

Итоговое занятие: интеграция методов

CNE 8972 Биоинформатика · 6B05105 Биотехнология
Неделя 14 · Жаниязов Ж.А., старший преподаватель

План лекции

- Цель мини-проекта
- Две линии проекта
- Команды и схемы анализа
- Воспроизводимость и отчёт
- Критерии оценки

14 / 15

неделя · мини-проект

Цель мини-проекта

- Объединить навыки курса: Linux, контроль качества, сборка, аннотация, популяционные данные
- Работа в группах на реальных данных
- Результат: отчёт, команды и презентация
- Применить PLINK и данные 1000 Genomes

Проект

анализ NGS + PLINK/1000G

Две линии проекта

А: геном микроорганизма

NGS → качество → тримминг → сборка → QUAST
→ аннотация

В: популяционные данные

VCF 1000 Genomes → фильтрация → PLINK →
частоты аллелей

Общая схема



Каждый этап оформляется командами и результатами в отчёте.

Линия А: команды

Шаг	Программа
Качество	fastqc, multiqc
Тримминг	trimmomatic PE
Сборка	spades.py --isolate
Оценка сборки	quast.py
Аннотация	prokka

Линия В: команды

Шаг	Команда
Скачивание	wget (VCF 1000 Genomes)
Просмотр	zcat file.vcf.gz head
Фильтрация	bcftools view -v snps
Бинарный формат	plink --vcf file.vcf.gz --make-bed
Частоты	plink --bfile data --freq

Организация проекта

data/raw

Исходные данные: не менять

results

Результаты каждого шага

logs

Журналы программ

README

Описание шагов и версий программ

Воспроизводимость

- Фиксируйте версии: `conda env export`
- Записывайте команды в скрипт `.sh`
- Сохраняйте журналы и параметры
- Другой человек должен повторить анализ

README

как повторить анализ

Интерпретация результатов

Качество данных

Достаточно ли для сборки

Геном

Размер, GC, число генов

Гены интереса

Функции и гомологи

Популяции

Частоты и различия

Структура отчёта

Раздел	Содержание
Введение	Цель и задачи
Данные и методы	Источники, программы, параметры
Результаты	Таблицы и рисунки
Выводы	Итоги и ограничения
Источники	Литература и ссылки на программы

Критерии оценивания

Критерий	Вес
Аккуратность и точность	30%
Творчество и креативность	30%
Полнота и зрелость	40%
Оригинальность	коэффициент 1,0; 0,5 или 0

Итоги и контрольные вопросы

- Какие этапы входят в линию А и линию В?
- Как обеспечить воспроизводимость анализа?
- Что включить в отчёт?
- Практика: ПР №14 «Мини-проект: анализ NGS данных + PLINK/1000G»

121–129

стр. · [2] Baxevanis