

### Домашнее задание №11-13 - Правила

Для практикумов №№11-13 предлагается одно домашнее задание, которое для удобства заполнения ведомости формально разбито на 3 серии. Содержание серии не всегда соответствует лекционному материалу занятия.

Обратите внимание, что вы работаете с реальными данными, которые могут быть не идеальны, что-то может пойти не так, но в этом нет ничего ужасного.

Пожалуйста, не стесняйтесь задавать вопросы, уточнять формулировки и указывать на ошибки в задании при обнаружении таковых.

В рамках четырех занятий мы решим две задачи:

- поиск и аннотация вариантов одного человека по данным экзомного секвенирования на примере одной хромосомы (задания к практикумам №№ 11-12)
- составление экспрессионного профиля одного человека по данным секвенирования РНК на примере одной хромосомы (задания к практикуму №13)

Общий **дедлайн** полного домашнего задания:

мягкий дедлайн - 9 декабря 23:59

жесткий дедлайн - 15 декабря 23:59

Коллоквиум - 16 декабря

### **Внимание!**

На коллоквиуме один из обязательных вопросов - ответить на вопросы по скрипту к ДЗ 11-12. Этот вопрос можно сдавать только Анастасии Жариковой.

## Инструкция

Работа на kodoמו

Зайдите на kodoמו через командную строку

В папке **/mnt/scratch/NGS** создайте директорию, которая будет называться, как ваш логин на kodoמו. Используйте команду `mkdir`.

**!!!Далее работайте только в этой директории!!!**

=====

В ведомости у каждого указан номер хромосомы. См. Практикум 11

Это будет вашим референсным геномом в ходе всего задания.

=====

## Данные

Все данные лежат в папке **/mnt/scratch/NGS**

Не работайте в директории с данными, копируйте необходимые файлы в **СВОИ** директории.

### Последовательность генома человека

Референсная последовательность генома человека версии GRCh38.p14 (GCA\_000001405.29) получен из ensembl

Необходимые файлы лежат тут:

**/mnt/scratch/NGS/DATA/hg38**

Далее файлы типа: Homo\_sapiens.GRCh38.dna.chromosome.N.fa, где N - номер хромосомы

В каждом файле представлена последовательность одной хромосомы генома человека.

### Чтения

**/mnt/scratch/NGS/DATA/dna\_reads** - парноконцевые чтения секвенирования экзона человека

Далее файлы типа: SRRXXXXXXXXX\_1.fastq.gz (“прямые” чтения / forward reads) и SRRXXXXXXXXX\_2.fastq.gz (“обратные” чтения / reverse reads), где SRRXXXXXXXXX - ID вашего образца (см. Ведомость, Практикум 11)

**/mnt/scratch/NGS/DATA/rna\_reads** - одноконцевые чтения секвенирования рнк человека

Далее файлы типа: ENCFFXXXXXXXXX.fastq.gz, где SRRXXXXXXXXX - ID вашего образца (см. Ведомость, Практикум 13)

## Генная разметка

Разметка генов версии Ensembl release 110 для hg38 в формате gtf  
**/mnt/scratch/NGS/DATA/genes/Homo\_sapiens.GRCh38.110.chr.gtf**

seqcap\_hg38.bed – координаты таргета (совокупность экзонов белок-кодирующих генов)

seqcap\_hg38\_50.bed – координаты таргета (совокупность экзонов белок-кодирующих генов), расширенные на 5 нукл с двух сторон

## Оформление домашнего задания

Сдавайте домашнее задание в виде отчета, оформленного в документе с расширением .pdf.

Называйте документ согласно шаблону (т.е. нужно поменять “IvanovI” на СВОЮ фамилию и первую букву имени):

**FBB\_y24\_HW11\_13\_IvanovI\_2025.pdf**. Документ, названный иначе, не будет проверен. Файл, присланный не в том формате, или нечитаемый файл не будет проверен.

Внимательно читайте формулировку задания и выполняйте только то, что от вас требуется.

Для вашего удобства организуйте директории так, чтобы вам было удобно ориентироваться в данных и чтобы вы не запутались в выходных файлах. Крайне не рекомендуется скидывать все файлы в одну директорию. Это не будет никак проверяться и остается на вашей совести.

### **Документ с отчетом должен содержать:**

- описание всех ваших манипуляций (т.е. ЧТО было сделано и ЗАЧЕМ)
- описание входных и выходных данных при использовании программ (описывайте КАК выглядят ваши файлы, КАКУЮ информацию вы можете из них почерпнуть, можно приводить примеры в виде нескольких первых строк файла)
- все использованные команды с пояснением каждого параметра
- ссылки на внешние источники, если они были использованы
- ответы на ВСЕ обязательные пункты задания
- иллюстрации при необходимости
- ваши комментарии и рассуждения (это самое важное, просто факт запуска программы и получение каких-то чисел не является выполненным заданием)

Некорректно оформленные отчеты проверены не будут.

Формально выполненные задания (т.е. без единого вашего комментария) либо не будут проверены, либо будут проверены с большим штрафом. Запуск команд - дело не хитрое, есть мануалы, форумы, откуда можно просто скопировать нужную вам команду и даже при минимальном опыте корректно запустить ее. Вы получаете результаты, картинки, числа. Эти данные кажутся вам адекватными? Программа отработала корректно? Что изменилось? Как устроены выходные файлы? Пригодны ли данные для дальнейшей работы? Например, вы получаете количество чтений после триммирования. Это много или мало, хорошо или плохо, достаточно или совсем не годится? Приветствуются любые рассуждения, картинки (можно даже мемасики, только уместные и приличные), ссылки на внешние источники.

После прочтения вашего отчета проверяющий должен быть убежден, что задание выполнено корректно, логика и ход ваших рассуждений понятны, отчет в достаточном объеме проиллюстрирован там, где это необходимо и уместно. Также проверяющий не должен обращаться к дополнительным источникам (например, уточнять формулировки заданий, если они отсутствуют в отчете или выяснять авторство отчета). Для демонстрации корректности выполненного задания можно, например, выводить небольшой фрагмент исследуемого файла, выводить результаты работы программ, если это возможно, исходя из объема вывода (т.е. не нужно выводить содержимое

файла на 3 страницы), уточняйте, на что стоит обращать внимание при проверке.

Проверяйте, пожалуйста, форматирование. Убедитесь, что при конвертации документа в pdf не «поехали» картинки, надписи и пр.

Документ можно сдавать ОДИН раз.

После проверки в ведомости вы сможете найти вопросы или комментарии к конкретным пунктам. Если за задание не набрано баллов, достаточных для зачета, нужно проработать ошибки и подойти к преподавателю для краткого обсуждения исправлений (именно исправлений сделанного, а не презентации несделанных ранее пунктов!!!).

Ошибки – это не страшно! Ошибки можно и нужно проработать, понять, рассказать, в чем они заключались и как были исправлены. Первая сдача доделок в случае хорошего объяснения не карается штрафами.

До мягкого дедлайна можно спрашивать, уточнять формулировки, задавать любые вопросы, касающиеся задания (даже если они кажутся вам дурацкими). Но прежде постарайтесь найти ответ самостоятельно.

Внутри документа сохраняйте (!!!!) структуру, предложенную в формулировках задания. Используйте заголовки разного уровня и прочие возможности форматирования для удобства чтения вашего отчета. Должно быть понятно, какой пункт задания вы выполняете, с какими данными вы работаете, что вам было выдано на вход, что получаете на выходе и что вы об этом всем думаете.

Рисунки и таблицы, которые вы приводите в своем отчете, должны быть пронумерованы, подписаны, а в тексте указана ссылка на соответствующий объект.

Если вы заметили какие-то ошибки и\или неточности в заданиях, если какие-то формулировки можно трактовать неоднозначно или они вызывают вопросы, если что-то совсем неясно - обязательно пишите мне письма!

Задания, отмеченные (\*), не обязательны, но приносят дополнительные баллы.

## **ФОРМА ДЛЯ ОТЧЕТА**

Поздравляю! Вы выполнили наше первое задание на чтение. Теперь вернитесь в начало и прочитайте его еще раз более внимательно. Это убережет нас всех от лишних страданий, вопросов, непониманий и ошибок.