

# Russian Exomes

Измайлова Екатерина  
Полещук Ольга

Руководитель:  
Барбитов Юрий  
(Институт Биоинформатики)

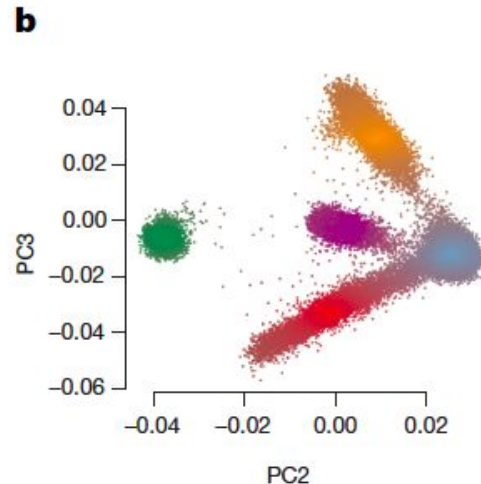
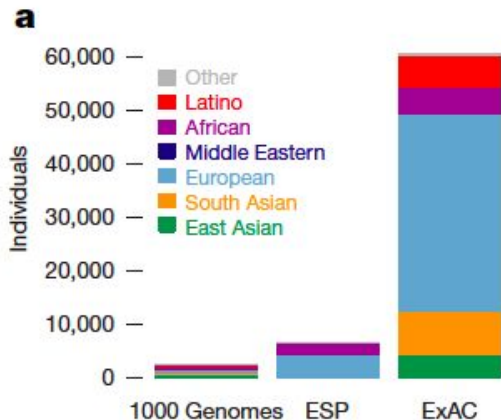
# Анализ экзонов и популяционные исследования в мире

ARTICLE

OPEN

doi:10.1038/nature19057

## Analysis of protein-coding genetic variation in 60,706 humans



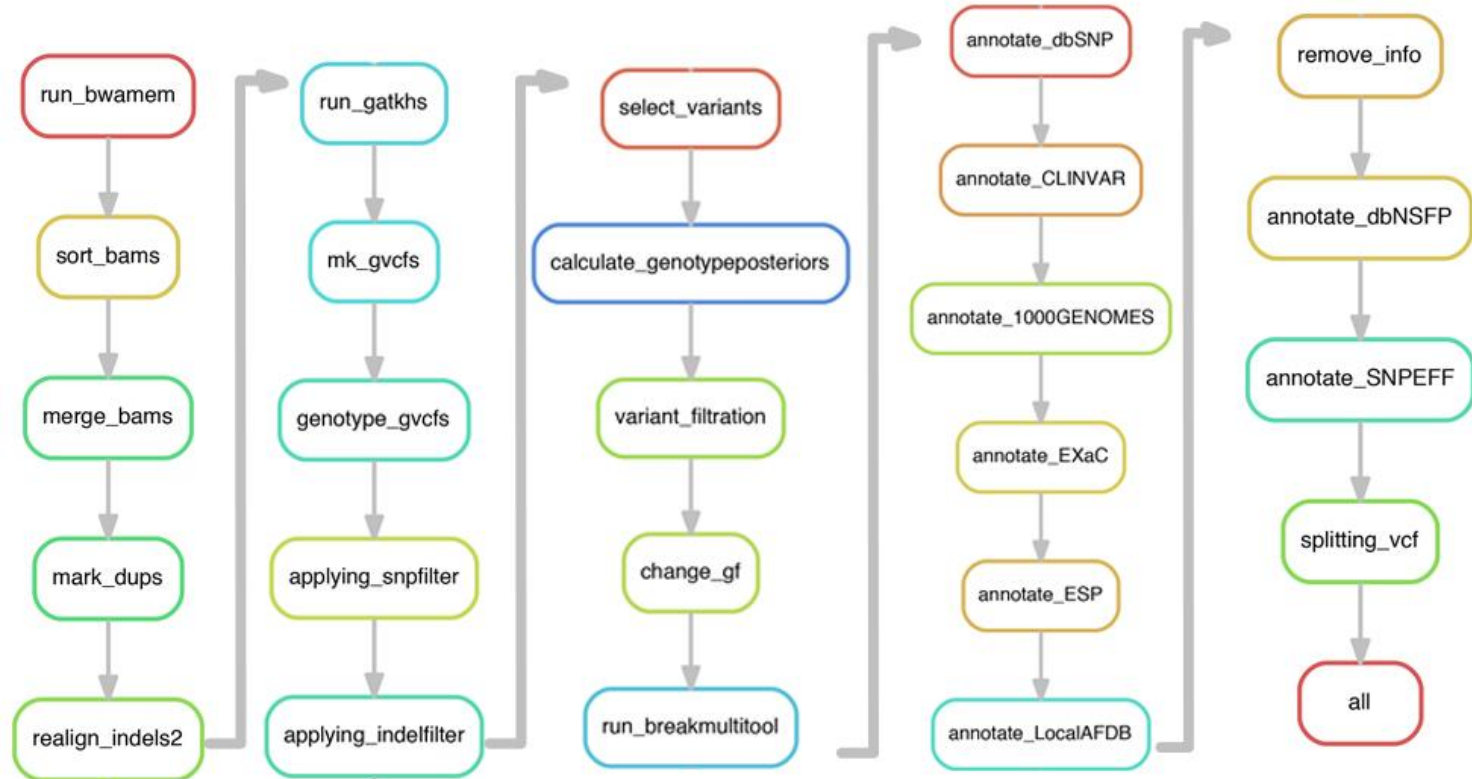
# Цель

Обработка и анализ данных, полученных путем секвенирования экзомов среди представителей российской популяции.

## Задачи

1. Изучить этапы работы с данными секвенирования, пайплайны GATK.
2. Освоить языки, на которых созданы инструменты для написания пайплайнов (SnakeMake).
3. Создать пайплайн, позволяющий из сырых данных получить VCF-файлы.
4. Произвести обработку как можно большего числа образцов с целью выявления особенностей распределения частот аллелей исследуемых генов в российской популяции. Проанализировать полученные данные на предмет наличия выраженных различий частот некоторых аллелей между российской и изученными ранее популяциями.

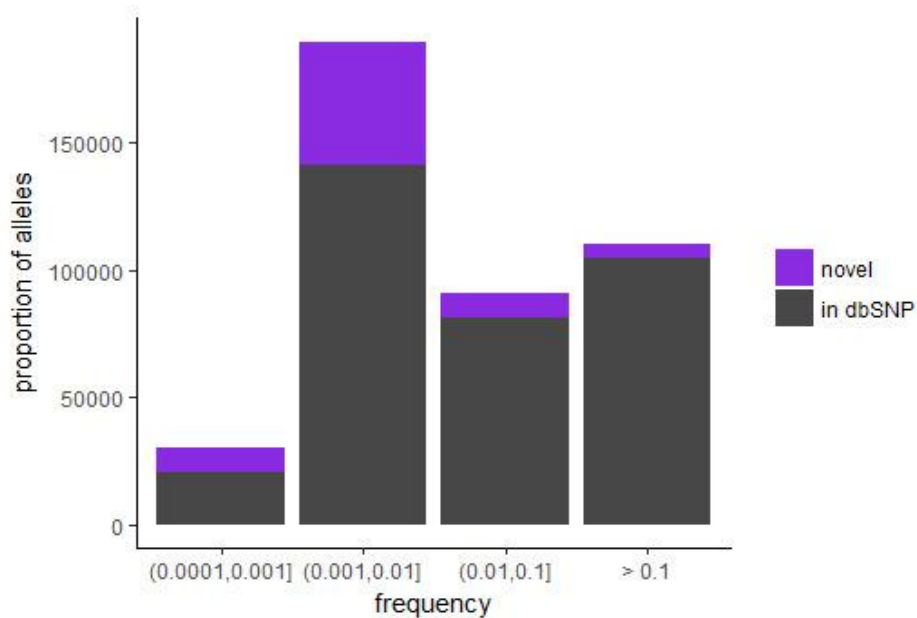
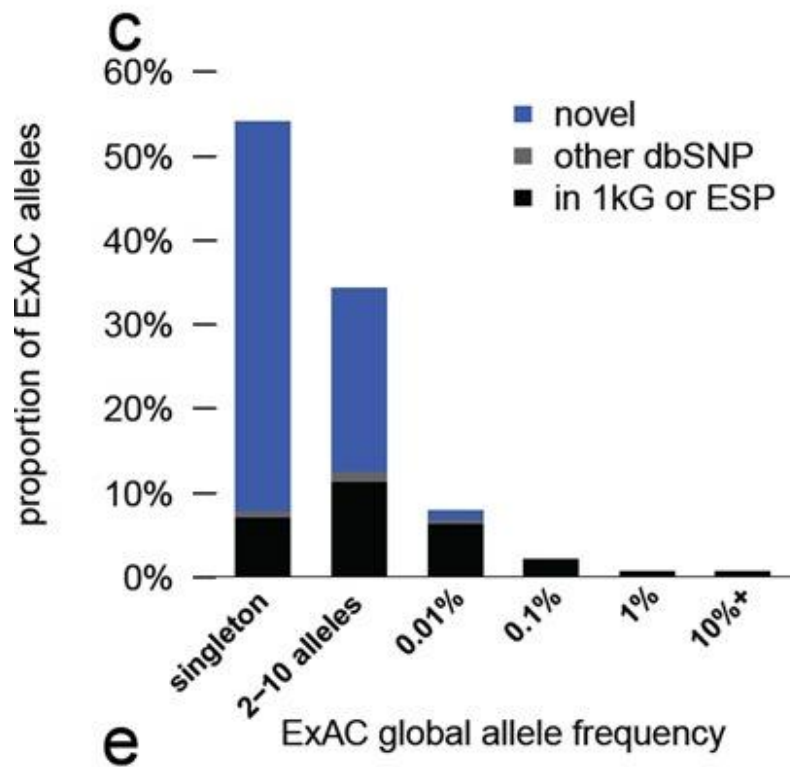
# Пайплайн для обработки данных экзомов





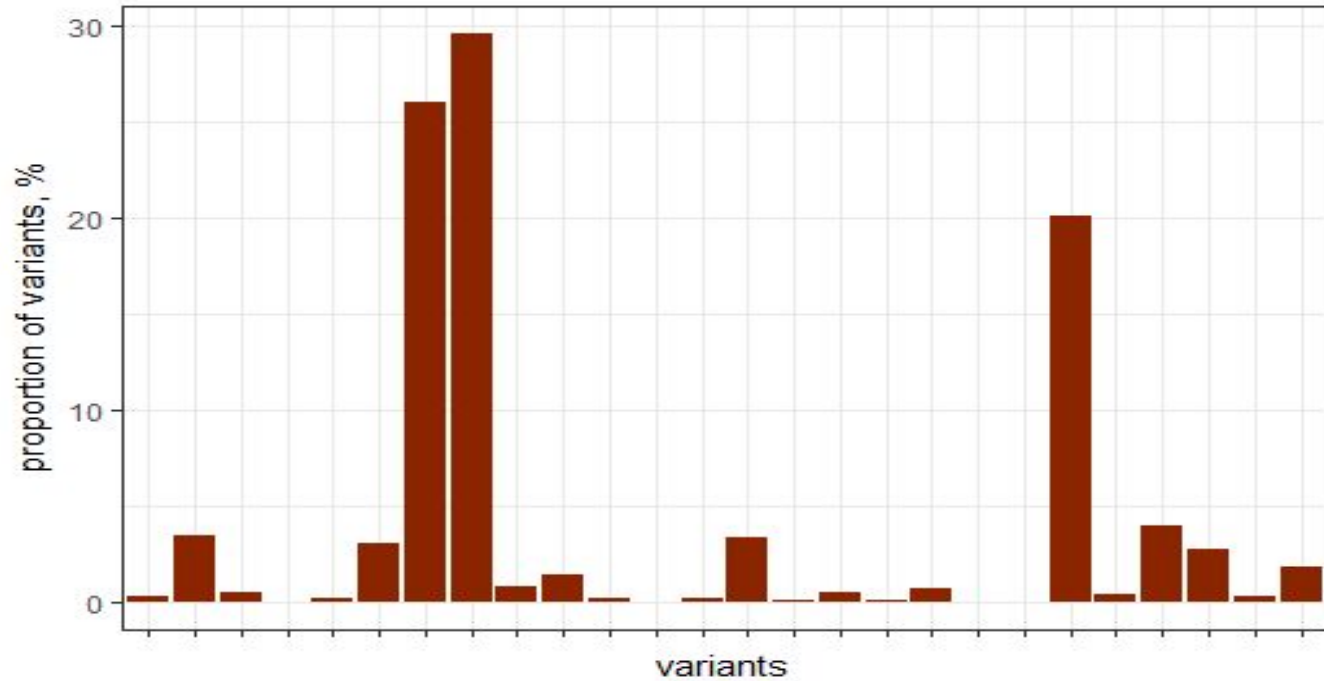
# Анализ данных секвенирования

# Число новых и известных вариантов: ExAC и REX

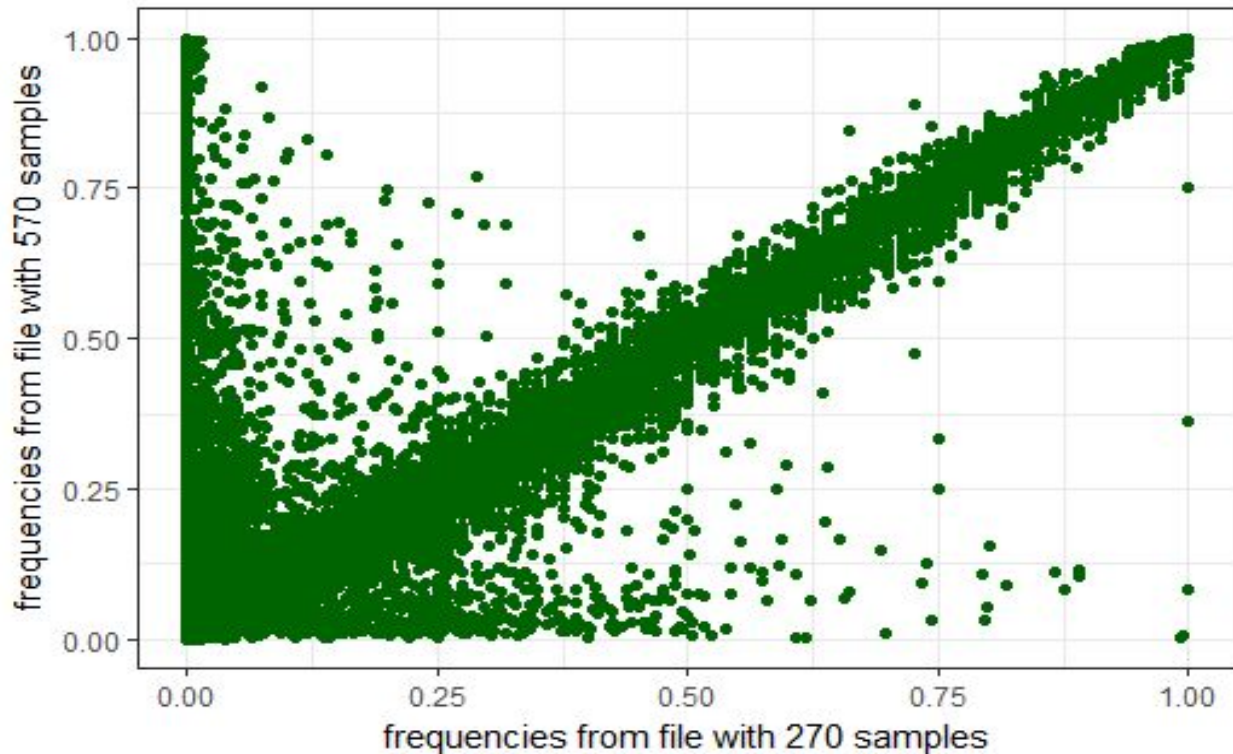


**e**

# Распределение вариантов по классам



# Частоты вариантов в выборках размером 270 и 570 образцов





# Выводы

---

1. В REX при разбиении обнаруженных вариантов по частотам во всех группах отмечается преобладание известных вариантов, занесенных в базу данных dbSNP
2. Наблюдается сравнительно низкое число вариантов с небольшой частотой (группа 0.0001-0.001)
3. Большая часть найденных вариантов принадлежит одной из трех групп: миссенс-мутации, варианты, обнаруженные в интронах и синонимичные замены.
4. С увеличением объема выборки возрастает число вариантов, которые не встречались в выборке меньшего размера. Также изменяется частота некоторых вариантов.

**Ольга:** работа со Snakemake, построение графа

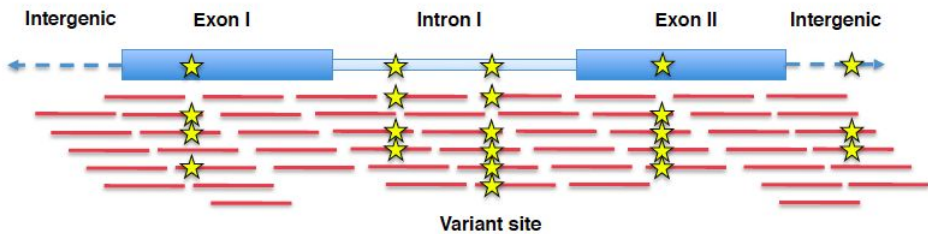
**Екатерина:** анализ данных секвенирования



**БЛАГОДАРИМ ЗА ВНИМАНИЕ!**

# Секвенирование экзонов: как и зачем

Whole  
Genome



Exome  
(+ gene panels)

