить шаг слияния; 1: Объединить одинаковые частичные выравнивания; 2: Объединить перекрывающиеся частичные выравнивания [По умолчанию: 2]

--match Оценка за опосредование с игнором [По умолчанию: 5] --fpr Штраф

за дополнительный сигнал [По умолчанию: 2]

--fpr Штраф за отсутствие сигнала [По умолчанию: 2] --local Включить

локальное выравнивание [По умолчанию: true]

5.1.2 Параметры фильтра результатов

--filter mode Режим фильтра: 0: Без фильтра; 1: Фильтровать по всем следующим параметрам; 2: Фильтровать по минимуму только оценки [По умолчанию: 0]

--minmatch Минимальное количество совпадений частичного выравнивания [По умолчанию: 3] --maxfr Максимальное

количество дополнительных сигналов частичного выравнивания [По умолчанию: 10000]

--maxfn Максимальное количество пропущенных сигналов частичного выравнивания [По умолчанию: 10000]

--maxfpr Максимальная скорость дополнительных сигналов частичного выравнивания [По умолчанию: 1.0E-4] --maxfnr

Максимальная скорость пропущенных сигналов частичного выравнивания [По умолчанию: 0.5]

--minscore Минимальная оценка частичного выравнивания [По умолчанию: 0,0] --minsubfratio

Минимальное отношение субфрагментов частичного выравнивания [По умолчанию: 0,0]

--minsigratio Минимальное отношение выровненных сигналов при частичном выравнивании [По умолчанию: 0,0]

--trim mode Режим обрезки: 0: Режим обрезки отключен; 1: Режим обрезки включен [По умолчанию: 0] --maxtrim Максимальное

количество шагов обрезки частичного выравнивания [По умолчанию: 5]

--match Оценка за опосредование с игнором [По умолчанию: 5] --fpr Штраф

за дополнительный сигнал [По умолчанию: 2]

--fpr Штраф за отсутствие сигнала [По умолчанию: 2]

## 5.1.3 Варианты соединения выравнивания

--alignmentjoinmode Режим. 0: Без объединения. 1: Стандартное соединение indel. 2: Стандартное соединение indel-inv.

3: Стандартное транслокационное присоединение [По умолчанию: 0]

--closefrag Максимальное расстояние (ссылка) между двумя частичными выравниваниями, которые необходимо объединить [По умолчанию: 250000]

--closefrag Максимальное расстояние (запрос) между двумя частичными выравниваниями, которые необходимо объединить [По умолчанию: 250000]

--minmatch Минимальное количество символов соответствия, которые следует считать допустимым частичным сопоставлением. [По умолчанию: 3]

--maxtrim Максимальные шаг и обрезки для частичного выравнивания [По умолчанию: 5]

--trimear Допустимая погрешность масштабирования во время обрезки [По умолчанию: 0,1]

--match Оценки сопоставления символов [По умолчанию: 5]

--fpp Штраф за дополнительный символ [По умолчанию: 2]

--fnp Штраф за отсутствие символа [По умолчанию: 2]

--indelp Штраф за присоединение частичных выравниваний с отношением indel [По умолчанию: 10]

--invp Штраф за объединение частичных выравниваний с инверсионным отношением [По умолчанию: 30]

--transp Штраф за присоединение частичных выравниваний с транслокационной связью [По умолчанию: 50]

--localpenalty Включить штраф за локальное выравнивание для окончательного выравнивания (расматривается как глобальное выравнивание) [По умолчанию: ложно]

--minjoinscore Минимальная оценка объединенного окончательного выравнивания [По умолчанию: 30]

--minconf Минимальная достоверность (уникальность) окончательного выравнивания [По умолчанию: 0,4]

--minjoinedfragratio Минимальное отношение выровненной длины к длине запроса [По умолчанию: -1,0]

--minjoinedsigratio Минимальное отношение количества выровненных символов к общему количеству символов запроса [По умолчанию: -1,0]

--overlapalign Разрешить вывод перекрывающихся конечных выравниваний [По умолчанию: true]

--maxalignitem Максимальное количество выводимых окончательных выравниваний. -1: Нет ограничений на количество окончательных выравниваний. выравнивания [По умолчанию: 1]

## 5.1.4 Параметры читателя результатов

--optresin Bx одной файл результатов выравнивания для повторного выравнивания

--optresinformat -1: Автоматически определяется по расширению файла; 0: Формат выравнивания OM (OMA); 1: Формат подробного выравнивания OM (OMD); 2: Формат XMAP (XMAP); 3: Формат Valouev et al.; 4: Формат уникального соответствия SOMA v2; 5: Формат Twin PSL; 6: Формат Maligner ALN; [По умолчанию: -1]



## 5.2 Параметры OMBlastMapper --local

Включить локальное выравнивание [По умолчанию: true]

--allowequalrefquery Разрешить равную ссылку и запросы в выравнивании. Отключение этой опции запрещает запрос от выравнивания по себе как по ссылке [По умолчанию: true]

--allowdiffrefquery Разрешить различные ссылки и запросы в выравнивании. В отличие от allowequalref-query, отключение этой опции запрещает запросу выравниваться по другим запросам в качестве ссылок [По умолчанию: true]

--meas Ошибка измерения [По умолчанию: 500]

--eag Допустимый диапазон ошибок (допустимая погрешность масштабирования) [По умолчанию: 0,1]

--match Оценка за сопоставление с игнором [По умолчанию: 5]

--fpr Штраф за дополнительный сигнал [По умолчанию: 2]

--fnr Штраф за отсечение сигнала [По умолчанию: 2]

--falselimit Максимальное количество последовательных дополнительных/отсеченных сигналов [По умолчанию: 5]

--maxseedno Максимальное количество возможных начальных значений в запросе [По умолчанию: 10]

### 5.2.1 Параметры заполнения

--seedingmode Режим заполнения: 1: Оптимизировано для длинных k-меров (обычно для k больше 10); 2: Оптимизировано для коротких k-меров (обычно для k меньше или равно 10); -1: Автоматический выбор. [По умолчанию: -1]

--k Длина К-мера. [По умолчанию: 3]

--maxnosignal Максимальная область отсечения сигнала между сигналами для заполнения. [По умолчанию: 10000000]

## 5.3 Параметры многопоточности

--thread Количество потоков [По умолчанию: 1]

## 5.4 Параметры чтения ссылок --refmapin

Входной файл карты ссылок [Обязательно]

--refmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

#### 5.5 Параметры считывателя данных --

optmapin Вх одной файл оптической карты [обязательно]

--optmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS);  
3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ  
Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OpGen XML; 10: Формат данных Валуева;  
11: Формат карт Maligner; [По умолчанию: -1]

--bnxsnr Значение фильтра BNХ SNR [по умолчанию: 3.0]

#### 5.6 Параметры записи и результатов --optresout

Вывод файла результатов выравнивания [Обязательно]

--optresoutformat Формат файла результата -1: Автоматически определяется по расширению файла; 0: Формат выравнивания OM (OMA); 1: Формат подробного выравнивания OM (OMD); 2: Формат XMAP (XMAP); 3: Формат Valouev et al.; 4: Формат уникального соответствия SOMA v2; 5: Формат Twin PSL; 6: Формат Maligner ALN; [По умолчанию: -1]

--writeunmap Записывать отброшенные или неотображенные молекулы. [По умолчанию: true]

--multiple Записывать несколько карт для молекулы. [По умолчанию: true]

--writeinfo Записывать информацию о молекуле. [По умолчанию: true]

## 6 OMHMapper

Выполняет выравнивание данных оптического картирования. Алгоритм OMHA использует эвристический подход для выравнивания оптических карт.

### 6.1 Общие параметры картографа

--minsig Минимальный сигнал запроса для выравнивания. [По умолчанию: 5]

--minsize Минимальный размер запроса для выравнивания. [По умолчанию: 50000]

--exactmatch Включить точное соответствие запроса с ссылкой. Отключить эту опцию при выполнении самовыравнивания. [По умолчанию: true]

#### 6.1.1 Параметры модуля слияния перекрывающихся выравниваний

--overlapmergemode Режим: 0: Отключить шаг слияния; 1: Объединить одинаковые частичные выравнивания; 2: Объединить перекрывающиеся частичные выравнивания. [По умолчанию: 2]

--match Оценка за сопоставление сигнала [По умолчанию: 5]

--fpr Штраф за дополнительный сигнал [По умолчанию: 2]

--fnp Штраф за отсутствующие сигналы [По умолчанию: 2]

--local Включить локальное выравнивание. [По умолчанию: true]

#### 6.1.2 Параметры фильтра результатов

--filtermode Режим фильтра: 0: Без фильтра; 1: Фильтровать по всем следующим параметрам; 2: Фильтровать по минимуму только оценки. [По умолчанию: 0]

--minmatch Минимальное количество совпадений частичного выравнивания. [По умолчанию: 3]

--maxfr Максимальное количество дополнительных сигналов частичного выравнивания. [По умолчанию: 10000]

--maxfn Максимальное количество пропущенных сигналов частичного выравнивания. [По умолчанию: 10000]

--maxfpr Максимальная скорость дополнительных сигналов частичного выравнивания. [По умолчанию: 1.0E-4]

--maxfnp Максимальная скорость пропуска сигналов частичного выравнивания. [По умолчанию: 0,5]

--minscore Минимальная оценка частичного выравнивания. [По умолчанию: 0,0]

--minsubfragratio Минимальное отношение субфрагментов частичного выравнивания. [По умолчанию: 0,0]

--minsigratio Минимальное отношение выровненных сигналов при частичном выравнивании. [По умолчанию: 0,0]

--trimmode Режим обрезки: 0: Режим обрезки отключен; 1: Режим обрезки включен. [По умолчанию: 0]

--maxtrim Максимальное количество шагов обрезки частичного выравнивания. [По умолчанию: 5]

--match Оценка за сопоставление сигнала [По умолчанию: 5]

--fpr Штраф за дополнительный сигнал [По умолчанию: 2]

--fnp Штраф за отсутствующие сигналы [По умолчанию: 2]

## 6.1.3 Варианты соединения выравнивания

--alignmentjoinmode Режим. 0: Без объединения. 1: Стандартное соединение indel. 2: Стандартное соединение indel-inv.

3: Стандартное транслокационное присоединение [По умолчанию: 0]

--closefrag Максимальное расстояние (ссылка) между двумя частичными выравниваниями, которые необходимо объединить [По умолчанию: 250000]

--closefrag Максимальное расстояние (запрос) между двумя частичными выравниваниями, которые необходимо объединить [По умолчанию: 250000]

--minmatch Минимальное количество символов соответствия, которые следует считать допустимым частичным сопоставлением. [По умолчанию: 3]

--maxtrim Максимальные шаг и обрезки для частичного выравнивания [По умолчанию: 5]

--trimear Допустимая погрешность масштабирования во время обрезки [По умолчанию: 0,1]

--match Оценки сопоставления символов [По умолчанию: 5]

--fpp Штраф за дополнительный символ [По умолчанию: 2]

--fnp Штраф за отсутствие символа [По умолчанию: 2]

--indelp Штраф за присоединение частичных выравниваний с отношением indel [По умолчанию: 10]

--invp Штраф за объединение частичных выравниваний с инверсионным отношением [По умолчанию: 30]

--transp Штраф за присоединение частичных выравниваний с транслокационной связью [По умолчанию: 50]

--localpenalty Включить штраф за локальное выравнивание для окончательного выравнивания (расматривается как глобальное выравнивание) [По умолчанию: ложно]

--minjoinscore Минимальная оценка объединенного окончательного выравнивания [По умолчанию: 30]

--minconf Минимальная достоверность (уникальность) окончательного выравнивания [По умолчанию: 0,4]

--minjoinedfragratio Минимальное отношение выровненной длины к длине запроса [По умолчанию: -1,0]

--minjoinsigratio Минимальное отношение количества выровненных символов к общему количеству символов запроса [По умолчанию: -1,0]

--overlapalign Разрешить вывод перекрывающихся конечных выравниваний [По умолчанию: true]

--maxalignitem Максимальное количество выводимых окончательных выравниваний. -1: Нет ограничений на количество окончательных выравниваний. выравнивания [По умолчанию: 1]

## 6.1.4 Параметры читателя результатов

--optresin Bx одной файл результатов выравнивания для повторного выравнивания

--optresinformat -1: Автоматически определяется по расширению файла; 0: Формат выравнивания OM (OMA); 1: Формат подробного выравнивания OM (OMD); 2: Формат XMAP (XMAP); 3: Формат Valouev et al.; 4: Формат уникального соответствия SOMA v2; 5: Формат Twin PSL; 6: Формат Maligner ALN; [По умолчанию: -1]

## 6.2 Параметры OMHMapper --local Включить

локальное выравнивание [По умолчанию: true]

--localstart Локальная начальная позиция для выравнивания, 0: начинается с каждого сигнала (исчерпывающий), x: начинается с первого сигнала, -x: запускается без последних сигналов. [По умолчанию: 0]

--scorefilter Первичный фильтр оценок во время выравнивания [По умолчанию: 30]

--deg Вырождение с сигналов закрытия для обработки ошибки разрешения. [По умолчанию: 1500]

--meas Ошибка измерения [По умолчанию: 500]

--eag Допустимый диапазон ошибок (допустимая погрешность масштабирования) [По умолчанию: 0,1]

--match Оценка за сопоставление с сигналом [По умолчанию: 5]

--fpr Штраф за дополнительный сигнал [По умолчанию: 2]

--fnp Штраф за отсутствие сигнала [По умолчанию: 2]

--falseimit Максимальное количество ложных сигналов [по умолчанию: 5]

## 6.3 Параметры многопоточности

--thread Количество потоков [По умолчанию: 1]

## 6.4 Параметры чтения ссылок --refmapin Вх одной

файл карты ссылок [Обязательно]

--refmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

## 6.5 Параметры считывателя данных --

optmapin Вх одной файл оптической карты [обязательно]

--optmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

--bnxsnr Значение фильтра BNХ SNR [по умолчанию: 3.0]

## 6.6 Параметры записи и результатов --

optresout Вывод файла результатов выравнивания [Обязательно]

--optresoutformat Формат файла результата -1: Автоматически определяется по расширению файла; 0: Формат выравнивания OM (OMA); 1: Формат подробного выравнивания OM (OMD); 2: Формат XMAP (XMAP); 3: Формат Valouev et al.; 4: Формат уникального ответа SOMA v2; 5: Формат Twin PSL; 6: Формат Maligner ALN; [По умолчанию: -1]

--writeunmap Записывать отброшенные или не отображенные молекулы. [По умолчанию: true]

--multiple Записывать несколько карт для молекулы. [По умолчанию: true]

--writeinfo Записывать информацию о молекуле. [По умолчанию: true]

## 7 OMFMMapper

Выполняет выравнивание данных оптического картирования. Алгоритм OMFMM использует индексный подход для выравнивания оптических карт.

### 7.1 Общие параметры картографа

--minsig Минимальный сигнал запроса для выравнивания. [По умолчанию: 5]

--minsize Минимальный размер запроса для выравнивания. [По умолчанию: 50000]

--exactmatch Включить точное соответствие запроса ссылке. Отключить эту опцию при выполнении самовыравнивания. [По умолчанию: true]

#### 7.1.1 Параметры модуля слияния перекрывающихся выравниваний

--overlapmergemode Режим: 0: Отключить шаг слияния; 1: Объединить одинаковые частичные выравнивания; 2: Объединить перекрывающиеся частичные выравнивания. [По умолчанию: 2]

--match Оценка с опосредованием сигнала [По умолчанию: 5]

--fpr Штраф за дополнительный сигнал [По умолчанию: 2]

--fnp Штраф за отсутствие сигнала [По умолчанию: 2]

--local Включить локальное выравнивание [По умолчанию: true]

#### 7.1.2 Параметры фильтра результатов

--filtermode Режим фильтра: 0: Без фильтра; 1: Фильтровать по всем следующим параметрам; 2: Фильтровать по минимуму только оценок [По умолчанию: 0]

--minmatch Минимальное количество совпадений частичного выравнивания. [По умолчанию: 3]

--maxfr Максимальное количество дополнительных сигналов частичного выравнивания. [По умолчанию: 10000]

--maxfn Максимальное количество пропущенных сигналов частичного выравнивания. [По умолчанию: 10000]

--maxfpr Максимальная скорость дополнительных сигналов частичного выравнивания. [По умолчанию: 1.0E-4]

--maxfnp Максимальная скорость пропуска сигналов частичного выравнивания. [По умолчанию: 0,5]

--minscore Минимальная оценка частичного выравнивания. [По умолчанию: 0,0]

--minsubfragratio Минимальное отношение субфрагментов частичного выравнивания. [По умолчанию: 0,0]

--minsigratio Минимальное отношение выровненных сигналов при частичном выравнивании. [По умолчанию: 0,0]

--trimmode Режим обрезки: 0: Режим обрезки отключен; 1: Режим обрезки включен [По умолчанию: 0]

--maxtrim Максимальное количество шагов обрезки частичного выравнивания. [По умолчанию: 5]

--match Оценка с опосредованием сигнала [По умолчанию: 5]

--fpr Штраф за дополнительный сигнал [По умолчанию: 2]

--fnp Штраф за отсутствие сигнала [По умолчанию: 2]

## 7.1.3 Варианты с объединения выравнивания

--alignmentjoinmode Режим. 0: Без объединения. 1: Стандартное с объединением indel. 2: Стандартное с объединением indel-inv.

3: Стандартное транслокационное присоединение [По умолчанию: 0]

--closefrag Максимальное расстояние (ссылка) между двумя частичными выравниваниями, которые необходимо объединить [По умолчанию: 250000]

--closefrag Максимальное расстояние (запрос) между двумя частичными выравниваниями, которые необходимо объединить [По умолчанию: 250000]

--minmatch Минимальное количество совпадений, которые следует считать допустимым частичным сопоставлением. [По умолчанию: 3]

--maxtrim Максимальные шаг и обрезки для частичного выравнивания [По умолчанию: 5]

--trimear Допустимая погрешность масштабирования во время обрезки [По умолчанию: 0,1]

--match Оценка сопоставления с игнорированием [По умолчанию: 5]

--fpp Штраф за дополнительный сигнал [По умолчанию: 2]

--fnp Штраф за отсутствие сигнала [По умолчанию: 2]

--indelp Штраф за присоединение частичных выравниваний с отношением indel [По умолчанию: 10]

--invp Штраф за объединение частичных выравниваний с инверсионным отношением [По умолчанию: 30]

--transp Штраф за присоединение частичных выравниваний с транслокационной связью [По умолчанию: 50]

--localpenalty Включить штраф за локальное выравнивание для окончательного выравнивания (расматривается как глобальное выравнивание) [По умолчанию: ложно]

--minjoinscore Минимальная оценка объединенного окончательного выравнивания [По умолчанию: 30]

--minconf Минимальная достоверность (уникальность) окончательного выравнивания [По умолчанию: 0,4]

--minjoinedfragratio Минимальное отношение выровненной длины к длине запроса [По умолчанию: -1,0]

--minjoinedsigratio Минимальное отношение количества выровненных сигналов к общему количеству сигналов запроса [По умолчанию: -1,0]

--overlapalign Разрешить вывод перекрывающихся конечных выравниваний [По умолчанию: true]

--maxalignitem Максимальное количество выводимых окончательных выравниваний. -1: Нет ограничений на количество окончательных выравниваний. выравнивания [По умолчанию: 1]

## 7.1.4 Параметры считывателя результатов

--optresin Bx одной файл результатов выравнивания для повторного выравнивания

--optresinformat -1: Автоматически определяется по расширению файла; 0: Формат выравнивания OM (OMA); 1: Формат подробного выравнивания OM (OMD); 2: Формат XMAP (XMAP); 3: Формат Valouev et al.; 4: Формат уникального сопоставления SOMA v2; 5: Формат Twin PSL; 6: Формат Maligner ALN; [По умолчанию: -1]



## 7.2 Параметры OMFMMapper --meas Ошибка

измерения [По умолчанию: 500]

--eag Допустимый диапазон ошибок (допустимая погрешность масштабирования) [По умолчанию: 0,1]

--match Оценка с опосредованием сигнала [По умолчанию: 5]

--frr Штраф за дополнительный сигнал [По умолчанию: 2]

--fnr Штраф за отсутствие сигнала [По умолчанию: 2]

--rfalseimit Максимальное количество ложных сигналов на ссылку [По умолчанию: 5]

--qfalseimit Максимальное количество ложных сигналов на запрос [По умолчанию: 5]

--cfalseimit Максимальное количество ложных сигналов для ссылки и запроса [По умолчанию: 5]

--minalignscore Минимальная оценка на этапе выравнивания [По умолчанию: 20]

### 7.2.1 Параметры заполнения

--seedingmode Режим заполнения: 1: Оптимизировано для длинных k-меров (обычно для k больше 10); 2: Оптимизировано для коротких k-меров (обычно для k меньше или равно 10); -1: Автоматический выбор. [По умолчанию: -1]

--k Длина Кмера. [По умолчанию: 3]

--maxnosignal Максимальная область отсутствия сигнала между сигналами для заполнения. [По умолчанию: 10000000]

## 7.3 Параметры многопоточности --thread

Количество потоков [По умолчанию: 1]

## 7.4 Параметры считывателя ссылок --refmapin

Вх одной файл карты ссылок [Обязательно]

--refmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

## 7.5 Параметры считывателя данных --

optmapin Вх одной файл оптической карты [Обязательно]

--optmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

--bnxsnr Значение фильтра BNХ SNR [по умолчанию: 3.0]

## 7.6 Параметры записи и результатов --

optresout Вывод файла результатов выравнивания [Обязательно]

--optresoutformat Формат файла результата -1: Автоматически определяется по расширению файла; 0: Формат выравнивания OM (OMA); 1: Формат подробного выравнивания OM (OMD); 2: Формат XMAP (XMAP); 3: Формат Valouev et al.; 4: Формат уникального ответа SOMA v2; 5: Формат Twin PSL; 6: Формат Maligner ALN; [По умолчанию: -1]

--writeunmap Записывать отброшенные или не отображенные молекулы. [По умолчанию: true]

--multiple Записывать несколько карт для молекулы. [По умолчанию: true]

--writeinfo Записывать информацию о молекуле. [По умолчанию: true]

## 8 Парное выравнивание

Выполняет парное выравнивание файлов данных на основе OMBlastMapper. Введите несколько файлов данных для парного выравнивания между каждой парой из них.

### 8.1 Общие параметры картографа

--minsig Минимальный сигнал запроса для выравнивания. [По умолчанию: 5]

--minsize Минимальный размер запроса для выравнивания. [По умолчанию: 50000]

--exactmatch Включить точное соответствие запроса с ссылкой. Отключить эту опцию при выполнении самовыравнивания. [По умолчанию: true]

#### 8.1.1 Параметры модуля слияния перекрывающихся выравниваний

--overlapmergemode Режим: 0: Отключить шаг слияния; 1: Объединить одинаковые частичные выравнивания; 2: Объединить перекрывающиеся частичные выравнивания. [По умолчанию: 2]

--match Оценка с опосредованием сигнала [По умолчанию: 5]

--fpr Штраф за дополнительный сигнал [По умолчанию: 2]

--fnp Штраф за отсутствие сигнала [По умолчанию: 2]

--local Включить локальное выравнивание [По умолчанию: true]

#### 8.1.2 Параметры фильтра результатов

--filtermode Режим фильтра: 0: Без фильтра; 1: Фильтровать по всем следующим параметрам; 2: Фильтровать по минимуму только оценка [По умолчанию: 0]

--minmatch Минимальное количество совпадений частичного выравнивания. [По умолчанию: 3]

--maxfr Максимальное количество дополнительных сигналов частичного выравнивания. [По умолчанию: 10000]

--maxfn Максимальное количество пропущенных сигналов частичного выравнивания. [По умолчанию: 10000]

--maxfpr Максимальная скорость дополнительных сигналов частичного выравнивания. [По умолчанию: 1.0E-4]

--maxfnp Максимальная скорость пропуска сигналов частичного выравнивания. [По умолчанию: 0,5]

--minscore Минимальная оценка частичного выравнивания. [По умолчанию: 0,0]

--minsubfragratio Минимальное отношение субфрагментов частичного выравнивания. [По умолчанию: 0,0]

--minsigratio Минимальное отношение выровненных сигналов при частичном выравнивании [По умолчанию: 0,0]

--trimmode Режим обрезки: 0: Режим обрезки отключен; 1: Режим обрезки включен [По умолчанию: 0]

--maxtrim Максимальное количество шагов обрезки частичного выравнивания. [По умолчанию: 5]

--match Оценка с опосредованием сигнала [По умолчанию: 5]

--fpr Штраф за дополнительный сигнал [По умолчанию: 2]

--fnp Штраф за отсутствие сигнала [По умолчанию: 2]

## 8.1.3 Варианты соединения выравнивания

--alignmentjoinmode Режим. 0: Без объединения. 1: Стандартное соединение indel. 2: Стандартное соединение indel-inv.

3: Стандартное транслокационное присоединение [По умолчанию: 0]

--closefrag Максимальное расстояние (ссылка) между двумя частичными выравниваниями, которые необходимо объединить [По умолчанию: 250000]

--closefrag Максимальное расстояние (запрос) между двумя частичными выравниваниями, которые необходимо объединить [По умолчанию: 250000]

--minmatch Минимальное количество символов соответствия, которые следует считать допустимым частичным сопоставлением. [По умолчанию: 3]

--maxtrim Максимальные шаг и обрезки для частичного выравнивания [По умолчанию: 5]

--trimear Допустимая погрешность масштабирования во время обрезки [По умолчанию: 0,1]

--match Оценка сопоставления символов [По умолчанию: 5]

--fpp Штраф за дополнительный символ [По умолчанию: 2]

--fnp Штраф за отсутствие символа [По умолчанию: 2]

--indelp Штраф за присоединение частичных выравниваний с отношением indel [По умолчанию: 10]

--invp Штраф за объединение частичных выравниваний с инверсионным отношением [По умолчанию: 30]

--transp Штраф за присоединение частичных выравниваний с транслокационной связью [По умолчанию: 50]

--localpenalty Включить штраф за локальное выравнивание для окончательного выравнивания (расматривается как глобальное выравнивание) [По умолчанию: ложно]

--minjoinscore Минимальная оценка объединенного окончательного выравнивания [По умолчанию: 30]

--minconf Минимальная достоверность (уникальность) окончательного выравнивания [По умолчанию: 0,4]

--minjoinedfragratio Минимальное отношение выровненной длины к длине запроса [По умолчанию: -1,0]

--minjoinedsigratio Минимальное отношение количества выровненных символов к общему количеству символов запроса [По умолчанию: -1,0]

--overlapalign Разрешить вывод перекрывающихся конечных выравниваний [По умолчанию: true]

--maxalignitem Максимальное количество выводимых окончательных выравниваний. -1: Нет ограничений на количество окончательных выравниваний. выравнивания [По умолчанию: 1]

## 8.1.4 Параметры считывателя результатов

--optresin Bx одной файл результатов выравнивания для повторного выравнивания

--optresinformat -1: Автоматически определяется по расширению файла; 0: Формат выравнивания OM (OMA); 1: Формат подробного выравнивания OM (OMD); 2: Формат XMAP (XMAP); 3: Формат Valouev et al.; 4: Формат уникального соответствия SOMA v2; 5: Формат Twin PSL; 6: Формат Maligner ALN; [По умолчанию: -1]

## 8.2 Параметры OMBlastMapper --local

Включить локальное выравнивание [По умолчанию: true]

--allowequalrefquery Разрешить равную ссылку и запрос в выравнивании. Отключение этой опции запрещает запрос от выравнивания по себе как по ссылке [По умолчанию: true]

--allowdiffrefquery Разрешить различные ссылки и запросы в выравнивании. В отличие от allowequalref-query, отключение этой опции запрещает запросу выравниваться по другим запросам в качестве ссылок [По умолчанию: true]

--meas Ошибка измерения [По умолчанию: 500]

--ear Допустимый диапазон ошибок (допустимая погрешность масштабирования) [По умолчанию: 0,1]

--match Оценка с опосредованием сигнала [По умолчанию: 5]

--fpr Штраф за дополнительный сигнал [По умолчанию: 2]

--fnr Штраф за отсечение сигнала [По умолчанию: 2]

--falselimit Максимальное количество последовательных дополнительных /отсеченных сигналов [По умолчанию: 5]

--maxseedno Максимальное количество похожих начальных значений в запросе [По умолчанию: 10]

### 8.2.1 Параметры заполнения

--seedingmode Режим заполнения: 1: Оптимизировано для длинных k-меров (обычно для k больше 10); 2: Оптимизировано для коротких k-меров (обычно для k меньше или равно 10); -1: Автоматический выбор. [По умолчанию: -1]

--k Длина Кмера. [По умолчанию: 3]

--maxposignal Максимальная область отсечения сигнала между сигналами для заполнения. [По умолчанию: 10000000]

## 8.3 Параметры многопоточности

--thread Количество потоков [По умолчанию: 1]

## 8.4 Параметры считывателя данных

--optmapin Вх одной файл оптической карты [обязательно]

--optmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OpGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

--bnxsnr Значение фильтра BNХ SNR [по умолчанию: 3.0]

## 8.5 Варианты парного выравнивания

--output префикс вывода [Обязательно]

--rerun Повторно запускать, даже если файл результатов существует [По умолчанию: false]

## Часть 2

# Моделирование

## 9 OptMapDataGenerator

Генерирует моделированные данные из эталона.

### 9.1 Параметры чтения с правочной литературы

--refmapin Вх одной файл с правочной карты

--refmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

### 9.2 Несколько вариантов чтения с правочных материалов

--refmaplistin Вх одной список файлов с правочной карты с соотношением сторон

--refmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

### 9.3 Параметры генератора данных --rsln Ошибка

разрешения [По умолчанию: 1200]

--meas Ошибка измерения [По умолчанию: 500]

--fsize Средний размер фрагмента [По умолчанию: 200000]

--fubound Верхняя граница размера, включительно [По умолчанию: 1000000]

--flbound Размер нижней границы, включительно [По умолчанию: 100000]

--median Медиана для шкалы [По умолчанию: 1.0]

--scalesd SD для масштаба [по умолчанию: 0,04]

--subound Верхняя граница шкалы, включительно [По умолчанию: 1.3]

--slbound Масштаб нижней границы, включительно [По умолчанию: 0,7]

--fpr час тотальных срабатываний [По умолчанию: 1.0E-5]

--fnr час тотальных отрицательных результатов [По умолчанию: 0,1]

--seed Случайное семя

--indelsize Случайный размер вставки/удаления [По умолчанию: 0]

--inversionmode Режим инверсии. 0: без инверсии. 1: инверсия второй половины [По умолчанию: 0]

--cov Охват вывода данных [По умолчанию: 10.0]

--moleno Количество молекул, которые будут сгенерированы. Переопределяет параметр покрытия, если установлено положительное число [По умолчанию: -1]

#### 9.4 Параметры записи данных --optmapout

Выходной файл оптической карты [обязательно]

--optmapoutformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: Формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNХ); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OpGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

## 10 Генератор случайных ссылок

Генерирует случайные справочные карты путем перемешивания порядка сегментов во всех справочных картах.

### 10.1 Параметры чтения справочной литературы

--refmapin Вх: одна файл с справочной карты [обязательно]

--refmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

### 10.2 Параметры референсного автора

--refmapout Вывод файла с справочной карты [обязательно]

--refmapoutformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: Формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]



## Часть 3

## Обнаружение СВ

### 11 SVDetection

Представляет базовый модуль обнаружения СВ для оптического картographic выравнивания

#### 11.1 Параметры чтения с справочной литературы

--refmapin Вх одной файл с справочной карты [обязательно]

--refmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

#### 11.2 Параметры считывателя данных

--optmapin Вх одной файл оптической карты

--optmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

--bnxsnr Значение фильтра BNХ SNR [по умолчанию: 3.0]

#### 11.3 Параметры считывателя результатов

--optresin Вх одной файл результатов выравнивания [Обязательно]

--optresinformat -1: Автоматически определяется по расширению файла; 0: Формат выравнивания OM (OMA); 1: Формат подробного выравнивания OM (OMD); 2: Формат XMAP (XMAP); 3: Формат Valouev et al.; 4: Формат уникального ответа SOMA v2; 5: Формат Twin PSL; 6: Формат Maligner ALN; [По умолчанию: -1]

#### 11.4 Стандартные параметры SV Writer

--svout Выходной SV-файл

#### 11.5 Обнаружение СВ

--svmode Режим SV. 1: подход с разделением карты (экспериментальный); 2: подход на основе сигнала; 3: подход с разделением карты и подход на основе сигнала [По умолчанию: 3]

--minsupport Минимальная поддержка молекул для SV [По умолчанию: 5]

--flanksig Минимальное количество фланговых сигналов для выравнивания [По умолчанию: 5]

#### 11.5.1 Параметры обнаружения на основе разделенной карты

--closeref Максимальное расстояние (ссылка) между двумя результатами, которые следует разматывать в одном классе. [Определение ошибки: 250000]

--closefrag Максимальное расстояние (фрагмент) между двумя результатами, которые следует разматывать в одном классе. [По умолчанию: 250000]

--closesv Закрыть SV (SV объединяются на основе равенства bp1bp2. [По умолчанию: 10000]

--mergesv Объединить SV (SV объединяются на основе простого равенства регионов). [По умолчанию: true]

--flanksize Минимальный размер фланга в выравнивании (используется только при обнаружении разделенной карты) [По умолчанию: 0]

--minindelsize Минимальный размер вставки для SV [по умолчанию: 1500]

--maxindelsize Максимальный размер вставки для SV [по умолчанию: 1000000]

--maxsupport Максимальная поддержка молекул для SV [По умолчанию: 100]

--minsvscore Минимальный балл sv (с отношением поддержки/опозиции) [По умолчанию: 0,5]

#### 11.5.2 Параметры обнаружения на основе сигнала

--mininvsig Минимальный сигнал, участвующий в инверсии [По умолчанию: 4]

--maxinvsizе Максимальный размер инверсии [По умолчанию: 100000]

--meas Ошибка измерения [По умолчанию: 500]

--eag Допустимый диапазон ошибок (допустимая погрешность масштабирования) [По умолчанию: 0,1]

--deg Вырождение сигналов закрытия для обработки ошибки разрешения. [По умолчанию: 1500]

## Часть 4

### Инструменты Fasta

## 12 FastaToOM

Выполняет *in silico*-расщепление последовательности ДНК.

### 12.1 Параметры потокового Fasta Reader

--fastain *вх* одной файл fasta [обязательно]

### 12.2 Варианты ввода фермента

--enzyme *Вс* троечные ферменты [BspQI, BbvCI, AlwI, BsmAI, BstNBI, BsmI, BsrDI, BssSI, BtsI] (поддержка ввода нескольких ферментов)

--enzymestring *Последовательность фермента* (например, GCTCTTC)

### 12.3 Параметры автора ссылок

--refmapout Вывод файла с правочной карты [обязательно]

--refmapoutformat *-1*: Автоматически определяется по расширению файла; *0*: Справочный стандартный формат (REF) (эквивалент формата SILICO); *1*: формат fasta-01 (FA01); *2*: Формат файла Spots (SPOTS); *3*: Стандартный формат молекулы (DATA); *4*: Формат моделирования молекулы (SDATA); *5*: Файл BNХ Формат (BNX); *6*: Формат файла CMAP (CMAP); *7*: Формат SOMA opt (OPT); *8*: Формат SOMA silico (SILICO) (эквивалент формата REF); *9*: Формат OpGen XML; *10*: Формат данных Валугеа; *11*: Формат карт Maligner; [По умолчанию: *-1*]

### 12.4 Параметры прогнозирования разрыва с айт-анинг а

--nsbout Потенциальный с айт-разрыва прерывает вывод (прогноз полезен для данных на основе фермента разрыва) только)

## Часть 5

### Инструменты данных

## 13 DataTools

Предоставляет базовые функции для фильтрации и обработки данных оптического картирования.

### 13.1 Параметры инструментов данных

--idprefix Добавить префикс ко всем идентификаторам

--idmodify Преобразовать все идентификаторы в  $x \dots x + n - 1$  (x: входное значение, n: количество оптических карт в данных файл). Отрицательное значение отключает эту функцию. [По умолчанию: -1]

--idmodifylog Файл журнала, содержащий преобразования идентификаторов

--staticid Использовать статический идентификатор для всех оптических карт (перезаопределять все остальные параметры, связанные с идентификатором)

--fix Исправить данные (отрицательная коррекция расстояния с игнорированием и т.д.) [По умолчанию: true]

--condense Объединить несколько сегментов, расположенных ближе, чем параметр, в один сегмент [По умолчанию: 0]

--removeseg Удалить сегменты, меньшие, чем параметр [По умолчанию: -1]

--minsize Данные с минимальным размером для сохранения [По умолчанию: 0]

--minsig Данные с минимальным сигналом для сохранения [По умолчанию: 0]

--dataid Список идентификаторов данных, которые необходимо извлечь

--region Список регионов для извлечения.

--shift Сдвиг вперед (вправо) x bp (предположим, круговой) [По умолчанию: 0]

--randdata Количество случайных данных, которые необходимо извлечь

--seed Начальное число, используемое при извлечении случайных данных.

--concat Объединить все записи данных в одну запись. -1: не активировано; Неотрицательное значение: пробел (сегмент без сигнала) между каждой записью данных. Игнорировать любые функции модификации данных [По умолчанию: -1]

#### 13.1.1 Параметры считывателя ConcatInfo

--concatin Вх одной файл ConcatInfo.

#### 13.1.2 Параметры записи ConcatInfo

--concatout Вывод файла ConcatInfo.

#### 13.1.3 Фильтрация низкой сложности

--lowscot Сократить/удалить молекулы с низкой сложностью -1: Сократить низкую сложность; 0: Ничего не делать; 1: Сократить высокую сложность [По умолчанию: 0]

--maxdensity Максимальная плотность на 100 кбит/с для фильтрации [по умолчанию: 25,0]

--maxseed Максимальное семя для фильтрации [по умолчанию: 5]

### 13.2 Параметры с читывателя данных

--optmapin Вх одной файл оптической карты [обязательно]

--optmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OpGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

--bnxsnr Значение фильтра BNХ SNR [по умолчанию: 3.0]

### 13.3 Параметры записи и данных --

optmapout Выходной файл оптической карты [обязательно]

--optmapoutformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: Формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OpGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

## 14 Проверка качества данных

Выполняет базовую проверку качества данных, включая распределение длины молекул, плотности сигнала и длины сегмента.

### 14.1 Параметры чтения с справочной литературы

--refmapin Вх одной файл с справочной карты

--refmapinform -1: Автоматически определяются по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

### 14.2 Параметры считывателя данных

--optmapin Вх одной файл оптической карты [обязательно]

--optmapinform -1: Автоматически определяются по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

--bnxsnr Значение фильтра BNХ SNR [по умолчанию: 3.0]

### 14.3 Параметры проверки качества данных

--name Ис использовать имя набора данных вместо имени файла.

--prefix Префикс вывода статистики [По умолчанию: вывод]

--gradcolor Ис использовать градиентный цвет [По умолчанию: false]

--imageformat Форматы сохранения изображения. [svg; png; jpg; ] [По умолчанию: png]

## 15 Данные Статистика

Формирует статистику файла данных.

### 15.1 Параметры статистики данных

--winsize Размер окна для размера молекулы [По умолчанию: 25000]

--maxsize Максимальный размер. -1 для автоматической настройки [По умолчанию: -1]

--maxsignal Максимальный сигнал. -1 для автоматической настройки [По умолчанию: -1]

--statout Вывод статистики [Обязательно]

### 15.2 Параметры читателя данных

--optmapin Вх одной файл оптической карты [обязательно]

--optmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

--bnxsnr Значение фильтра BNХ SNR [по умолчанию: 3.0]

### 15.3 Параметры чтения справочной литературы

--refmapin Вх одной файл с справочной карты

--refmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

## 16 Дублированных Молекул Обнаружение

Обнаруживает дублированные молекулы в наборе данных оптической карты. Дублированные молекулы содержат одинаковое количество сегментов, а разницы между размерами каждого сегмента очень мала (обычно менее 100 пар оснований)

### 16.1 Параметры многопоточного заполнения

--meas Ошибки измерения. Обычно это намного меньше обычных ошибок измерения при обнаружении дублированных молекул [По умолчанию: 100]

--eag Допустимый диапазон ошибок [По умолчанию: 0,0]

--thread Количество потоков [По умолчанию: 1]

### 16.2 Параметры заполнения

--seedingmode Режим заполнения: 1: Оптимизировано для длинных k-меров (обычно для k больше 10); 2: Оптимизировано для коротких k-меров (обычно для k меньше или равно 10); -1: Автоматический выбор. [По умолчанию: -1]

--k Длина Кмера. [По умолчанию: 3]

--maxnosignal Максимальная область отсутствия сигнала между сигналами для заполнения. [По умолчанию: 10000000]

### 16.3 Параметры обнаружения дублированных молекул

--dupout Файлы, содержащие дублированные молекулы [Обязательно]

--minseg Минимальное количество сегментов, которые следует считать дублированными [По умолчанию: 15]

### 16.4 Параметры считывателя данных

--optmapin Вх одной файл оптической карты [обязательно]

--optmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS);

3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNX

Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF);

9: Формат OrGen XML; 10: Формат данных Валуева;

11: Формат карт Maligner; [По умолчанию: -1]

--bnxsnr Значение фильтра BNX SNR [по умолчанию: 3.0]



## 17 Дублированных Молекул Удаление

Удаление обнаруженных дублированных молекул из файла данных

### 17.1 Параметры удаления дублированных молекул

--dupin Файлы, содержащие дублированные молекулы [Обязательно]

### 17.2 Параметры считывателя данных

--optmapin Вх одной файл оптической карты [обязательно]

--optmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

--bnxsnr Значение фильтра BNХ SNR [по умолчанию: 3.0]

### 17.3 Параметры записи данных --optmapout

Выходной файл оптической карты [обязательно]

--optmapoutformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: Формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

## 18 FrequentKmerВыделение

Возвращает идентификаторы сегментов в километрах с числом совпадений, превышающим минимальное значение, по одному и тому же запросу.

### 18.1 Параметры считывателя данных

--optmapin Вх одной файл оптической карты [обязательно]

--optmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF)

(эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS);

3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ

Формат (BNХ); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF);

9: Формат OrGen XML; 10: Формат данных Валуева;

11: Формат карт Maligner; [По умолчанию: -1]

--bnxsnr Значение фильтра BNХ SNR [по умолчанию: 3.0]

### 18.2 Параметры записи и идентификатора сегмента

--siout Выходной файл идентификатора сегмента

### 18.3 Параметры заполнения

--seedingmode Режим заполнения: 1: Оптимизировано для длинных k-меров (обычно для k больше 10); 2: Оптимизировано для коротких k-меров (обычно для k меньше или равно 10); -1: Автоматический выбор. [По умолчанию: -1]

--k Длина Кмера. [По умолчанию: 3]

--maxnosignal Максимальная область отсутствия сигнала между сигналами для заполнения. [По умолчанию: 10000000]

## Часть 6

### Инструменты для выравнивания

#### 19 Результаты Инструменты

Предоставляет базовые функции для фильтрации и обработки результатов выравнивания

##### 19.1 Параметры считывателя результатов

--optresin Вх одной файл результатов выравнивания [Обязательно]

--optresinformat -1: Автоматически определяется по расширению файла; 0: Формат выравнивания OM (OMA); 1: Формат подробного выравнивания OM (OMD); 2: Формат XMAP (XMAP); 3: Формат Valouev et al.; 4: Формат уникального соответствия SOMA v2; 5: Формат Twin PSL; 6: Формат Maligner ALN; [По умолчанию: -1]

##### 19.2 Параметры записи результатов

--optresout Вывести файл результатов выравнивания

--optresoutformat Формат файла результата -1: Автоматически определяется по расширению файла; 0: Формат выравнивания OM (OMA); 1: Формат подробного выравнивания OM (OMD); 2: Формат XMAP (XMAP); 3: Формат Valouev et al.; 4: Формат уникального соответствия SOMA v2; 5: Формат Twin PSL; 6: Формат Maligner ALN; [По умолчанию: -1]

--writeunmap Записать отброшенные или неотображенные молекулы. [По умолчанию: true]

--multiple Записать несколько карт для молекулы. [По умолчанию: true]

--writeinfo Записать информацию о молекуле. [По умолчанию: true]

##### 19.3 Параметры чтения с справочной литературы

--refmapin Вх одной файл с справочной карты

--refmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валюева; 11: Формат карт Maligner; [По умолчанию: -1]

##### 19.4 Параметры считывателя данных

--optmapin Вх одной файл оптической карты

--optmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валюева; 11: Формат карт Maligner; [По умолчанию: -1]

--bnxsnr Значение фильтра BNХ SNR [по умолчанию: 3.0]

## 19.5 Параметры вывода данных

--mapout Вывод картированных молекул

--unmapout Вывод не отображенных молекул

--optmapoutformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: Формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNX Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OpGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

## 19.6 Параметры инструментов результатов

--qprefix Добавить префикс к имени запроса.

--rprefix Добавить префикс к имени ссылки.

--disinvalid Отменить недействительные результаты. [По умолчанию: true]

--conf Пересчет достоверности результата [По умолчанию: false]

--dataid Список идентификаторов данных, которые необходимо извлечь

--region Переход в формате chrN:start-end или chrN:start

--refnamemodify Изменить имя ссылки в соответствии с целевым файлом в формате: src\target

--dataremoval Удалить результат с именами запросов в файле

--joinresult Представить частичные выравнивания в одном выравнивании (пробел заполняется дополнительными и отсутствующими символами). Работает только с частичными выравниваниями с отношением indel [По умолчанию: false]

### 19.6.1 Опция прерывания результатов

--breakermode Режим 0: Отключить функцию прерывания; 1: Разбить выравнивание на сегменты запроса/ссылки, размер которых слишком сильно отклоняется, на несколько частичных выравниваний [По умолчанию: 0]

--meas Ошибка измерения [По умолчанию: 500]

--eag Допустимый диапазон ошибок (допустимая погрешность масштабирования) [По умолчанию: 0,1]

--match Оценка за сопоставление символа [По умолчанию: 5]

--fpp Штраф за дополнительный символ [По умолчанию: 2]

--fnp Штраф за отсутствие символа [По умолчанию: 2]

### 19.6.2 Параметры фильтра результатов

--filtermode Режим фильтра. 0: Без фильтра; 1: Фильтровать по всем следующим параметрам; 2: Фильтровать по минимуму только оценка [По умолчанию: 0]

--minmatch Минимальное количество совпадений частичного выравнивания [По умолчанию: 3]

--maxfp Максимальное количество дополнительных символов частичного выравнивания [По умолчанию: 10000]

--maxfn Максимальное количество пропущенных символов частичного выравнивания [По умолчанию: 10000]

--maxfpr Максимальная скорость дополнительных сигналов частичного выравнивания [По умолчанию: 1.0E-4] --maxfpr

Максимальная скорость пропущенных сигналов частичного выравнивания [По умолчанию: 0.5]

--minscore Минимальная оценка частичного выравнивания [По умолчанию: 0,0] --minsubfragratio

Минимальное отношение субфрагментов частичного выравнивания [По умолчанию: 0,0]

--minsigratio Минимальное отношение выровненных осигнала частичного выравнивания [По умолчанию: 0,0] --trimmode Режим подстройки. 0: Режим подстройки отключен; 1: Режим подстройки включен [По умолчанию: 0] --maxtrim Максимальное количество шагов подстройки частичного выравнивания [По умолчанию: 5]

--match Оценка с опосредованием сигнала [По умолчанию: 5] --fpr Штраф за дополнительный сигнал [По умолчанию: 2]

--fpr Штраф за отсутствие сигнала [По умолчанию: 2]

19.6.3 Параметры соединения выравнивания

--alignmentjoinmode Режим. 0: Без объединения. 1: Стандартное соединение indel. 2: Стандартное соединение indel-inv. 3: Стандартное транслокационное присоединение [По умолчанию: 0]

--closeref Максимальное расстояние (ссылка) между двумя частичными выравниваниями, которые необходимо объединить [По умолчанию: 250000]

--closefrag Максимальное расстояние (запрос) между двумя частичными выравниваниями, которые необходимо объединить [По умолчанию: 250000]

--minmatch Минимальное количество сигналов с ответствия, которые следует считать допустимым частичным с опосредованием. [По умолчанию: 3]

--maxtrim Максимальные шаг и обрезки для частичного выравнивания [По умолчанию: 5] --trimear Допустимая порешность масштабирования во время обрезки [По умолчанию: 0,1] --match Оценка с опосредующим сигналом [По умолчанию: 5]

--fpr Штраф за лишний сигнал [По умолчанию: 2] --fpr Штраф за отсутствие сигнала [По умолчанию: 2]

--indelp Штраф за объединение частичных выравниваний с отношением indel [По умолчанию: 10] --invp Штраф за объединение частичных выравниваний с отношением инверсии [По умолчанию: 30] --transp Штраф за объединение частичных выравниваний с отношением транслокации [По умолчанию: 50]

--localpenalty Включить штраф за локальное выравнивание для окончательного выравнивания (расматривается как глобальное выравнивание) [По умолчанию: ложно]

--minjoinscore Минимальная оценка объединенного окончательного выравнивания [По умолчанию: 0]

--minconf Минимальная достоверность (уникальность) окончательного выравнивания [По умолчанию: 0.0]

--minjoinedfragratio Минимальное отношение выровненной длины к длине запроса [По умолчанию: -1,0]

--minjoinedsigratio Минимальное отношение количества выровненных сигналов к общему количеству сигналов запроса [По умолчанию: -1,0]

--overlapalign Разрешить вывод перекрывающихся конечных выравниваний [По умолчанию: true] --maxalignitem

Максимальное количество выводимых конечных выравниваний. -1: Нет ограничений на количество конечных выравниваний [По умолчанию: 1]

#### 19.6.4 Варианты подъема

--liftoverin Вх одной файл liftOver, Ф ормат: х ромос omalт координата\т размер\r\n

#### 19.7 Варианты анализа результатов моделирования

--rocout Вывес ти таблиц у для пос троения ROC-кривой

## 20 Результат Объединение

Объединяет результаты выравнивания, полученные разными методами

20.1 Параметры считывателя результатов --optresin

Вх одной файл результатов выравнивания [Обязательно]

--optresinformat -1: Автоматически определяется по расширению файла; 0: Формат выравнивания OM (OMA); 1: Формат подробного выравнивания OM (OMD); 2: Формат XMAP (XMAP); 3: Формат Valouev et al.; 4: Формат уникального соответствия SOMA v2; 5: Формат Twin PSL; 6: Формат Maligner ALN; [По умолчанию: -1]

20.2 Параметры сливания результатов --resultkey

Ключи (имена) для представления файлов результатов [Обязательно]

--gapallowed Разрешены пробелы между результатами [По умолчанию: 0]

--analyzeall Анализировать только если запрос присутствует во всех результатах [По умолчанию: false]

--prefix Префикс выходного файла [Обязательно]

--outtype Тип выходного файла [По умолчанию: .omd]

## 21 РезультатСтатистика

Формирует статистику результатов выравнивания

### 21.1 Параметры чтения ссылок --refmapin Вх одной

файл карты ссылок [Обязательно]

--refmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

### 21.2 Параметры считывателя данных

--optmapin Вх одной файл оптической карты [обязательно]

--optmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

--bnxsnr Значение фильтра BNХ SNR [по умолчанию: 3.0]

### 21.3 Параметры считывателя результатов

--optresin Вх одной файл результатов выравнивания [Обязательно]

--optresinformat -1: Автоматически определяется по расширению файла; 0: Формат выравнивания OM (OMA); 1: Формат подробного выравнивания OM (OMD); 2: Формат XMAP (XMAP); 3: Формат Valouev et al.; 4: Формат уникального соответствия SOMA v2; 5: Формат Twin PSL; 6: Формат Maligner ALN; [По умолчанию: -1]

### 21.4 Параметры статистики результата--

checkstrand Проверка цепочки на корректность [По умолчанию: true]

--covout Вывод покрытия (в разработке)

--statout Вывод статистики



## 22 PrecisionRecallGraphDataGenerator

Генерирует таблицу данных для графиков точного отзыва. Этот модуль предполагает одно выравнивание (может содержать несколько частичных выравниваний) на один запрос. Вам необходимо использовать те же параметры модуля при соединении выравнивания, если файл выравнивания генерируется с помощью картографа OMTtools. Если вы используете другие инструменты выравнивания, установите alignmentjoinmode на 0.

### 22.1 Параметры чтения с правочной литературы

--refmapin Вх одной файл с правочной карты [обязательно]

--refmapinformat -1: Автоматически определяется по расширению файла; 0: Правочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNХ); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OpGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

### 22.2 Параметры считывателя данных

--optmapin Вх одной файл оптической карты [обязательно]

--optmapinformat -1: Автоматически определяется по расширению файла; 0: Правочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNХ); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OpGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

--bnxsnr Значение фильтра BNХ SNR [по умолчанию: 3.0]

### 22.3 Параметры считывателя результатов

--optresin Вх одной файл результатов выравнивания [Обязательно]

--optresinformat -1: Автоматически определяется по расширению файла; 0: Формат выравнивания OM (OMA); 1: Формат подробного выравнивания OM (OMD); 2: Формат XMAP (XMAP); 3: Формат Valouev et al.; 4: Формат уникального ответа SOMA v2; 5: Формат Twin PSL; 6: Формат Maligner ALN; [По умолчанию: -1]

### 22.4 Варианты соединения выравнивания

--alignmentjoinmode Режим. 0: Без объединения. 1: Стандартное соединение indel. 2: Стандартное соединение indel-inv. 3: Стандартное транслокационное присоединение [По умолчанию: 0]

--closeref Максимальное расстояние (ссылка) между двумя частичными выравниваниями, которые необходимо объединить [По умолчанию: 250000]

--closefrag Максимальное расстояние (запрос) между двумя частичными выравниваниями, которые необходимо объединить [По умолчанию: 250000]

--minmatch Минимальное количество сигналов ответа, которые следует считать допустимым частичным сопоставлением. [По умолчанию: 3]

--maxtrim Максимальные шаг и обрезки для частичного выравнивания [По умолчанию: 5]

--trimear Допустимая погрешность масштабирования во время обрезки [По умолчанию: 0,1]

--match Оценка за сопоставление с игнором [По умолчанию: 5]

--fpp Штраф за дополнительный с игнором [По умолчанию: 2]

--fnp Штраф за отсутствие с игнором [По умолчанию: 2]

--indelp Штраф за присоединение частичных выравниваний с отношением indel [По умолчанию: 10]

--invp Штраф за объединение частичных выравниваний с инверсионным отношением [По умолчанию: 30]

--transp Штраф за присоединение частичных выравниваний с транслокационной связью [По умолчанию: 50]

--localpenalty Включить штраф за локальное выравнивание для окончательного выравнивания (расматривается как глобальное выравнивание)  
[По умолчанию: ложно]

--minjoinscore Минимальная оценка объединенного окончательного выравнивания [По умолчанию: 30]

--minconf Минимальная достоверность (уникальность) окончательного выравнивания [По умолчанию: 0,4]

--minjoinedfratio Минимальное отношение выровненной длины к длине запроса [По умолчанию: -1,0]

--minjoinsratio Минимальное отношение количества выровненных с игнором к общему количеству с игнором запроса [По умолчанию: -1,0]

--overlapalign Разрешить вывод перекрывающихся конечных выравниваний [По умолчанию: true]

--maxalignitem Максимальное количество выводимых окончательных выравниваний. -1: Нет ограничений на количество окончательных выравниваний.  
выравнивания [По умолчанию: 1]

22.5 Параметры графика точности отзыва --prgout Вывод таблиц графика

точности отзыва [Обязательно]

--checkstrand Проверка цепочки на корректность [По умолчанию: true]

--sortstrat Сортировать по «оценке» или «уверенности» [По умолчанию: оценка]

## 23.3 запрос Обратный

Изменяет запрос в соответствии с результатами выравнивания так, чтобы они имели ту же цепочку, что и ссылка. В результатах выравнивания для каждого запроса допускается только одна ссылка и одна цепочка (иначе будет выдано исключение). Инверсии не поддерживаются.

### 23.1 Параметры считывателя результатов

--optresin Выходной файл результатов выравнивания [Обязательно]

--optresinformat -1: Автоматически определяется по расширению файла; 0: Формат выравнивания OM (OMA); 1: Формат подробного выравнивания OM (OMD); 2: Формат XMAP (XMAP); 3: Формат Valouev et al.; 4: Формат уникального ответа SOMA v2; 5: Формат Twin PSL; 6: Формат Maligner ALN; [По умолчанию: -1]

### 23.2 Параметры считывателя данных --

optmapin Выходной файл оптической карты [Обязательно]

--optmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OpGen XML; 10: Формат данных Валюева; 11: Формат карт Maligner; [По умолчанию: -1]

--bnxsnr Значение фильтра BNХ SNR [по умолчанию: 3.0]

### 23.3 Параметры записи данных

--optmapout Выходной файл оптической карты [обязательно]

--optmapoutformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: Формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OpGen XML; 10: Формат данных Валюева; 11: Формат карт Maligner; [По умолчанию: -1]

## 24 Выравнивание Выделение

Выделяет сегменты, используемые в выравнивании x в качестве идентификаторов сегментов.

### 24.1 Параметры считывателя данных --optmapin

Вх одной файл оптической карты [Обязательно]

--optmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNХ); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OrGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

--bnxsnr Значение фильтра BNХ SNR [по умолчанию: 3.0]

### 24.2 Параметры считывателя результатов

--optresin Вх одной файл результатов выравнивания [Обязательно]

--optresinformat -1: Автоматически определяется по расширению файла; 0: Формат выравнивания OM (OMA); 1: Формат подробного выравнивания OM (OMD); 2: Формат XMAP (XMAP); 3: Формат Valouev et al.; 4: Формат уникального ответа SOMA v2; 5: Формат Twin PSL; 6: Формат Maligner ALN; [По умолчанию: -1]

### 24.3 Параметры записи идентификатора сегмента

--siout Вых одной файл идентификатора сегмента

### 24.4 Параметры BED-писателя

--bedout Вых одной BED-файл

### 24.5 Параметры подветки выравнивания

--mode Выделить сегменты в - 1: Только ссылка; 2: Только запрос; 3: И запрос, и ссылка [По умолчанию: 3]

--overlap Требуется перекрывающийся запрос и ссылку (по крайней мере с одним игналом) [По умолчанию: false]

--strict Выделять сегменты только в последовательном совпадении с игналом (без дополнительных или отсутствующих игналов между ними) [По умолчанию: false]

## Часть VII

## Множественное выравнивание

## 25 Множественное выравнивание

Выполняет множественное выравнивание, используя несколько оптических карт в качестве запросов

## 25.1 Параметры считывателя данных

--optmapin Вх одной файл оптической карты [обязательно]

--optmapinform -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF)

(эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS);

3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNH

Формат (BNH); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9:

Формат OpGen XML; 10: Формат данных Валуева;

11: Формат карт Maligner; [По умолчанию: -1]

--bnxsnr Значение фильтра BNH SNR [по умолчанию: 3.0]

## 25.2 Несколько вариантов выравнивания

--maconfig Файл конфигурации или множественного выравнивания

--minlinksize Минимальный размер сегмента в процессе ссылавания [По умолчанию: 1000]

--maref Ссылки для множественного выравнивания

--maorderby 1: Первое появление блока; 2: классификация по сходству [По умолчанию: 1]

## 25.3 Параметры коллинеарного блок-писца

--cblout Вывод нескольких коллинеарных блоков выравнивания .

## 25.4 Коллинеарный блок-писатель порядка

--cboout Несколько коллинеарных блоков выравнивания упорядочивают вывод

## 25.5 Конфигурация многократного выравнивания

Весь процесс многократного выравнивания можно разбить на три основных подмодуля — с вызовом, построение и слияние/слияние.

Первый подмодуль «link» обрабатывает входные файлы выравнивания, классификации и многократного выравнивания для создания потенциальных «сегментных связей» между сегментами из разных запросов. Эти связи необходимы для стратегий группировки сегментов в коллинеарные блоки. Связи, созданные для этого подмодуля, должны быть достоверными, поскольку они определяют глобальную структуру многократного выравнивания.

Второй подмодуль «build» компилирует сегментные связи, созданные из предыдущего подмодуля, в коллинеарные блоки.

Третий подмодуль «merge» и «mergeproximity» объединяет похожие коллинеарные блоки в один коллинеарный блок, не нарушая глобальной структуры многократного выравнивания. Поскольку консервативная стратегия обычно используется для обмена чувствительности на более высокую специфичность на этапе «link», некоторые сегменты, принадлежащие одному и тому же коллинеарному блоку, могут быть помещены в несколько коллинеарных блоков. Объединение основано на (1) менее надежных связях («merge») или (2) длине проксимальных блоков («mergeproximity»).

Альтернативный подмодуль «mask» маркирует и предотвращает слияние сегментов, которые приводят к неспецифическому многократному выравниванию, указанных выше подмодулях. Этот модуль используется только в том случае, если запросы состоят из множества повторяющихся сигнальных шаблонов или серьезных сегментных дубликатов, которые нарушают процесс многократного выравнивания.

### 25.5.1 Файл конфигурации многократного выравнивания

В файле конфигурации каждая строка начинается с команды, соответствующей модулю в нескольких процедурах выравнивания, за которой следуют параметры и входные файлы.

mask (Необязательно) mask [Вх одной файл]

Запретить обработку сегментов, используемых в файле выравнивания, на этапах «с вызова» и «объединения».

ssylка с сылками [Вх одной файл] [k=? (классификация); maxnosignal=? (классификация)]

Создать сегментные ссылки из файлов выравнивания, классификации и многократного выравнивания. Для ввода файла классификации требуются два параметра k и maxnosignal.

rearrange (allowrearrangement=true)

Объедините эти сегментные связи в коллинеарные блоки.

mergeproximity (Необязательно) mergeproximity [meas=? (ошибка измерения); ear=? (ошибка масштабирования)]

Объединить соседние блоки, если их длины достаточно близки с учетом допустимого уровня ошибок.

mergeproximitysimple (Необязательно) mergeproximitysimple [meas=? (погрешность измерения); ear=? (масштабирование ошибка)]

Объединить блоки в близости, если их длины блоков достаточно близки с учетом допустимого уровня ошибок. Эта команда использует упрощенную версию mergeproximity для сокращения времени обработки и памяти, но объединяет только близкие блоки с одинаковым и уникальным родительским блоком.

merge (Необязательно) merge [Вх одной файл] [mergechain=true; k=? (классификация); maxnosignal=? (классификация)]

Объединить коллинеарные блоки с помощью сегментных связей. Входные файлы и параметры аналогичны тем, что в команде link.

Файл конфигурации должен содержать как минимум одну команду «link», за которой следует одна команда «build».

Команды «mask», «mergeproximity», «mergeproximitysimple» и «merge» являются необязательными.

## Пример файла конфигурации

```
маска SegmentMaskFile.si

ссылка AlignmentFile1.oma AlignmentFile2.oma

ссылка ClusteringFile1.kc k=7;maxnosignal=1000000

ссылка MultipleAlignmentFile1.cbl

сборка allowrearrangement=true

mergeproximitysimple изм=500;ухо=0.1

мера близости для ния = 500;ухо=0,1

объединить AlignmentFile3.oma mergechain=true
```

## 25.5.2 Стратегия оптимизации конфигурации

Поскольку оптимизация параметров во многом зависит от характера запросов, сложно иметь одну единственную конфигурацию, которая подходит для всех случаев. В этом разделе обобщена и представлена стратегия оптимизации конфигурации. Нетерпеливые пользователи могут пропустить подробности здесь и прочитать разделы ниже для шаблонов конфигурации для конкретных случаев.

Основные усилия по оптимизации направлены на то, какие файлы должны быть предоставлены на этапах «связать» и «слить», где пользователи должны решить, что предоставлять в качестве подсказок для ссылок сегментов. Другие этапы менее сложны и не требуют особой оптимизации.

Маска сегменты, которые приводят к неспецифическому выравниванию, но с высоким баллом, должны быть замаскированы, поскольку они приводят к неправильным связям сегментов для первоначального построения коллинеарного облака и нарушают структуру множественного выравнивания. Обычно эти сегменты происходят из повторяющихся сигнальных схем или сегментарных дубликатов. Пользователи могут найти эти сегменты путем сэмплирования каждого запроса.

## Маска

```
OMBlastMapper -refmapin query.data -optmapin query.data -alignmentjoinmode 0 -filtermode 1 -minconf 0 -allowdiffrefquery
false -exactmatch false -maxalignitem -1 -minjoinscore 50 -fpp 10 -fnp 10 AlignmentHighlight -mode 3 -optmapin query.data
-optresin
queryselfalign.oma -siout query-mask.si
```

Пользователи могут уменьшить параметры minjoinscore, fpp и fnp для менее строгого выравнивания, чтобы исключить больше потенциальных сегментов, вовлеченных в сегментарные дубликаты. Этот шаг не является необходимым, если сегментарные дубликаты отсутствуют в запросах.

Ссылка «link» является наиболее важной частью оптимизации. Ссылки сегментов должны быть построены с использованием высоконадежных доказательств, поскольку каждая ссылка сегмента определяет окончательную глобальную структуру множественного выравнивания. Пользователи должны быть консервативны и использовать строгие параметры для выравнивания. Основная цель оптимизации — связать как можно больше запросов, сохраняя при этом корректную структуру.

Оптимизация может быть упрощена как процесс выбора правильных файлов. Во-первых, пользователи могут импортировать файлы выравнивания от высокой до низкой достоверности. Для неоднозначной общей структуры множественного выравнивания пользователи должны импортировать меньше файлов выравнивания с низкой достоверностью. Для множественного выравнивания с большим количеством похожих шаблонов, не объединенных в одинаковые коллинеарные блоки, вместо этого пользователи должны импортировать больше файлов выравнивания с низкой достоверностью.

Файл множественного выравнивания также может быть использован для связывания. Обычно это происходит при множественном выравнивании крупных эукариотических геномов, где запросы были точно выровнены в скользящих окнах.

Обратите внимание, что во время оптимизации шаг «link» пользователи могут отключить команды «merge» и «mergeproximity».

Пользователям не нужно оптимизировать этап «сборки».

**merge** На основе сегментных связей шаг «merge» объединяет два коллинеарных блока только в том случае, если сливание не нарушает глобальную структуру множественного выравнивания (т.е. не создает перегруппировку). Процесс оптимизации «merge» похож на процесс «link», за исключением того, что файлы для сегментных связей могут иметь более низкую достоверность в целом. Этот шаг настоятельно НЕ рекомендуется использовать в регионе с вариациями числа копий. Обратите внимание, что этот шаг требует очень много времени и памяти и неприменим к эукариотическим геномам.

**mergeproximity** Пользователям не обязательно учитывать только один фактор при оптимизации шага «mergeproximity» — толерантность к ошибкам (ошибка измерения и ошибка масштабирования). По сути, более высокая толерантность к ошибкам должна применяться при множественном выравнивании необработанных молекул оптического картирования. Напротив, более низкая толерантность к ошибкам может использоваться при множественном выравнивании сбранных контигов оптического картирования или последовательностей, переваренных *in-silico*. Пользователи найдут шаг «mergeproximity» полезным практически во всех обстоятельствах. Шаг требует много времени и памяти, упрощенная, но менее мощная версия «mergeproximitysimple» может использоваться в более крупных эукариотических геномах.

«mergeproximitysimple» — это упрощенная версия «mergeproximity».

Он ограничивает объединение соседних блоков только в том случае, если у них есть один родительский блок.



### 25.5.3 Шаблон конфигурации

## Геномы прокариот

## Выравнивание

```

OMBlastMapper -refmapin query.data -optmapin query.data -optresout query1.oma -filtermode 1 -alignmentjoinmode
0 -maxalignitem -1 -minconf 0 -thread 2 - allowequalrefquery false -exactmatch false -writeinfo false -writeunmap false
-fpp 15 -fnp 15 -ear 0.05 -meas 500 -minjoinscore 50 OMBlastMapper -refmapin query.data -optmapin query.data -
optresout query2.oma -filtermode 1 -alignmentjoinmode 0 -
maxalignitem -1 -minconf 0 -thread 2 - allowequalrefquery false -exactmatch false -writeinfo false -writeunmap false -
fpp 10 -fnp 10 -ear 0.05 -meas 500 -minjoinscore 50 OMBlastMapper -refmapin query.data -optmapin query.data -
optresout query3.oma -filtermode 1 -alignmentjoinmode 0 -maxalignitem -1 -minconf 0 -thread 2 - allowequalrefquery
false -exactmatch false -writeinfo false -writeunmap false -fpp
5 -fnp 5 -ear 0.05 -meas 500 -minjoinscore 50 OMBlastMapper -refmapin query.data -optmapin query.data -optresout
query4.oma -filtermode 1 -alignmentjoinmode 0 -maxalignitem -1 -minconf 0 -thread 2 - allowequalrefquery false -
exactmatch false -writeinfo false -writeunmap false -fpp 50 -fnp 50 -ear 0.05 -meas 500 -minjoinscore 30 OMBlastMapper
-refmapin query.data -optmapin query.data -optresout
query5.oma -filtermode 1 -alignmentjoinmode 0 -maxalignitem -1 -minconf 0 -thread 2 -allowequalrefquery false -
exactmatch false -writeinfo false -writeunmap false -fpp 50 -fnp 50 -ear 0.05 -meas 500 -minjoinscore 20 OMBlastMapper
-refmapin query.data -optmapin query.data -optresout query6.oma -filtermode 1 -alignmentjoinmode 0 -maxalignitem
-1 -minconf 0 -thread 2 -allowequalrefquery false -exactmatch
false -writeinfo false -writeunmap false -fpp 2 -fnp 2 -ear 0.05 -meas 500 -minjoinscore 50

```

## Файл конфигурации множественного выравнивания

```

ссылка запрос 1.oma
ссылка запрос 2.oma
ссылка запрос 3.oma
сборка allowrearrangement=true
mergeproximity изм.=500;yx o=0,05 объединение
запрос 1.oma объединение
запрос 2.oma объединение
запрос 3.oma объединение
запрос 4.oma объединение
запрос 5.oma объединение
запрос 6.oma
mergeproximity изм.=500;yx o=0,05

```

## Варианты номеров копий

## Выравнивание

```

OMBlastMapper -refmapin query.data -optmapin query.data -optresout query1.oma -filtermode 1 -alignmentjoinmode
0 -maxalignitem -1 -minconf 0 -thread 2 - allowequalrefquery false -exactmatch false -writeinfo false -writeunmap false
-fpp 10 -fnp 10 -ear 0.05 -meas 500 -minjoinscore 100 OMBlastMapper -refmapin query.data -optmapin query.data -
optresout query1.oma -filtermode 1 -alignmentjoinmode 0 -
maxalignitem -1 -minconf 0 -thread 2 - allowequalrefquery false -exactmatch false -writeinfo false -writeunmap false -
fpp 5 -fnp 5 -ear 0.05 -meas 500 -minjoinscore 50 OMBlastMapper -refmapin query.data -optmapin query.data -optresout
query1.oma -filtermode 1 -alignmentjoinmode 0 -maxalignitem -1 -minconf 0 -thread 2 -allowequalrefquery false -
exactmatch false -writeinfo false -writeunmap false -fpp 50 -
fnp 50 -ear 0.05 -meas 500 -minjoinscore 30

```

## Конфигурация множественного выравнивания

```

ссылка query1.oma
ссылка query2.oma
ссылка query3.oma
сброска allowrearrangement=true
mergeproximity meas=500;ear=0.05

```

## Часть VIII

## Несколько инструментов выравнивания

### 26 CBLTools

Представляет базовые функции для фильтрации и обработки результатов множественного выравнивания.

#### 26.1 Параметры считывателя данных

--optmapin Вх одной файл оптической карты

--optmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OpGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

--bnxsnr Значение фильтра BNХ SNR [по умолчанию: 3.0]

#### 26.2 Параметры коллинеарного считывателя блоков

--cblin Ввод нескольких коллинеарных блоков выравнивания.

#### 26.3 Параметры коллинеарного блок-писца

--cblout Вывод нескольких коллинеарных блоков выравнивания.

#### 26.4 Параметры инструментов CBL

--cboout Вывод коллинеарного порядка блоков

##### 26.4.1 Параметры фильтрации

--allblock Извлечь запросы, содержащие все блоки в этом списке

--reservequery Резервировать запросы, даже если они не соответствуют другим критериям

##### 26.4.2 Параметры сортировки

--sortblock Сортировать запросы по наличию блока

--sortdesflanklen Сортировать запросы по длине между двумя фланговыми блоками (FB1a FB1b) по убыванию.  
порядок

## 27 Анализ производительности множественного выравнивания

Анализирует эффективность множественного выравнивания. ОМ на основе множественного выравнивания последовательностей.

### 27.1 Параметры считывателя данных --

optmapin Вх одной файл оптической карты [Обязательно]

--optmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF) (эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS); 3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9: Формат OpGen XML; 10: Формат данных Валуева; 11: Формат карт Maligner; [По умолчанию: -1]

--bnxsnr Значение фильтра BNХ SNR [по умолчанию: 3.0]

### 27.2 Считыватель MAF

--mafin Ввод формата множественного выравнивания .

### 27.3 Параметры коллинеарного считывателя блоков

--cblin Ввод нескольких коллинеарных блоков выравнивания .

### 27.4 множественные варианты выравнивания

--statout Вывод статистики

## 28 Блок Соединение Граф Генерация

Создает файл точечной диаграммы, отображающий последовательность коллинеарных блоков.

### 28.1 Параметры коллинеарного считывателя блоков

--cblin Ввод нескольких коллинеарных блоков выравнивания.

### 28.2 Варианты генерации графа соединений блоков

--maref Ссылки для множественного выравнивания

--reverse Изменить направление прогрессии коллинеарных блоков на противоположное [По умолчанию: false]

--minedgeweight Фильтр с минимальным весом ребра [По умолчанию: 0]

--displayall Вывести все блоки (Установите false, чтобы программа не выводила блоки без какой-либо связи с другими блоками) [По умолчанию: true]

--dotout Вывод файла с точкой [обязательно]

## Часть 9

# Филогеноетика

## 29 UPGMATreeConstruction

Реконструирует филогеноетическое дерево на основе результатов множественного выравнивания с использованием подхода UPGMA

### 29.1 Варианты построения UPGMATree

```
--matrixout Вывести матрицу расстояний [обязательно]
--treeout Выводить деревья в формате newick [Обязательно]
--startblock Стартовый блок (Экспериментальные параметры)
--stopblock Стоп-блок (Экспериментальные параметры)
```

### 29.2 Параметры считывателя данных

```
--optmapin Вх одной файл оптической карты [обязательно]
--optmapinformat -1: Автоматически определяться по расширению файла; 0: Справочный стандартный формат (REF)
(эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS);
3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNХ
Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата
REF); 9: Формат OrGen XML; 10: Формат данных Валуева;
11: Формат карт Maligner; [По умолчанию: -1]
--bnxsnr Значение фильтра BNХ SNR [по умолчанию: 3.0]
```

### 29.3 Параметры коллинеарного считывателя блоков

```
--cblin Ввод нескольких коллинеарных блоков выравнивания .
```

## Часть X

# Визуализация

## 30 OMView

Визуализирует данные оптического картирования. OMView предоставляет графический интерфейс для визуализации данных оптического картирования для различных целей.

### 30.1 Загрузка данных

--viewrefin Загрузить сылки

--viewmapin Загрузить молекулы

--viewresin Загрузить результаты выравнивания

--viewcblin Загрузить коллинеарные блоки

--viewcboin Загрузить коллинеарные блоки (порядок)

--viewcbcin Загрузить коллинеарные блоки (цвет)

--viewannoин Загрузить аннотации

--viewin Автоматический ввод файлов

### 30.2 Просмотр открытия

--viewregion Показать определенный регион в региональном представлении

--viewanchor Показать определенный якорь в представлении якорей

--viewanchoregion Укажите регион для якоря в представлении якоря.

--viewalignment Показать определенное выравнивание

--viewma Автоматически открывать представление с несколькими выравниваниями [По умолчанию: false]

--viewtab Автоматически открывать представление блока множественного выравнивания [По умолчанию: false]

--viewmolecule Автоматически открывать вид молекулы [По умолчанию: false]

--viewsave Сохранять представления в определенном месте вместе запуском OMView

--viewsaveformat Форматы сохранения изображений. [svg; png; jpg; ] [По умолчанию: png]

### 30.3 Настройки просмотра--dnapratio

Соотношение ДНК по умолчанию [По умолчанию: 400,0]

--zoom Уровень масштабирования по умолчанию [По умолчанию: 1.0]

--viewbreakresult Включить прерыватель результатов [по умолчанию: false]

--viewunmap Включить не отображенную часть [по умолчанию: false]

--viewsettingin Вх одной файл настроек OMView

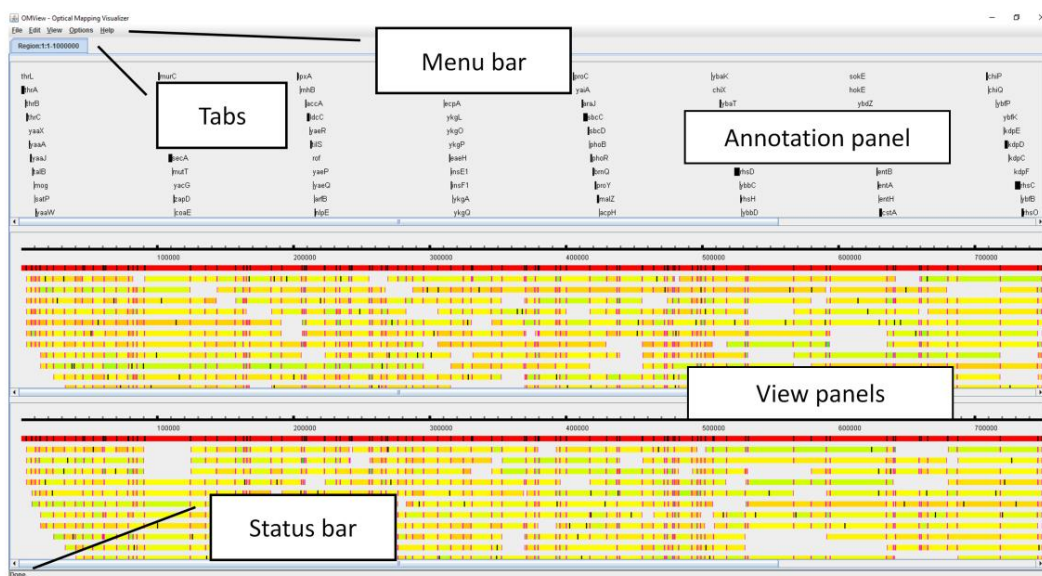


## 30.4 Помощь

--help Отобразить меню с правки

## 30.5 Проц едур ы визуализац ии

### 30.5.1 Макет OMView

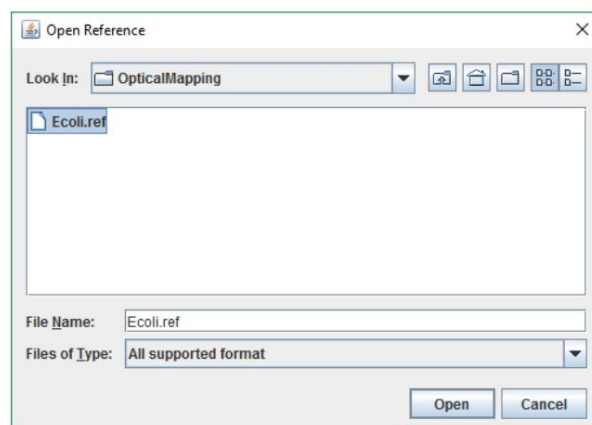


### 30.5.2 Заг рузка необх одимых данных

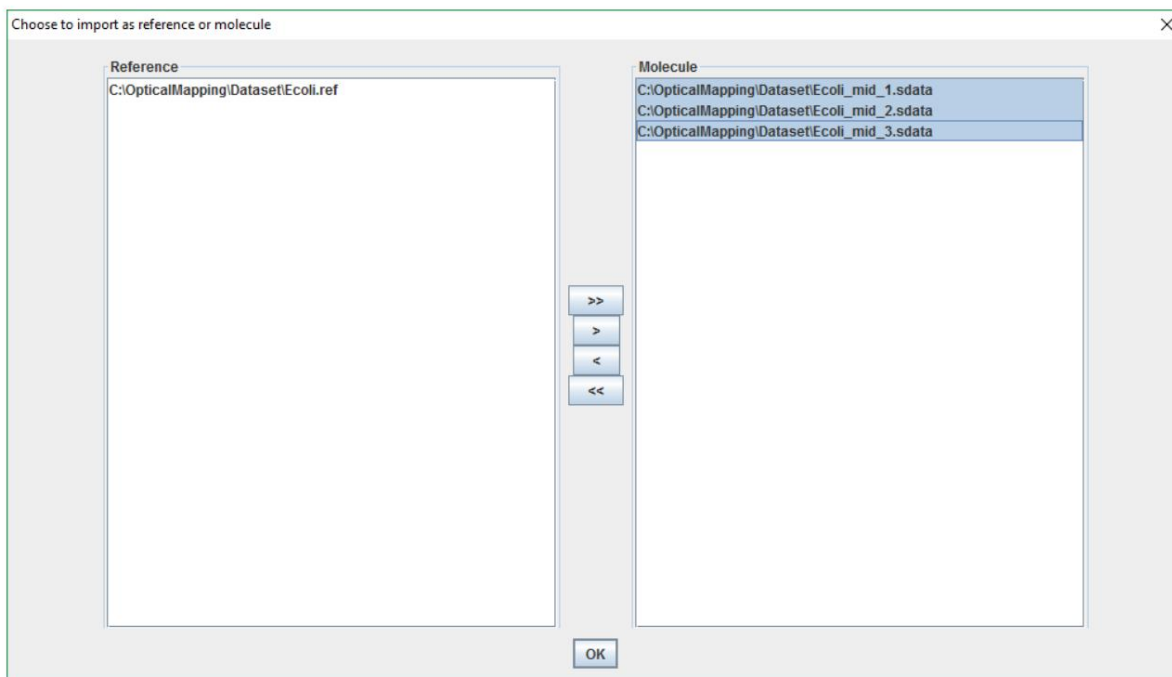
Сущес твует два способа загрузки данных в OMView: (1) загрузить наборы данных из меню и (2) перетаскивать наборы данных в программу.

Примечание: пользователи должны сначала загрузить файлы с сылками и молекулами, прежде чем загружать другие файлы (выравнивание, файлы аннотаций и множественных выравниваний).

Выберите файлы данных из меню. После выбора файла данных для загрузки (FileLoad) выберите целевой файл и нажмите «Открыть».



Перетаскивание наборов данных. В программу можно одновременно перетаскивать несколько файлов.



Поскольку форматы файлов ссылок и молекул одинаковы, пользователю не обязательно указывать, файлы загружаются как ссылки или молекулы.

Зависимость файла

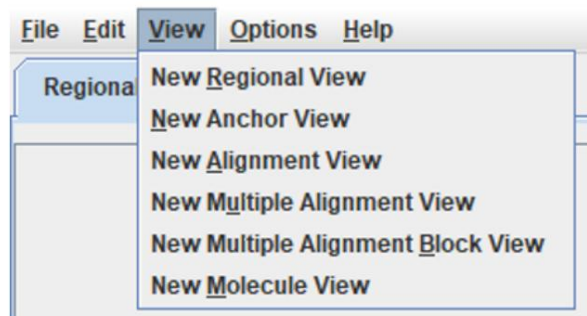
Ссылка и молекулы: Нет требований к другим файлам данных. Обратите внимание, что идентификатор ссылки не должен дублироваться среди всех файлов ссылок. Аналогично, идентификатор молекулы не должен дублироваться.

Выравнивания: требуются ссылки и молекулы (если файл выравнивания содержит информацию о молекулах, загрузка файла молекул не требуется)

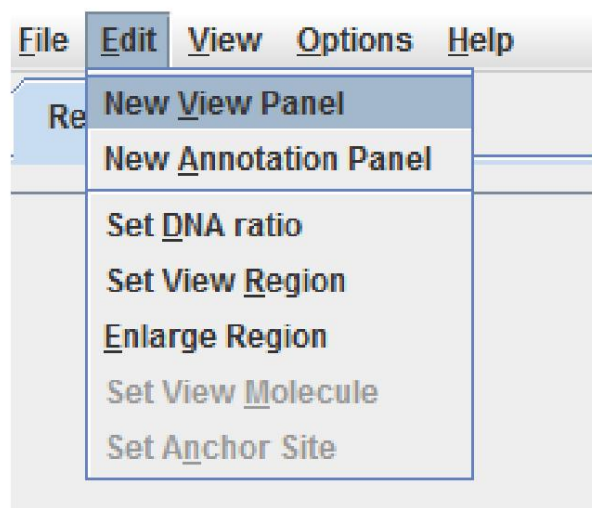
Аннотации: Требуется ссылка

Множественные выравнивания: требуются молекулы

### 30.5.3 Запуск представления



Пользователи могут открывать новые вкладки просмотра в меню Вид. По умолчанию OMView инициализируется с новым пустым региональным видом.

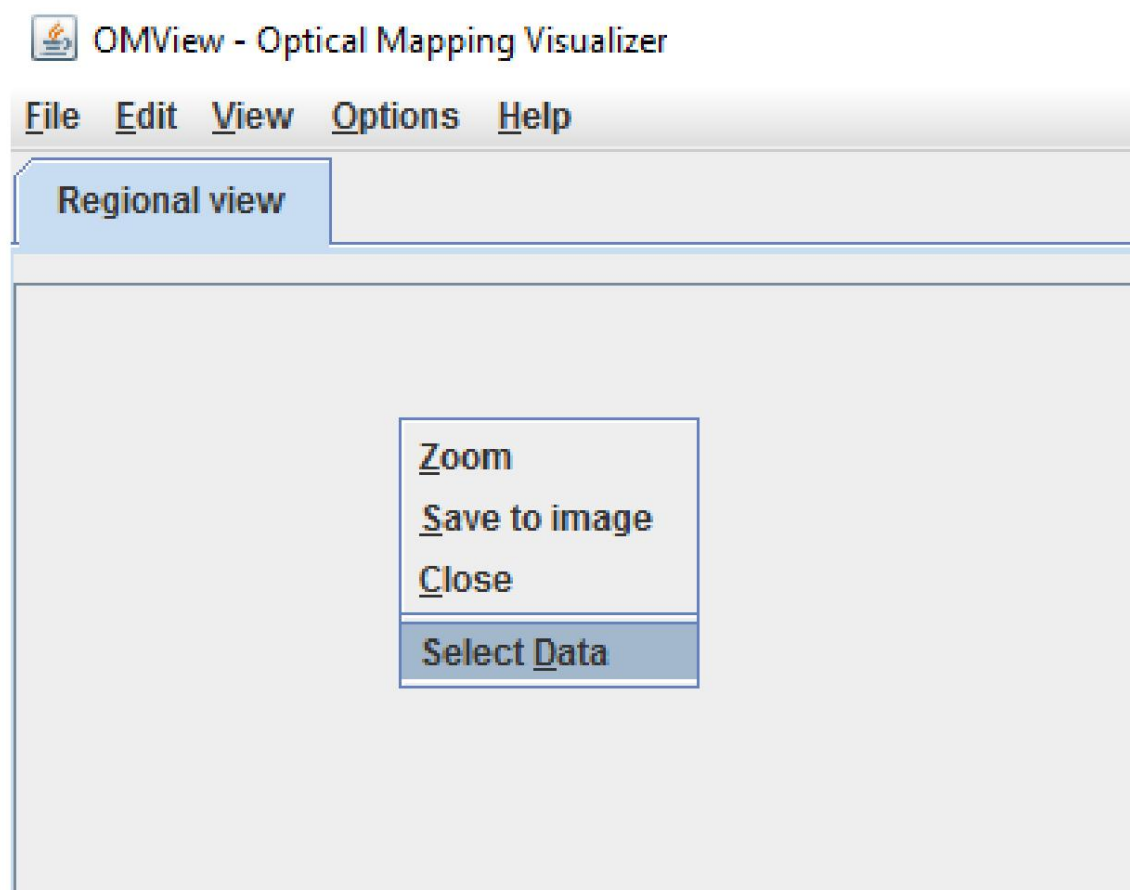


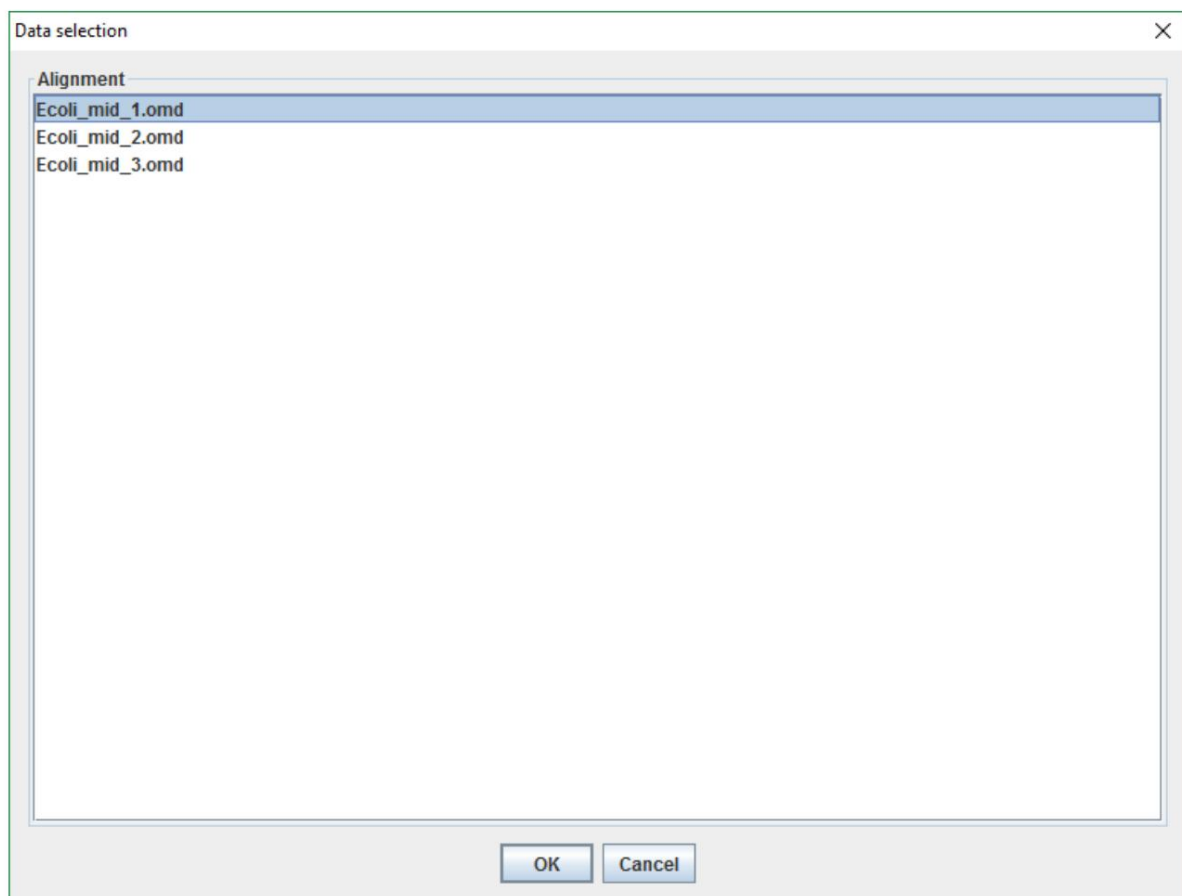
Каждая вкладка по умолчанию содержит панель просмотра. Некоторые представления поддерживают визуализацию более чем одной панели просмотра (региональный вид, видя коря и вид молекулы) и панелей аннотаций (региональный вид и видя коря) на одной вкладке. Пользователи могут вставлять новые панели в меню «Изменить».

#### 30.5.4 Выбор данных на панели просмотра

В контекстном меню каждой панели просмотра выберите «Выбрать данные», чтобы выбрать данные для отображения.

Откроется диалоговое окно, в котором пользователи могут выбрать данные. Обратите внимание, что если результаты выравнивания более одного набора загружены на одной панели просмотра, убедитесь, что их соответствующие идентификаторы молекул различны.





## 30.5.5 Процент едур по конкретным видам

Региональный вид Региональный вид отображает трассы в виде обзор в выбранном регионе.



После открытия вкладки регионального представления установите целевой регион как X:NNNNNN-NNNNNN (EditSet (Просмотр региона). Должна появиться ссылка (красный прямоугольник) с сигналами (черные вертикальные полосы).

Примечание: Имя ссылки должно соответствовать имени в наборе данных. Распространенная ошибка в

настройка региона портит имя ссылки «chr1» или «1».

После выбора выравниваний и аннотаций результаты будут отображены на панели просмотра.

Выровненные части молекул показаны в цветовом спектре (от зеленого до красного в зависимости от коэффициента масштабирования от 0,5 до 1,5, где желтый цвет означает коэффициент масштабирования 1) с розовыми и черными сигналами, указывающими на отображенные или не отображенные сигналы.

Наведите курсор на сигнал, чтобы отобразить его информацию. Чтобы отобразить детали выравнивания определенной молекулы, просто щелкните на целевой молекуле, и будет создана новая вкладка представления выравнивания. Чтобы просмотреть все молекулы, выровненные с определенным опорным сигналом, щелкните на этом сигнале, чтобы создать новое представление привязки.

Пример: (Ссылка Ecoli.ref, Ecoli mid 1.omd, Ecoli mid 2.omd, Ecoli mid 3.omd, Ecoli.gff

Представление «Якорь» Представление «Якорь» отображает выравнивания, соответствующие выбранным сигналам, для проверки структурных изменений.

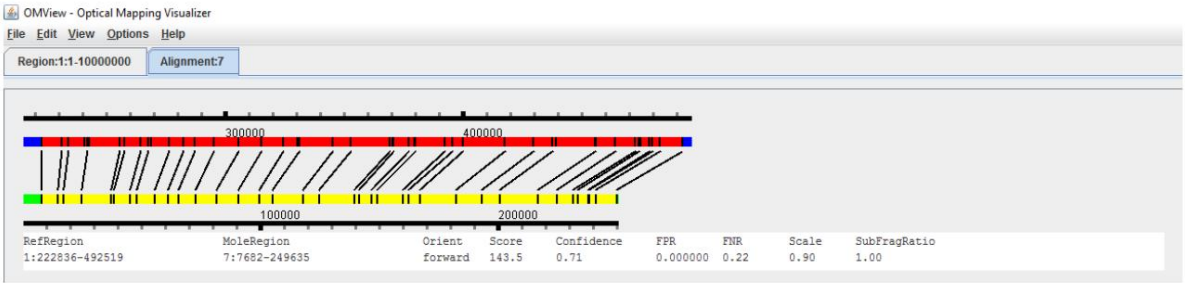


Процедура настраивается аналогично процедуре в региональном представлении. После открытия вкладки представления якоря установите якорный сайт как X:NNNNNN-NNNNNN (EditSet Anchor Site). Обратите внимание, что якорный сайт должен представлять положение одного или двух сигналов. Пользователи могут установить регион как X:NNNNNN-NNNNNN (EditSet Region). По умолчанию регион установлен на 200 кбит/с от якорных сайтов.

Пример: (Ссылка) Ecoli.ref, Ecoli mid 1.omd, Ecoli mid 2.omd, Ecoli mid 3.omd, Ecoli.gff — —



Вид выравнивания Вид выравнивания отображает детали выравнивания отдельной молекулы.



В представлении выравнивания установите идентификатор молекулы, чтобы просмотреть детали выравнивания (EditSet View Molecule). Пользователи получают предупреждение, если для выбранной молекулы выравнивание не существует.

Верхний и нижний прямоугольники представляют ссылки на молекулу соответственно. Два набора линеек указывают координаты относительно начала эталонной хромосомы и начала оптической молекулы. Обратите внимание, что молекула масштабируется и ориентируется в прямой ориентации, в то время как ссылка может быть масштабирована и перевернута для лучшей визуализации. Пользователи могут посмотреть на текстовое поле ниже, чтобы получить дополнительные сведения о выравнивании.

Пример: (Ссылка) Ecoli.ref, Ecoli mid 1.omd      -      -

Предс тавление множественног о выравнивания Предс тавление множественног о выравнивания отображает множес твенные выравнивания вс ех запрос ов для г еномног о с равнения .



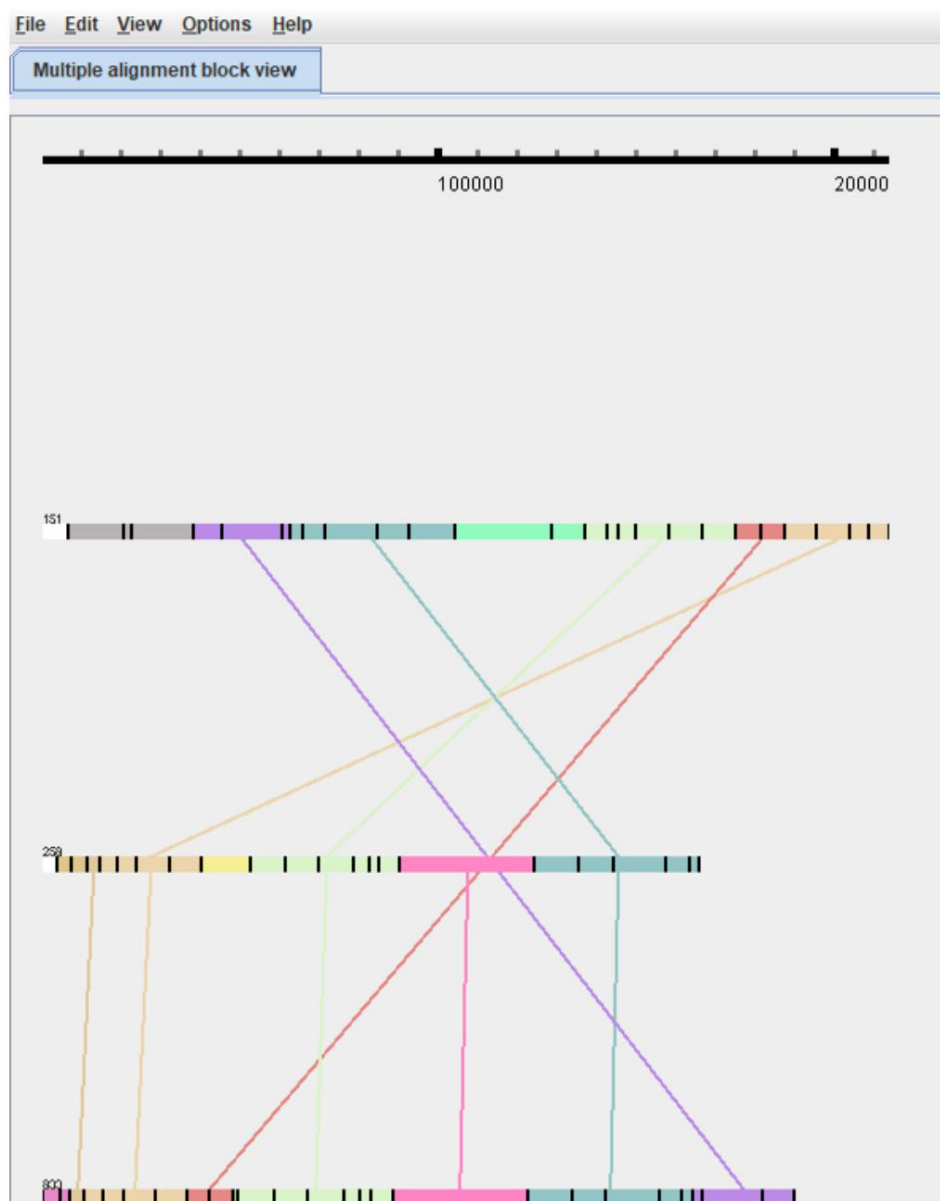
Вид множественног о выравнивания отображается автоматичес ки после ус тановки данных . Если файл ц вета (.cbs) не при этом коллинеарным блокам будет назначен с лучайный ц вет.



Щёлкните правой кнопкой мыши, чтобы с ортировать и управля ть множественными выравнивания ми. Пользователи мог ут с ох ранить текуще множественное выравнивание после манипуля ц ии. Также дос тупна функц ия подс чета, а статис тика выводится в конс оли.

Пример: (Молекула) Ecoli MA.sdata, Ecoli MA.cbl, Ecoli MA.cbo, Ecoli MA.cbc-

Вид блока множественного выравнивания. Вид блока множественного выравнивания отображает альтернативную визуализацию множественного выравнивания. Запросы остаются в исходном виде без изменений. Сегменты из одного коллинеарного блока соединены линиями одного цвета.



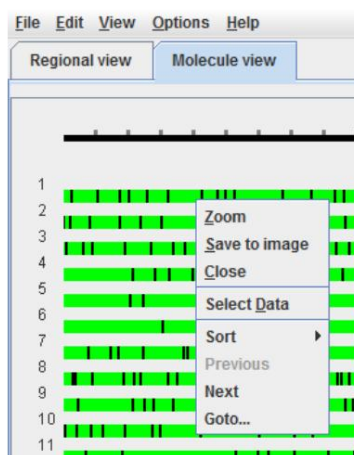
Вид блока множественного выравнивания отображается автоматически после установки данных. Если файл цвета(.cbc) не указан, коллинеарным блокам будет назначен случайный цвет.

Пример: (Молекула) Ecoli MA.sdata, Ecoli MA.cbl, Ecoli MA.cbo, Ecoli MA.cbc

–



Вид молекул В виде молекул панель отображает страницу, содержащую 100 молекул. Перейдите на другую страницу с помощью пунктов меню правой кнопки мыши Предыдущий, Следующий и Перейти. Также доступна функция сортировки для сортировки молекул по размеру молекулы, количеству сигналов в молекуле или имени молекулы.



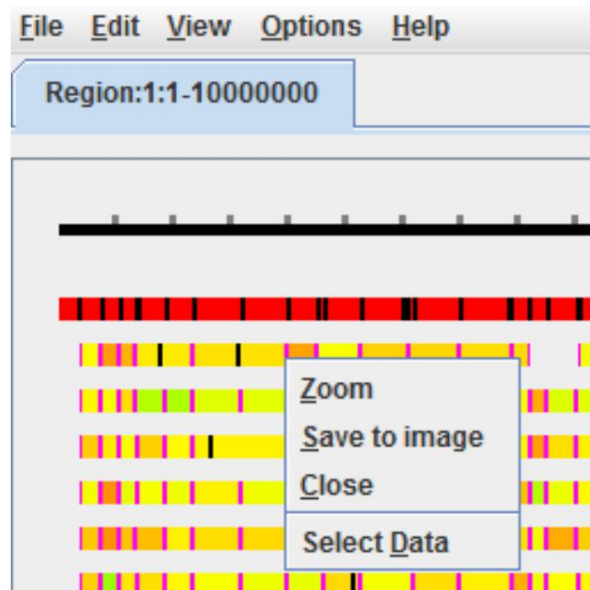
Щелкните правой кнопкой мыши, чтобы отсортировать молекулы.

Пример: (Молекула) Ecoli mid 1.sdata -

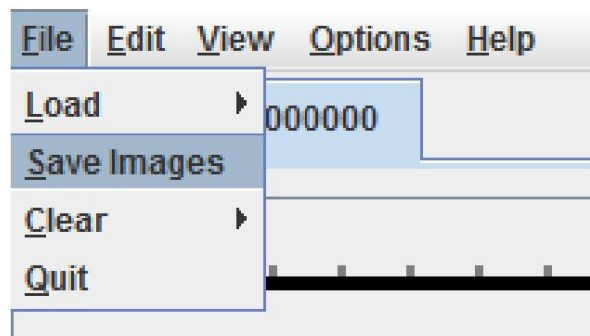
### 30.5.6 Экспорт изображения

Изображения можно экспортировать на отдельных панелях или на нескольких панелях в форматах SVG, PNG и JPG.

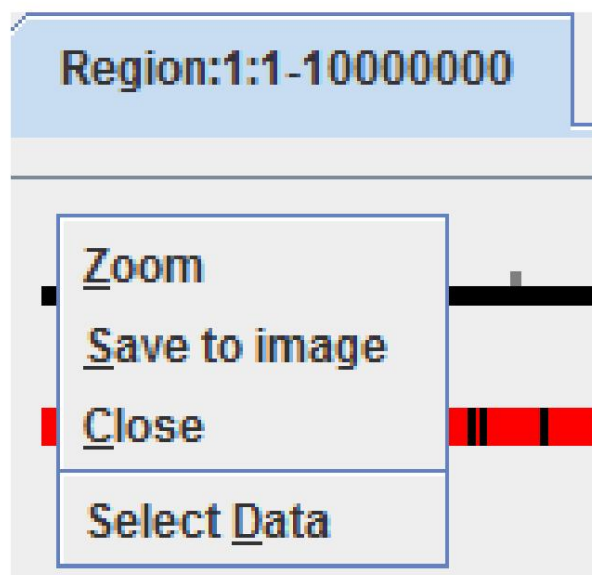
Отдельная панель В контекстном меню пользователи могут выбрать «Сохранить как изображение», чтобы сохранить текущую панель просмотра.



Все панели на одной вкладке Пользователи могут выбрать меню 'Файл' 'Сохранить изображения', чтобы сохранить все панели просмотра на текущей вкладке в одно изображение.



### 30.5.7 Закрытие вкладок и панелей



Чтобы закрыть вкладку, щелкните по ней с средней кнопкой мыши. Чтобы закрыть панель просмотра, щелкните правой кнопкой мыши панель и выберите «Закрыть».

## 30.6 Часто задаваемые вопросы (FAQ)

В: Существует ли пакетный режим генерации изображений?

А: Да. Вы можете использовать опцию `--viewsave` для пакетной генерации изображений.

В: Можно ли выбрать несколько элементов одновременно для импорта справочных или выборочных данных?

А: Да. Удерживайте клавишу Control при выборе нескольких элементов.

В: Я загрузил все файлы, включая ссылки, молекулы и результаты выравнивания, но ничего не вижу.  
В чем дело?

А: Чтобы визуализировать ссылки и результаты выравнивания, (1) необходимо выбрать данные на панели просмотра (Выбрать данные в контекстном меню); и (2) необходимо задать область просмотра с помощью EditSet View Region.

В: Я установил область просмотра или якорный сайт для визуализации своих данных, но получаю ошибку «Ссылка не найдена».  
Что не так? А: Наиболее распространенной причиной ошибки является неправильный ввод имени ссылки. Обратите внимание на разницу между «chr1», «Chr1», «CHR1» и «1».

В: Почему скорость загрузки файла очень низкая после загрузки половины моих данных?

А: Убедитесь, что у вас достаточно памяти для хранения всех данных. Используйте параметр `Xmx`, чтобы выделить больше памяти для Java-машины. Не загружайте слишком много данных в экземпляр OMView, если на вашей машине недостаточно памяти.

В: Скорость загрузки регионально вида низкая.

А: Попробуйте ограничить диапазон региона. Обычно региону больше 1 Мбит/с требуется некоторое время, чтобы полностью загрузиться.

В: Какие форматы принимает OMView?

А: Форматы файлов ссылок и молекул: REF, FA01, SPOTS, DATA, SDATA, BNX, CMAP, OPT, SILICO и OpGen XML. Форматы результатов выравнивания: OMA, OMD, XMAP, Valouev et al., SOMA v2 Unique Match и Twin PSL. Форматы аннотаций: BED, GVF и OSV, AGP.

## Часть XI

## Другие сценарии

### 31 TWINResultRepeatRemover

Удаляет повторные результаты выравнивания из TWIN.

#### 31.1 Параметры считывателя результатов

--optresin Вх одной файл результатов выравнивания [Обязательно]

--optresinformat -1: Автоматически определяется по расширению файла; 0: Формат выравнивания OM (OMA); 1: Формат подробного выравнивания OM (OMD); 2: Формат XMAP (XMAP); 3: Формат Valouev et al.; 4: Формат уникального ответа SOMA v2; 5: Формат Twin PSL; 6: Формат Maligner ALN; [По умолчанию: -1]

#### 31.2 Параметры записи результатов

--optresout Вывести файл результатов выравнивания [обязательно]

--optresoutformat Формат файла результата -1: Автоматически определяется по расширению файла; 0: Формат выравнивания OM (OMA); 1: Формат подробного выравнивания OM (OMD); 2: Формат XMAP (XMAP); 3: Формат Valouev et al.; 4: Формат уникального ответа SOMA v2; 5: Формат Twin PSL; 6: Формат Maligner ALN; [По умолчанию: -1]

--writeunmap Записать отброшенные или неотображенные молекулы. [По умолчанию: true]

--multiple Записать несколько карт для молекулы. [По умолчанию: true]

--writeinfo Записать информацию о молекуле. [По умолчанию: true]



## 32 Отдельный BNXScan

Разделяет файл bnx на несколько файлов bnx в соответствии с глобальным номером сканирования.

32.1 Параметры считывателя данных --optmapin

Вх одной файл оптической карты [Обязательно]

--optmapinformat -1: Автоматически определяется по расширению файла; 0: Справочный стандартный формат (REF)

(эквивалент формата SILICO); 1: формат fasta-01 (FA01); 2: формат файла Spots (SPOTS);

3: Стандартный формат молекулы (DATA); 4: Формат моделирования молекулы (SDATA); 5: Файл BNX

Формат (BNX); 6: Формат файла CMAP (CMAP); 7: Формат SOMA opt (OPT); 8: Формат SOMA silico (SILICO) (эквивалент формата REF); 9:

Формат OrGen XML; 10: Формат данных Валуева;

11: Формат карт Maligner; [По умолчанию: -1]

--bnxsnr Значение фильтра BNX SNR [по умолчанию: 3.0]

### 32.2 Отдельные параметры BNXScan

--prefix Префикс вывода [По умолчанию: сканирование]