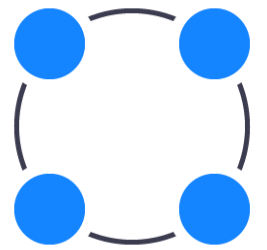


Поиск множественных ассоциаций в данных GWAS

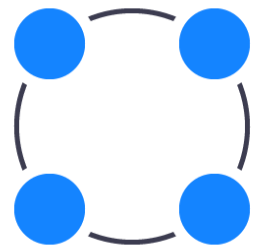
Руководитель: Юрий Барбитов,
Институт биоинформатики

Выполнил: Шиков Антон

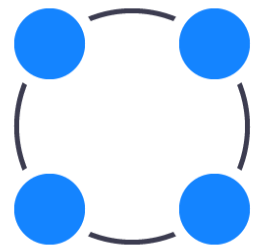
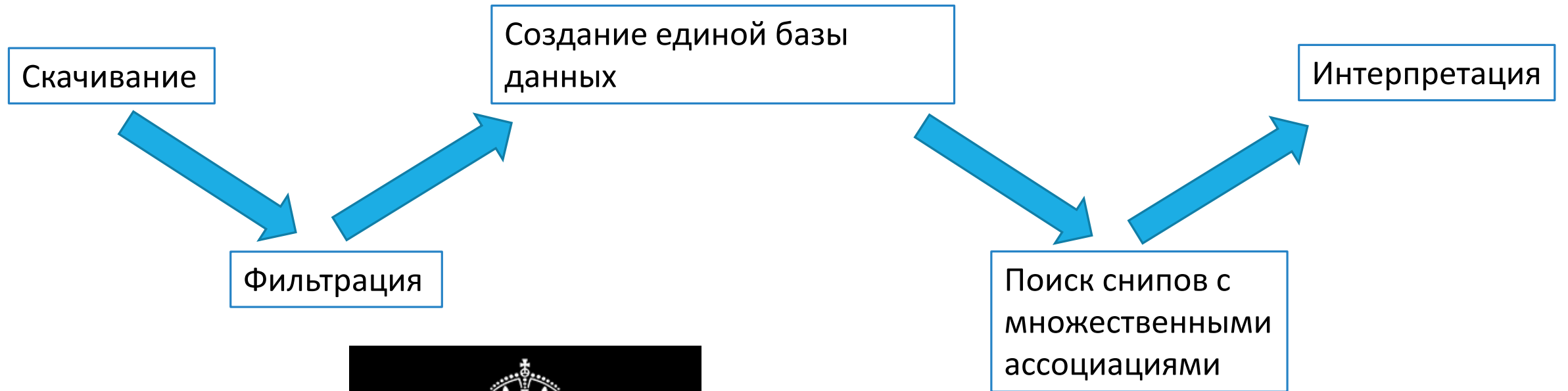


Цели проекта

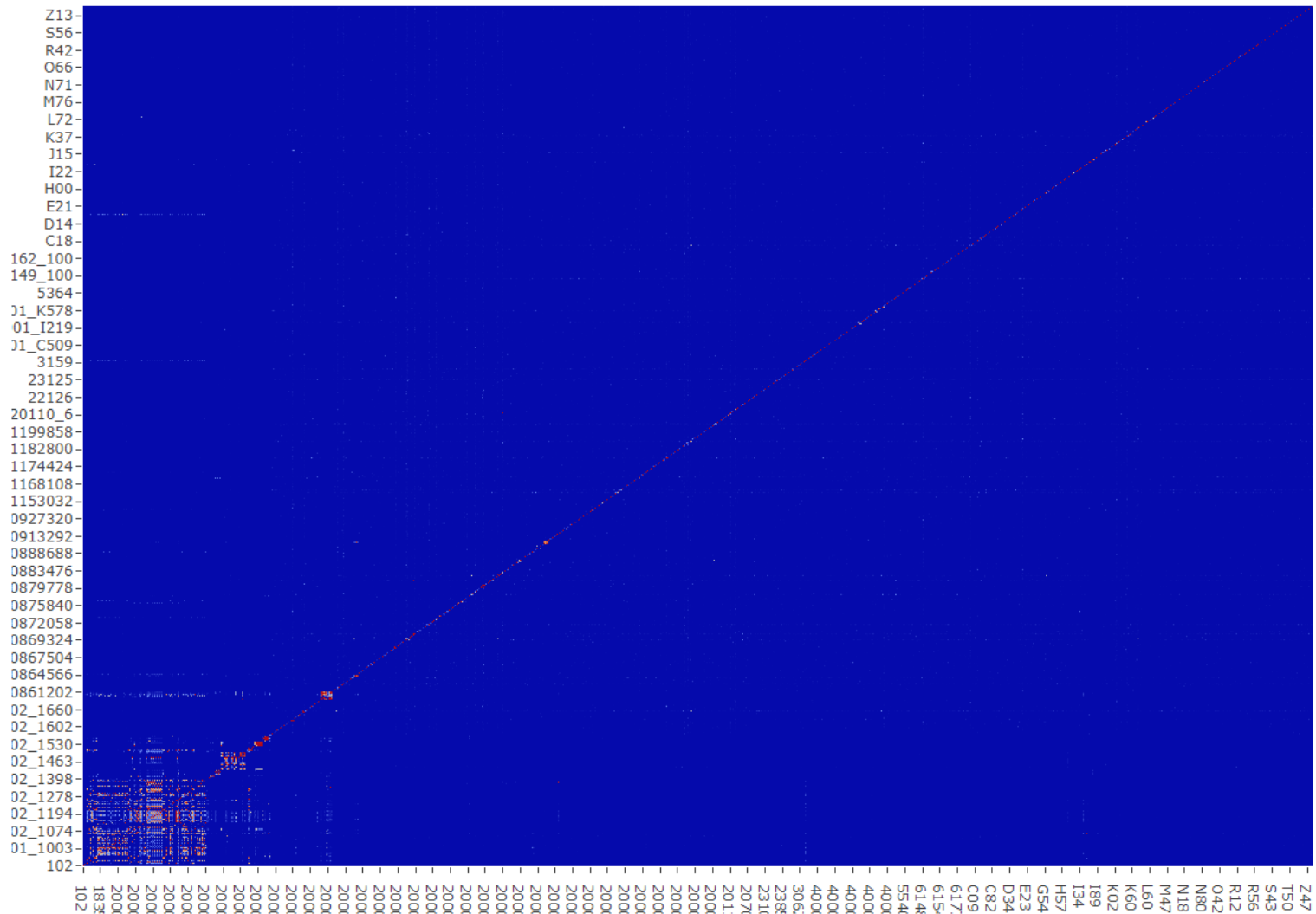
- Ознакомиться с данными GWAS-исследований когорты UK Biobank.
 - Скачать и отфильтровать значимые ассоциации по всем изучаемым фенотипам.
 - Определить GWAS-маркеры, имеющими множественные ассоциированные фенотипы.
-
- Изучить механизмы плейотропного действия GWAS-маркеров при помощи данных геномных экспериментов и анализа фенотипических корреляций.



Pipeline

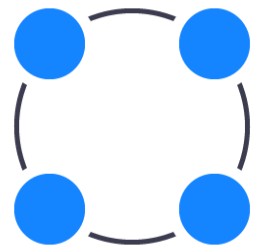


Методы снижения множественности

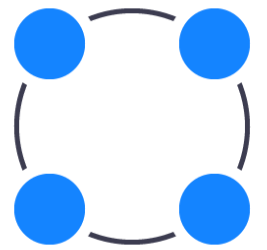
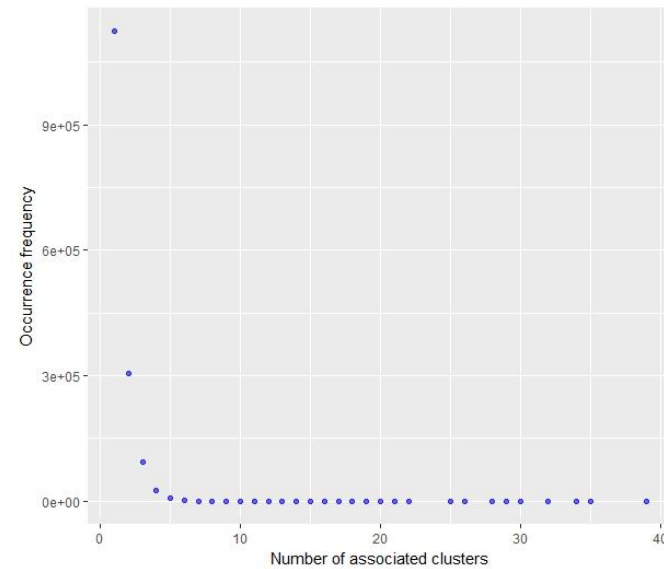
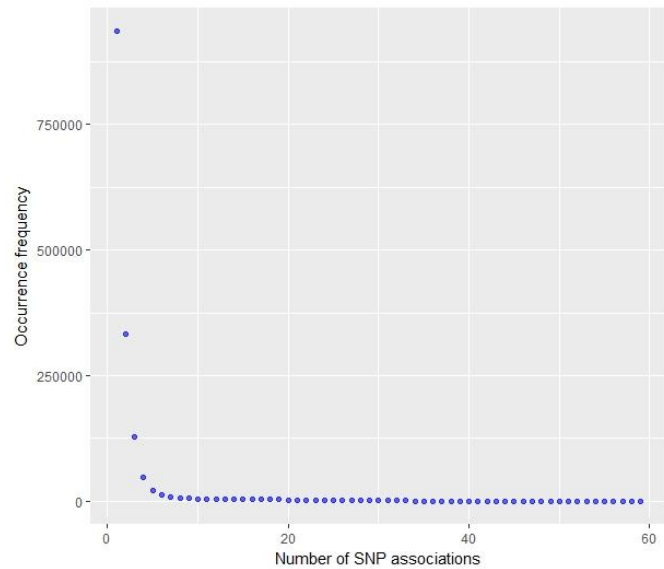
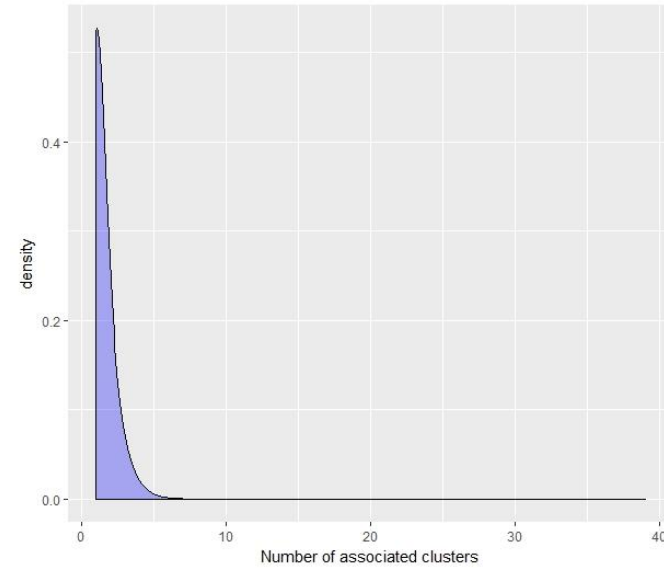
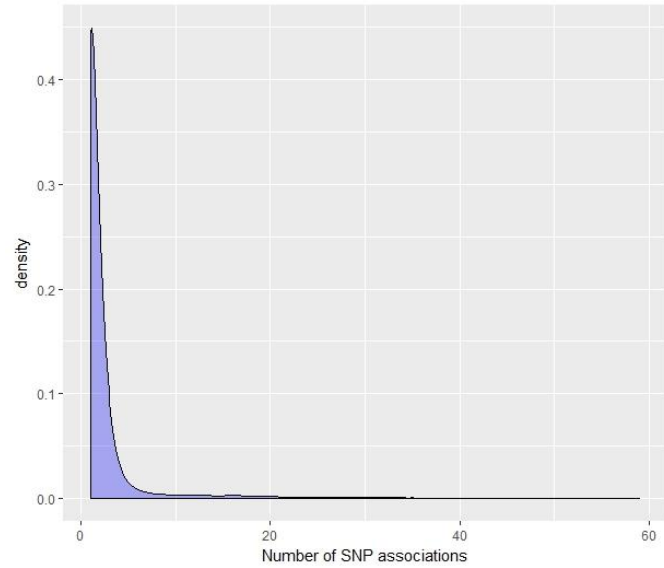


$$K_{0,+1} = \frac{n(A \cap B)}{\min[n(A), n(B)]}$$

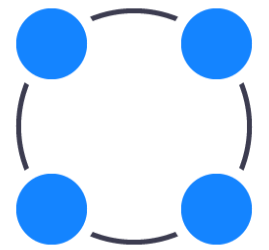
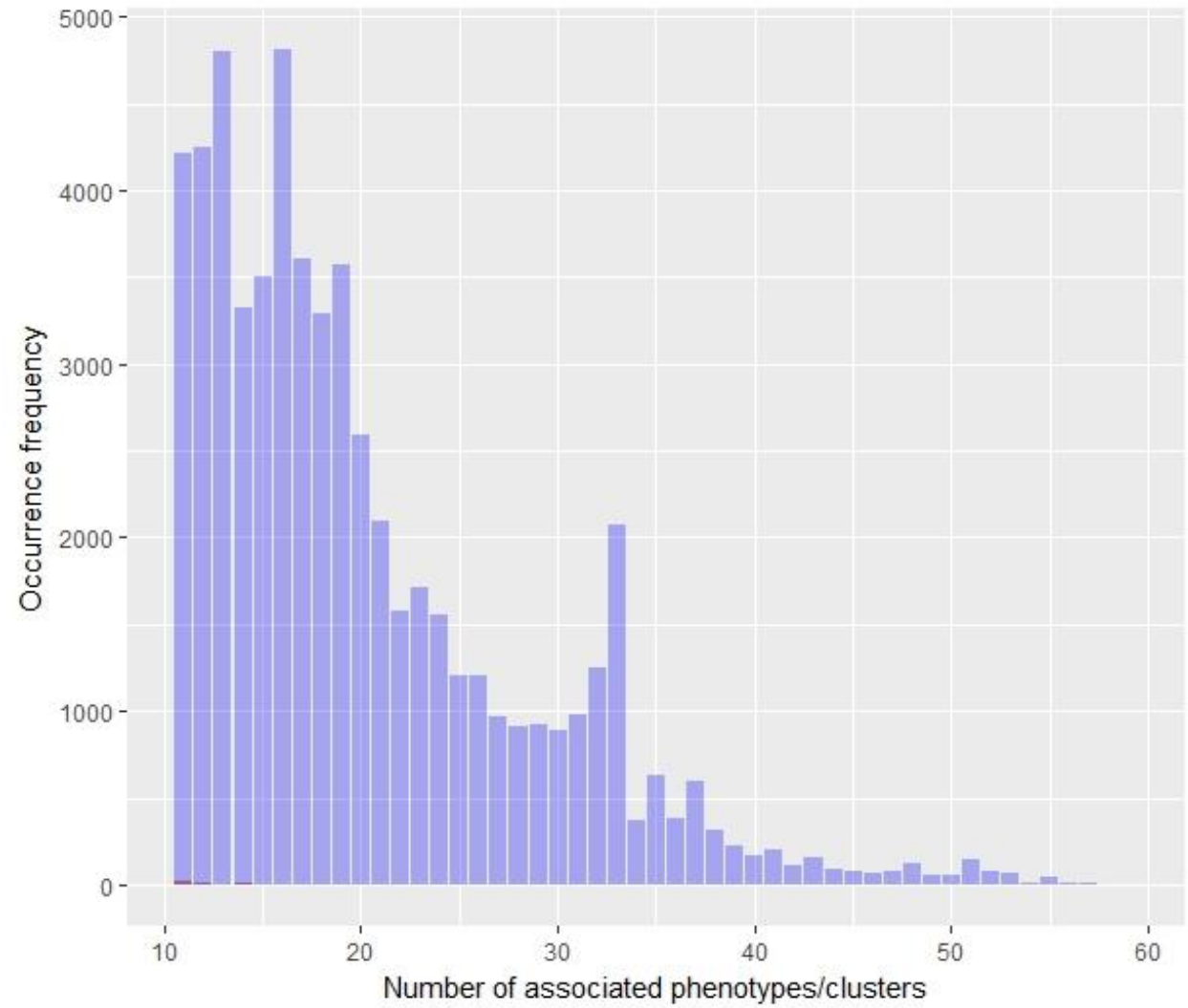
- Использование Меры Шимкевича-Симпсона для общих снипов по всем фенотипам
- Проведение иерархической кластеризации



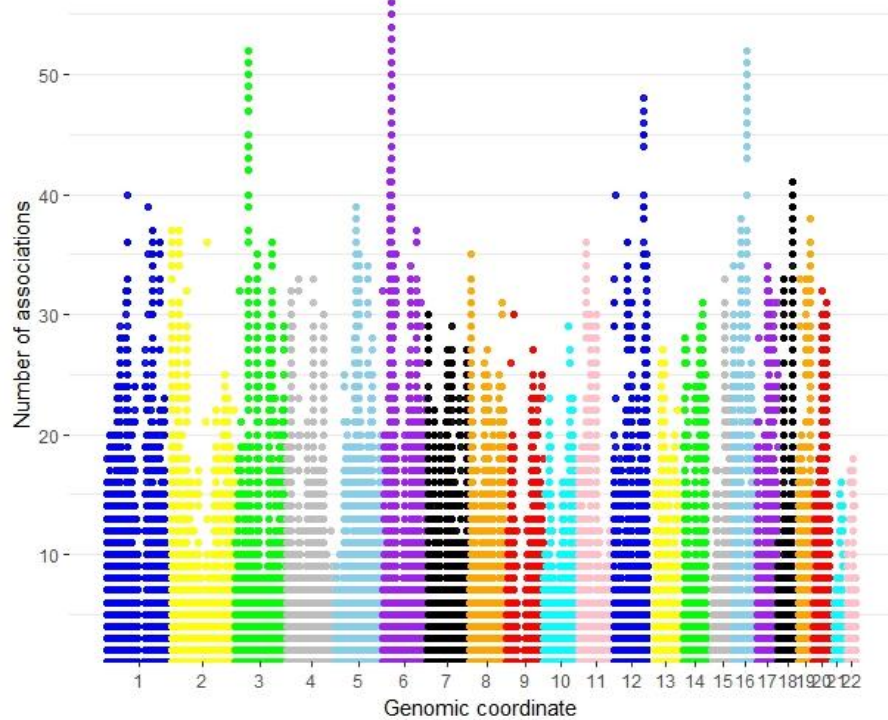
Методы снижения множественности



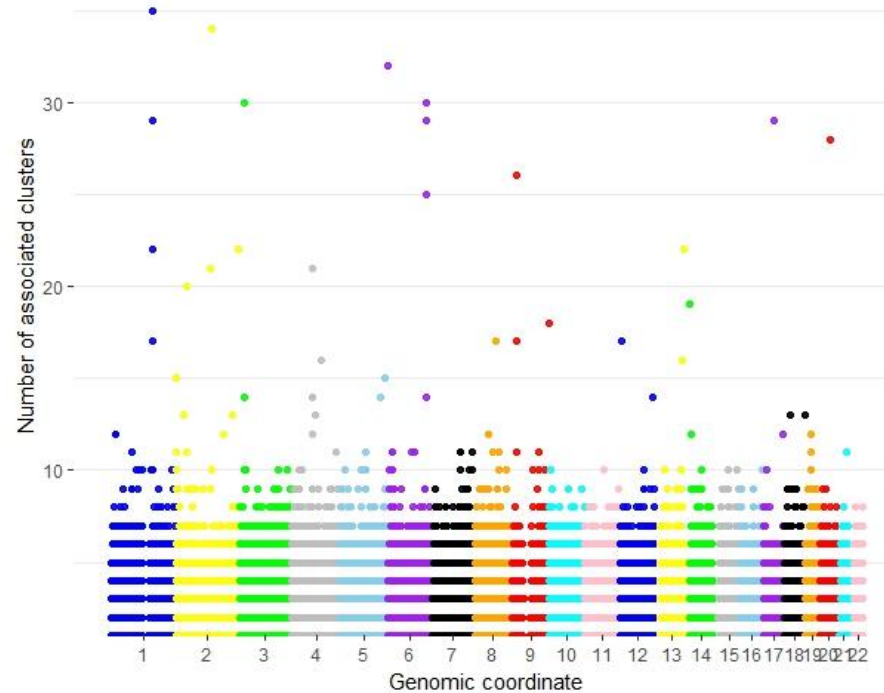
Методы снижения множественности



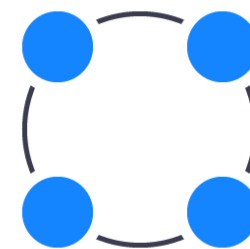
Методы снижения множественности



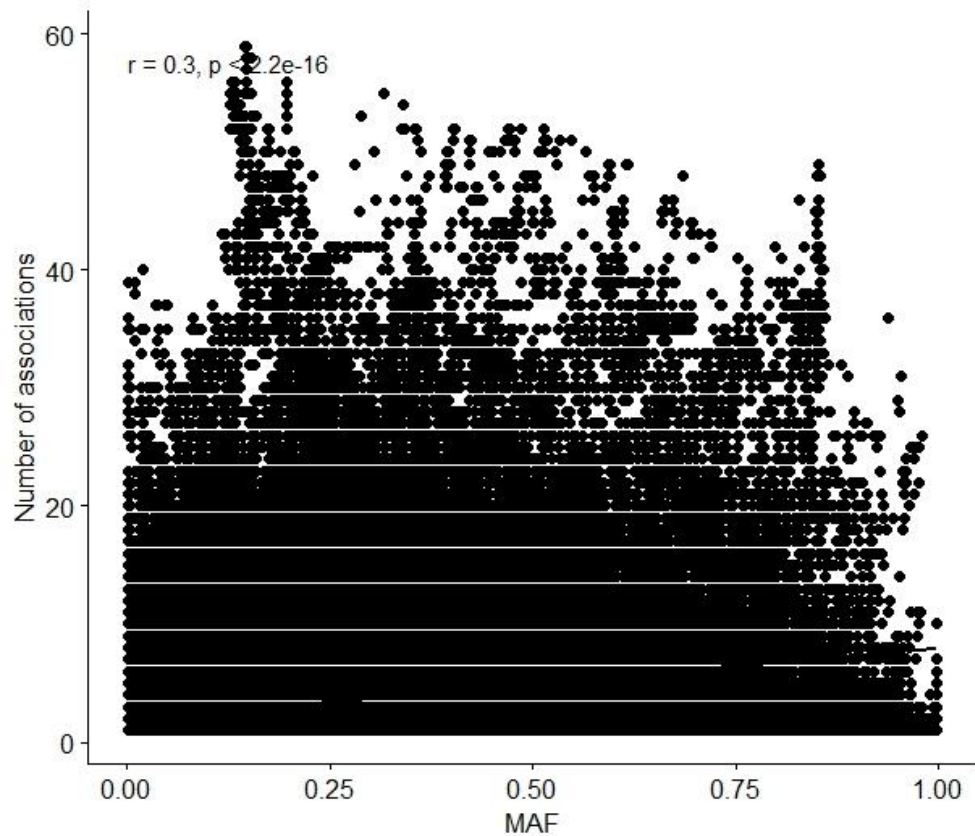
До кластеризации



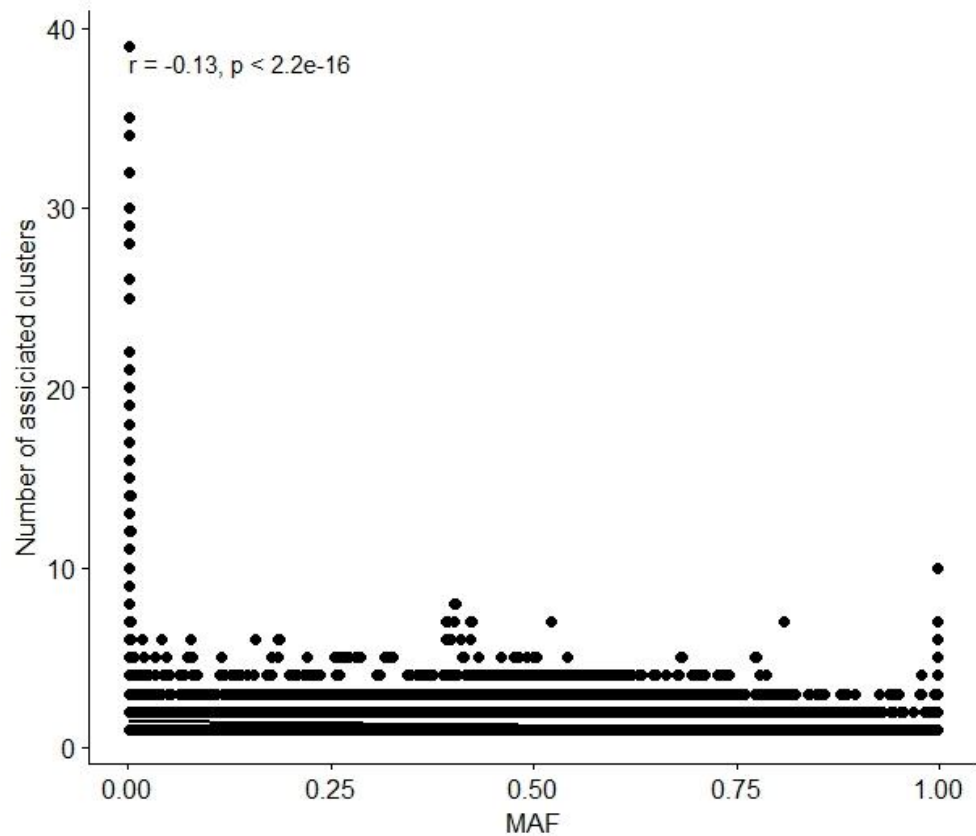
После кластеризации



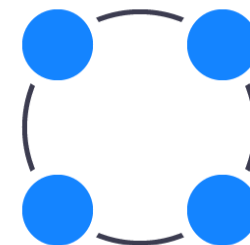
Методы снижения множественности



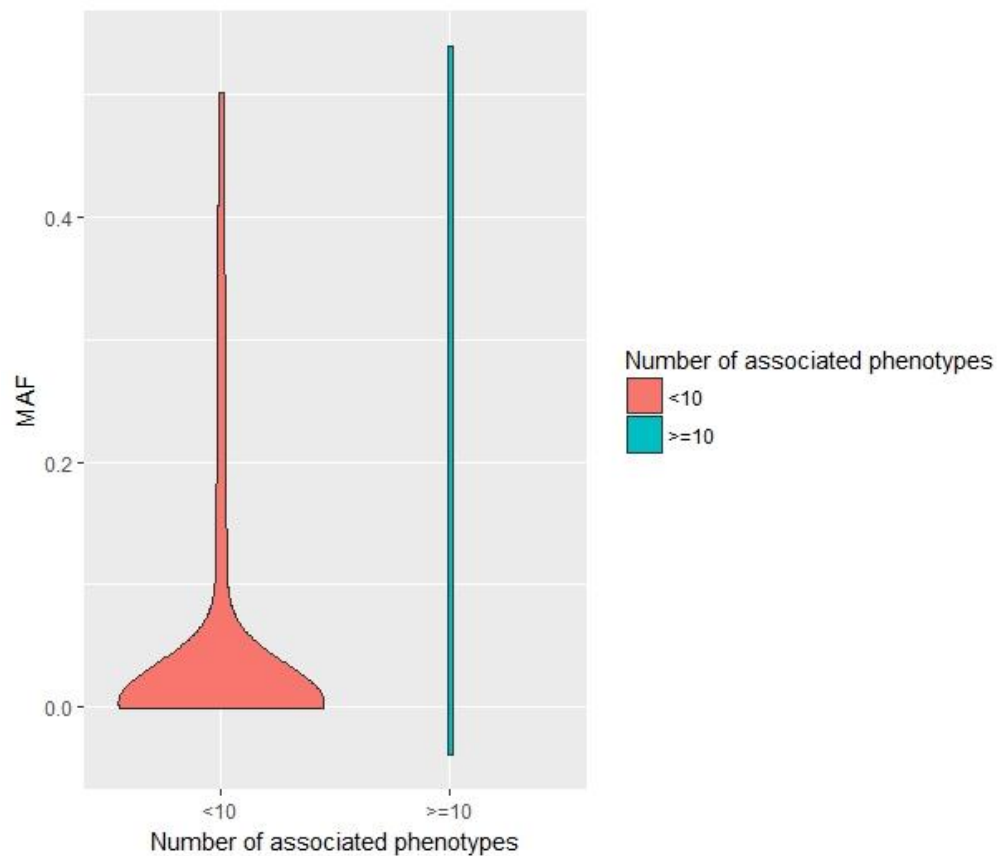
До кластеризации



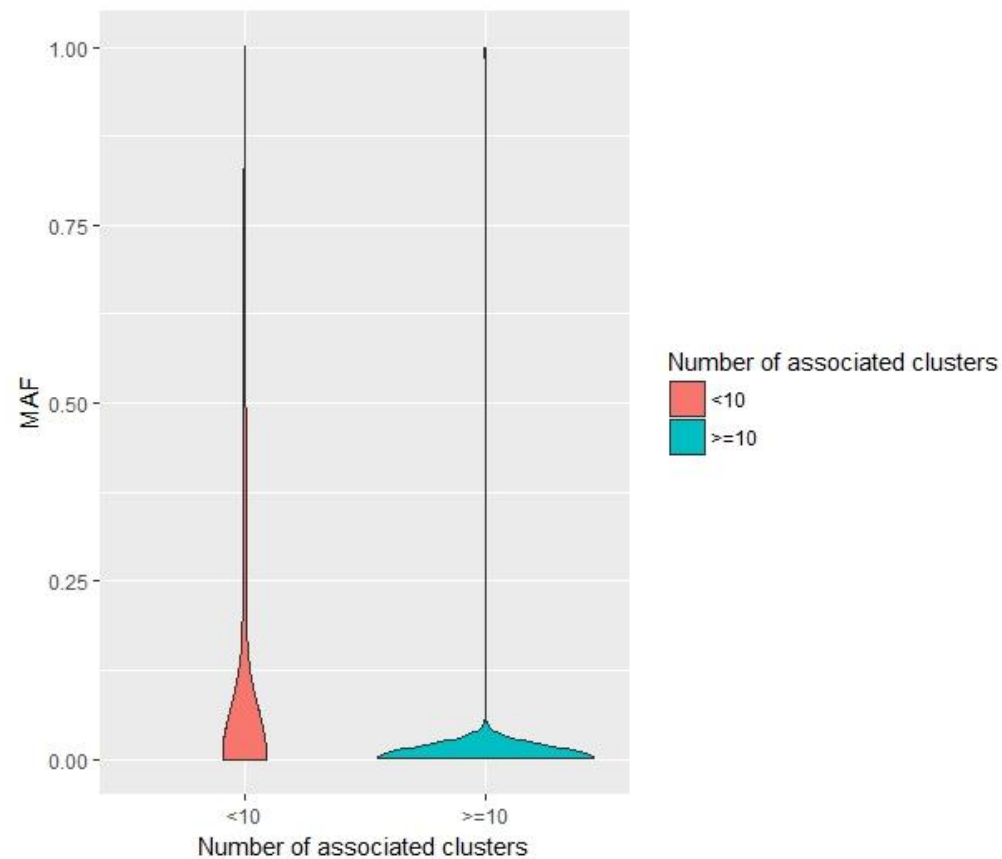
После кластеризации



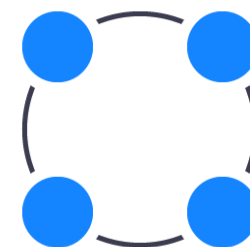
Методы снижения множественности



До кластеризации



После кластеризации



Функциональная аннотация в GREAT

⊖ Disease Ontology (6 terms)

Table controls: Export

Shown top rows in this table: 20 Set

Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Val
vascular disease	7	1.3189e-5	4.2110e-3
Varicellovirus infectious disease	9	4.5592e-5	1.1322e-2
retinal disease	14	9.6236e-5	1.5363e-2
hepatocellular carcinoma	17	1.8804e-4	2.4721e-2
liver and intrahepatic biliary tract carcinoma	19	2.3590e-4	2.7749e-2
primary immunodeficiency disease	21	3.1510e-4	3.3536e-2

Без кластеризации

+ GO Molecular Function (no terms)

+ GO Biological Process (no terms)

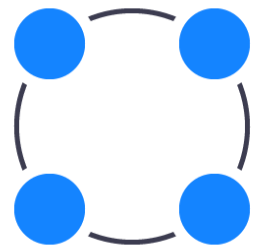
+ GO Cellular Component (no terms)

+ Mouse Phenotype (no terms)

+ Human Phenotype (no terms)

+ Disease Ontology (no terms)

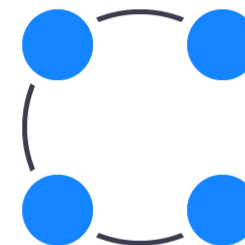
С кластеризацией



Некоторые снипы

SNP	Ген	Число ассоциаций	Тип эффекта
rs186009694	NCKAP5	51	intron_variat MODIFIER
rs563467513	SH3BP5	50	intron_variat MODIFIER
rs770943503	CHD6	37	intron_variat MODIFIER
rs540119985	RAB5C	39	intron_variat MODIFIER
rs538743783	CAMK2D	32	intron_variat MODIFIER

- Большая часть локальных максимумов попадает в область активного хроматина и дают энричмент на H3K27Ac
- Не удалось найти eQTL по данным GTEx из-за редкости снипов



Некоторые снипы

rs563467513 в гене SH3BP5

Ассоциация с тромбозом, гранулематозом, диабетической нейропатией, плантарным фасциитом, эрозией шейки матки, нейромой, альвеолитом и др.

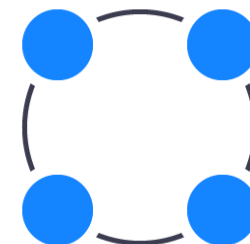
Ген вовлечен в работу иммунной системы

rs186009694 в гене NCKAP5

Ассоциация с миокардитом, нарушением митрального клапана, обсессивно-компульсивным расстройством, расстройством пищевода, талассемией, отказом сердца и др.

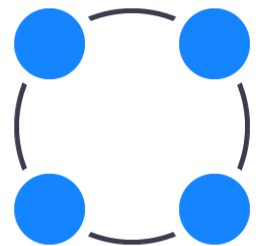
Ген связан с регуляцией циркадных ритмов (!)

- Для снипов со множественными ассоциациями характерна ассоциация с длительностью приема различных медикаментов (Treatment/medication code)
- Плейотропный эффект нуждается в дальнейшей проверке



Дальнейшие планы

- Симуляция поведения искусственного датасета с редкими снипами.
- Анализ большего числа синипов-локальных максимумов
- Функциональная аннотация



Спасибо за внимание!

-
- Все скрипты и необходимая информация доступна на репозитории
https://github.com/anton-shikov/GWAS_project

