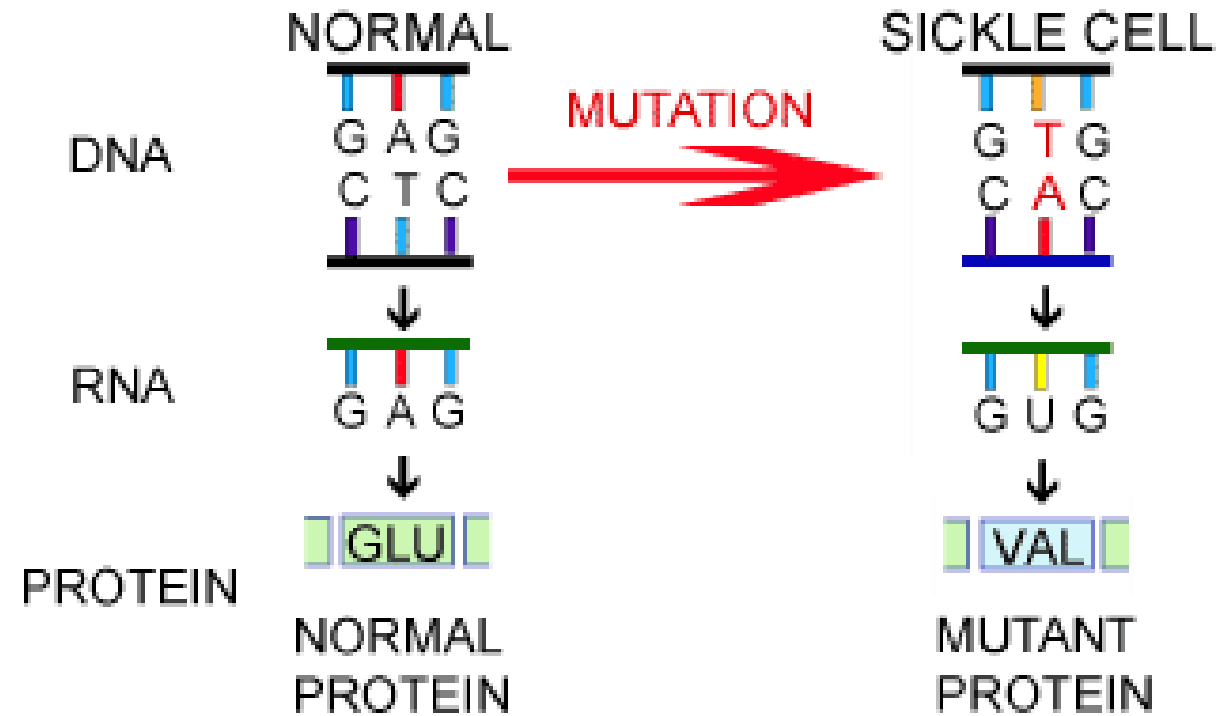
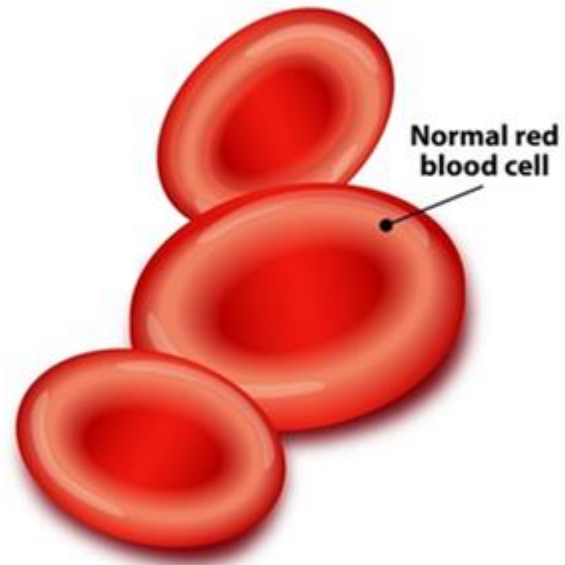


Поиск патогенных пар мутаций в геноме человека

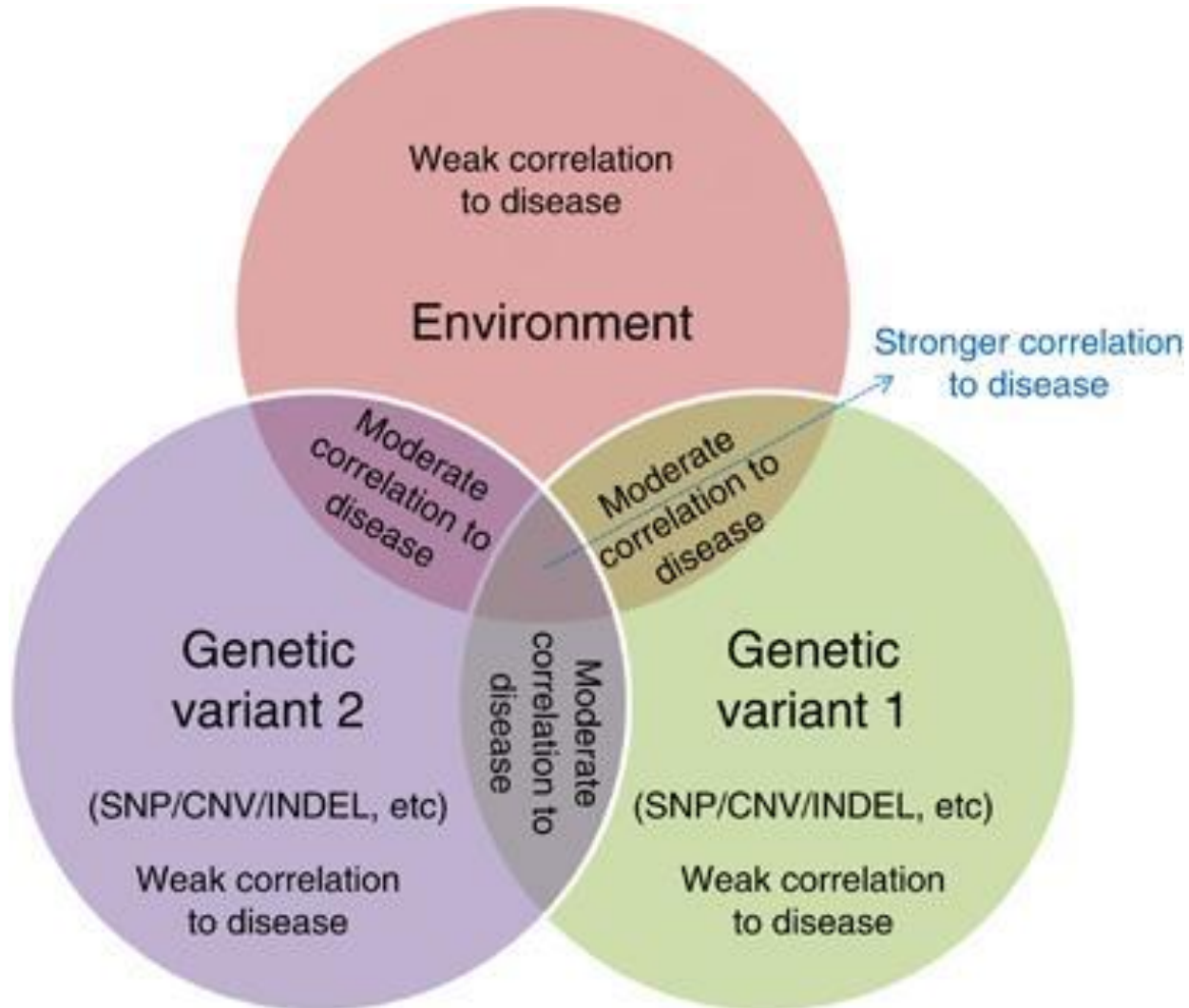
**Состав команды: Дмитрий Клабуков, Алексей Корнеев,
Юрий Лебеда, Игорь Лопатинский, Ксения Черенкова**

ПРИМЕР МОНОГЕННОГО ЗАБОЛЕВАНИЯ



Серповидноклеточная анемия

ПРИМЕРЫ МУЛЬТИФАКТОРНЫХ ЗАБОЛЕВАНИЙ



- Астма
- Диабет 2 типа
- Заболевания сердца и сосудов
- Артериальная гипертензия
- Метаболический синдром
- Бесплодие

ЦЕЛЬ ПРОЕКТА

- Поиск и анализ статистически значимой зависимости между парами генных мутаций

Задачи:

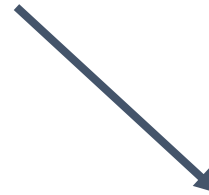
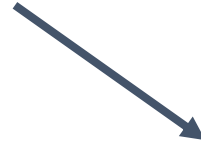
- Преобразовать данные проекта «1000 Genomes» в удобный для работы формат
- Провести попарный статистический анализ мутаций
- Соотнести полученную информацию с базой данных известных генных и белковых взаимодействий BioGRID

PIPELINE

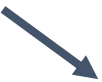
**ДААННЫЕ “1000
ГЕНОМОВ”**

**BEDTOOLS
BCFTOOLS**

**ОТФИЛЬТРОВАННЫЕ
ДААННЫЕ С
ХРОМОСОМАМИ**



**ОТФИЛЬТРОВАННЫЕ
ДААННЫЕ С
ХРОМОСОМАМИ**



**ПРОГРАММА
ЗАПОЛНЕНИЯ
БАЗЫ
ДААННЫХ**



**ПЕРЕБОР
ПАР
МУТАЦИЙ**



**СТАТИСТИЧЕСКИ
ЗНАЧИМЫЕ ДААННЫЕ**



**КОНТРОЛЬНАЯ
ГРУППА**

FDR

False Discovery Rate

0.10 ?

0.05 ?

0.01 ?

Малое число = 5 ?

FWER

Family-Wise Error Rate

0.0001 !

20 !

РЕЗУЛЬТАТЫ

- Обработано **2 млрд 360 млн** пар мутаций
- Получено **470 тыс** (около 0.2%) статистически значимых пар мутаций
 - Из них 644 (**0.14 %**) принадлежат к взаимодействующим генам по информации базы BioGrid
 - В 10 контрольных выборках такого же размера, полученных случайным образом из статистически незначимых пар, от 270 до 327 (**0.06...0.07 %**)

ВЫВОДЫ

- Реализован пайплайн, отбирающий пары мутаций, распределение генотипов которых не является независимым со статистической значимостью.
- Показано, что существует большое количество пар мутаций, распределение генотипов которых не соответствует гипотезе об их независимости.
- Пары мутаций аннотированы в соответствии с базой BioGRID.
- Установлено, что среди статистически значимых мутаций вдвое выше доля тех, для генов которых описано взаимодействие.

Supplementary 1. Пример статистически значимой позиции

Ожидаемая матрица				Наблюдаемая матрица				Отношение			
	0/0	0/1	1/1		0/0	0/1	1/1		0/0	0/1	1/1
0/0	46	289	10	0/0	137	457	771	0/0	3.0	1.6	77.4
0/1	128	796	27	0/1	140	378	328	0/1	1.1	0.5	11.9
1/1	163	1011	35	1/1	68	116	109	1/1	0.4	0.1	3.1