

1. [https://www.ncbi.nlm.nih.gov/sra/SRX7991621\[accn\]](https://www.ncbi.nlm.nih.gov/sra/SRX7991621[accn])
2. Скрипт на bash: pipeline.sh
3. Результат работы samtools flagstat: SRR11413027.txt
4. Скрипт разбора результата в самом скрипте на bash (строки: 11-20)
5. -
6. Несколько способов:
 - a. Предполагаю, что conda уже установлена
conda create -c conda-forge -c bioconda -n snakemake snakemake
conda activate snakemake
 - b. Если conda не установлена, то можно
pip install snakemake

Для запуска

snakemake -s <snake_file> <outputfile>

или (будет запущено первое правило)

snakemake -s <snake_file>

Чтобы pipeline печатал команды, можно добавить флаг -p

7. Код тестового pipeline: test_snakemake.smk
8. Результаты работы пайплайна: up_input.txt
Логи пайплайна: test-snakemake.log
9. -
10. Код пайплайна “оценки качества картирования” на фреймворке: pipeline.smk
11. Результаты работы: results.txt
12. Лог-файлы работы пайплайна на загруженных данных: snake_pipeline.log
13. Визуализация pipeline: dag.svg
14. Визуализацию сгенерировал с помощью команды (взял с официального туториала):
snakemake --dag -s pipeline.smk | dot -Tsvg > dag.svg
dot команда из Graphviz

Т.к. шаги fastQC и BWA независимы, то их можно выполнять одновременно

Также на схеме нет ветвления на шаге проверки качества, т.к. эта логика написана в .ru файле а на содержимое файла визуализация не смотрит

samtools sort не зависит от файлов из шагов samtools flagstat и check mapping quality поэтому пришлось сделать искусственную зависимость, чтобы сохранить логику проверки качества перед шагом samtools sort