

ПОДХОД К КЛАСТЕРИЗАЦИИ ПАЦИЕНТОВ ПО МИКРОЧИПОВЫМ ДАННЫМ ВНУТРИ ОТДЕЛЬНЫХ ЛОКУСОВ С ИСПОЛЬЗОВАНИЕМ КОМБИНАЦИЙ ВАРИАНТОВ

Л. Н. Юльметова, Н. А. Кулемин, Е. И. Шарова 

Федеральный научно-клинический центр физико-химической медицины имени Ю. М. Лопухина Федерального медико-биологического агентства, Москва, Россия

Дистрофия роговицы Фукса является социально значимым наследственным заболеванием. Более половины случаев в европейской популяции вызваны увеличением числа тринуклеотидных повторов в гене *TCF4*. Целью исследования было разработать и проверить подход разделения пациентов на группы на основе результатов чип-генотипирования и полногеномного ассоциативного исследования (GWAS). В качестве исходных данных использовали датасеты FECD Genetics Multi-center Study и проекта AREDS в количестве 1721 клинических случаев и 2408 контрольных пациентов. При анализе результатов GWAS было проведено разделение пациентов и группы контроля на две группы методом иерархической кластеризации с учетом предположения, что пациенты с увеличенным числом повторов в гене *TCF4* имеют определенные сочетания геномных вариантов (гаплотипов). Показано, что одиночные варианты не могут быть использованы для молекулярно-генетической классификации пациентов с увеличенным числом повторов в гене *TCF4* из-за рассогласованности результатов по вариантам. При этом гаплотипный подход превзошел анализируемые варианты по параметру отношения шансов, перекрывая 95%-й доверительный интервал выборки двух экспериментальных исследований. Предложенный метод позволяет продолжать поиск биологически обоснованных сочетаний геномных вариантов.


Ключевые слова: полногеномное ассоциативное исследование, эндотелиальная дистрофия роговицы, экспансия тринуклеотидных повторов, классификация пациентов, локус

Финансирование: работа выполнена в рамках гранта Президента РФ для молодых ученых-кандидатов наук МК-2951.2022.1.4.

Благодарности: авторы благодарят dbGaP за предоставление доступа к наборам данных phs000421.v1.p1 и phs000001.v3.p1. Набор данных с регистрационным номером dbGaP phs000421.v1.p1 получен из исследования генетики эндотелиальной дистрофии роговицы Фукса (FECD) https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000421.v1.p1. Авторы признают гранты, финансировавшие регистрацию случаев и контролей, которые будут использоваться в этом GWAS: R01EY016514 (DUEC, PI: Gordon Klintworth), R01EY016482 (CWRU, PI: Sudha Iyengar) и 1X01HG006619-01 (PI: Sudha Iyengar, Натали Афшари). Авторы благодарят участников FECD и исследовательскую группу FECD за их ценный вклад в это исследование. Набор данных с регистрационным номером dbGaP phs000001.v3.p1 получен из базы данных исследования возрастных заболеваний глаз (AREDS) по адресу https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000001.v3.p1. Финансовая поддержка AREDS была предоставлена Национальным глазным институтом (N01-EY-0-2127). Авторы благодарят участников AREDS и исследовательскую группу AREDS за их ценный вклад в это исследование. Авторы благодарят научного сотрудника ФГБУ ФНКЦ ФХМ имени Ю. М. Лопухина Л. О. Скородумову за ценные предложения, замечания и поддержку.

Вклад авторов: Е. И. Шарова — идея и подбор данных; Е. И. Шарова, Л. Н. Юльметова — планирование и выбор методов; Н. А. Кулемин — финансирование и администрирование проекта; Л. Н. Юльметова — разработка и вычисления; Е. И. Шарова, Л. Н. Юльметова, Н. А. Кулемин — обсуждение результатов, написание и редактирование статьи.

Соблюдение этических стандартов: исследование проведено с соблюдением принципов Хельсинкской декларации, с использованием наборов данных проектов phs000421.v1.p1 и phs000001.v3.p1., доступ к которым одобрен и получен через dbGaP в соответствии с их политикой одобрения и доступа к конкретным наборам данных.

 **Для корреспонденции:** Елена Ивановна Шарова
ул. Малая Пироговская, д. 1с3, г. Москва, 119435, Россия; sharova78@gmail.com

Статья получена: 12.12.2022 **Статья принята к печати:** 20.01.2023 **Опубликована онлайн:** 12.02.2023

DOI: 10.24075/vrgmu.2023.001

THE APPROACH TO PATIENT CLUSTERING BASED ON THE MICROCHIP DATA CONFINED TO DISTINCT LOCI USING THE COMBINATIONS OF VARIANTS

Iulmetova LN, Kulemin NA, Sharova EI 

Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine of the Federal Medical Biological Agency, Moscow, Russia

Fuchs' endothelial corneal dystrophy is a socially significant hereditary disease. More than a half of cases in the European population are caused by the increased number of trinucleotide repeats in the *TCF4* gene. The study was aimed to develop and test the approach of dividing patients into groups based on the chip-based genotyping and genome-wide association study (GWAS) results. The analysis was conducted using FECD Genetics Multi-center Study and AREDS project datasets containing the data of 1721 clinical cases and 2408 control patients. When analyzing the GWAS results, the patients and the control group were divided into two groups by means of hierarchical clustering suggesting that patients with the increased number of repeats in the *TCF4* gene are carriers of specific combinations of genomic variants (haplotypes). It was shown that individual variants cannot be used for the molecular genetic stratification of patients with the increased number of repeats in *TCF4* due to inconsistent results obtained for the variants. Furthermore, the haplotype-based approach outperformed the SNPs in terms of odds ratio. The paper proposes a method that enables further search for the biologically relevant combinations of genomic variants.


Keywords: genome wide association study, Fuchs endothelial corneal dystrophy, trinucleotide repeat expansion, patient stratification, locus

Funding: the study was supported by the President of the Russian Federation (grant for young postdocs MK-2951.2022.1.4)

Acknowledgements: the authors thank dbGaP for providing access to the phs000421.v1.p1 and phs000001.v3.p1 datasets. The dataset with the dbGaP registration number phs000421.v1.p1 was obtained from the genetic study of Fuchs' endothelial corneal dystrophy (FECD) available from https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000421.v1.p1. The authors recognize the grants that have been used to support registration of cases and controls and would be used in this GWAS: R01EY016514 (DUEC, PI: Gordon Klintworth), R01EY016482 (CWRU, PI: Sudha Iyengar) and 1X01HG006619-01 (PI: Sudha Iyengar, Natalia Afshari). The authors express their gratitude to the FECD study participants and the FECD research team for their valuable contribution to this study. The dataset with the dbGaP registration number phs000001.v3.p1 was obtained from the Age-Related Eye Disease Study (AREDS) database available from https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000001.v3.p1. AREDS was funded by the National Eye Institute (N01-EY-0-2127). The authors thank the AREDS participants and research team for their valuable contribution to this study. The authors express their gratitude to L.O. Skorodumova, research fellow at the Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine, for valuable suggestions, comments, and support.

Author contribution: Sharova EI — concept and selection of data; Sharova EI, Iulmetova LN — planning and selection of methods; Kulemin NA — project funding and management; Iulmetova LN — design and computation; Sharova EI, Iulmetova LN, Kulemin NA — discussion, manuscript writing and editing.

Compliance with ethical standards: the study was performed according to the principles of the Declaration of Helsinki using the data of the phs000421.v1.p1 and phs000001.v3.p1 projects, the access to which was approved and provided by dbGaP in accordance with the policy of approval and access to specific datasets.

 **Correspondence should be addressed:** Elena I. Sharova
Malaya Pirogovskaya, 1, str. 3, Moscow, 119435, Russia; sharova78@gmail.com

Received: 12.12.2022 **Accepted:** 20.01.2023 **Published online:** 12.02.2023

DOI: 10.24075/brsmu.2023.001

Поиск биологического обоснования наследуемости фенотипов является одной из основных задач современной медицинской генетики. Исторически все известные подходы к поиску патогенных геномных вариантов делят на две группы: биологические и математические. К биологическим методам относятся подходы объяснения фенотипа на основе изученных биохимических процессов, а к математическим — различные статистические подходы. В случае, когда невозможно напрямую проследить биохимическую цепь формирования фенотипа, но развитие заболевания явно отягощено семейным анамнезом, применяют различные статистические подходы: полногеномные ассоциативные исследования (Genome-Wide Association Studies, GWAS) [1], методы оценки полигенных рисков (Polygenic Risk Score, PRS) [2], подходы по выделению гаплотипов [3] и другие методы. Однако у каждого из методов есть свои особенности и ограничения. При вычислении по алгоритмам GWAS полиморфизм оценивают независимо, а затем производят фильтрацию по критерию значимости. Таким образом, результат любого заболевания, являющегося совокупностью редких наследуемых мутаций, с помощью данного подхода будет некорректным, так как максимально значимыми окажутся не относящиеся к фенотипу полиморфизмы или вообще ни один полиморфизм не преодолет критерии фильтрации. Подход PRS представляет собой расширение методологии GWAS, однако он тоже основан на предположении, что все используемые в модели однонуклеотидные варианты (single nucleotide polymorphisms, SNP) независимы. К тому же PRS-модели не воспроизводятся между популяциями при воспроизводимости локусов точек, на которых они построены, т. е. использования одиночных вариантов явно недостаточно для построения моделей и классификации пациентов. Кроме того, существует значительное количество заболеваний, генетическую структуру которых не удастся определить ни биологическими, ни популярными статистическими подходами. Чаще всего в основе алгоритма наследования таких фенотипов находятся гаплотипные структуры. Под гаплотипом подразумевается линейно сцепленная совокупность из нескольких (в том числе до нескольких сотен) вариабельных вариантов, которые совместно образуют небольшое (менее 100, в среднем 10–20) количество вариантов аллелей. Подход с выявлением конкретных вариантов гаплотипов активно используют в фармакогенетике при анализе цитохромов P450. Для CYP2D6, например, существует более 120 гаплотипных вариантов, которые обусловлены более чем 500 полиморфизмами [4]. Однако подобный подход очень редко обсуждают для большинства локусов полигенных заболеваний.

Особой популярностью GWAS-исследования пользуются при изучении заболеваний с вовлечением нервной системы, в том числе нарушений развития и нейродегенеративных патологий с полигенной природой, таких как боковой амиотрофический склероз, болезнь Паркинсона, шизофрения, патологии аутистического спектра. Они позволяют выявлять участки генома, изменения в которых перепредставлены у больных по сравнению с общей популяцией. Интересным свойством GWAS является возможность выявления локусов, патогенным вариантом в которых является не однонуклеотидный вариант из структуры чипа, а структурная вариация, напрямую чипом не детектируемая, но выявляемая по сцепленным с ней SNP. В частности, в GWAS бокового амиотрофического склероза детектируется локус в гене *C9orf72*, содержащий

увеличенный шестинуклеотидный повтор G4C2 (GGGGCC) [5], при этом варианты повтора непосредственно чипом не детектируются. В GWAS болезни Хантингтона выявляется локус гена *HTT* в 15-й хромосоме, содержащий тринуклеотидные повторы [6], при этом в чипе нет проб, соответствующих непосредственно участку повтора.

Эндотелиальная дистрофия роговицы Фукса (ДФ) — наследственное заболевание глаз, характеризующееся уменьшением количества эндотелиальных клеток роговицы, поддерживающих водный баланс стромы роговицы. ДФ является полигенным заболеванием и представляет значительный интерес для исследования генетики [7]. Существуют две формы ДФ: ранняя и поздняя. Они различаются генетическими причинами. Ранняя форма диагностируется в возрасте до 50 лет, встречается очень редко и связана с патогенными вариантами гена *COL8A2* [8]. Поздний дебют ДФ проявляется в возрасте старше 50 лет и является наиболее распространенной формой ДФ. Было показано, что поздний дебют ДФ ассоциирован с экспансией тринуклеотидных повторов CTG18.1 в интроне гена *TCF4* [9]. По нашим данным и данным зарубежных авторов, экспансия тринуклеотидных повторов CTG18.1 в гене *TCF4* является наиболее частым вариантом, ассоциированным с ДФ, среди популяций европейского происхождения. Экспансия по крайней мере одного аллеля тринуклеотидного повтора CTG18.1 была обнаружена примерно у двух третей пациентов с ДФ в когорте европейского происхождения. Позднее Афшари и др. [10] предприняли попытку с помощью GWAS в большой когорте найти другие варианты, связанные с ДФ. Они подтвердили ассоциацию установленного локуса *TCF4*, а также идентифицировали три новых локуса в генах *KANK4*, *LAMC1* и вблизи гена *ATP1B1*, однако их независимость от экспансии тринуклеотидного повтора не была проверена [10]. Обсуждается также роль мутаций в генах *ZEB1* [11], *SLC4A11* [12], *AGBL1* [13] и *LOXHD1* [14] как каузальных и независимых от экспансии повторов для реализации ДФ. Вопрос о том, является ли ДФ набором фенкопий или полигенным заболеванием, до сих пор остается открытым. Но известное бессимптомное носительство экспансии повторов [9, 15] и спорность четкой моногенной связи ДФ с некоторыми другими генами позволяют предположить, что ДФ с поздним началом — это набор полигенных фенкопий, что делает ее схожей с другими болезнями экспансий с поздней манифестацией.

В связи с этим представляется интересным вопрос — можно ли и с какой точностью разделить пациентов на группы внутри локусов по результатам GWAS. И если речь идет о ДФ позднего начала, можно ли по данным микрочипового анализа определить пациентов с фенкопией с экспансией и без нее? Сравнимы ли для этих целей использование гаплотипного подхода/комбинаций вариантов и носительство минорных аллелей отдельных генетических вариантов? Целью данной работы было разработать и проверить подход разделения пациентов на группы на основе результатов чип-генотипирования и полногеномного ассоциативного исследования (GWAS).

ПАЦИЕНТЫ И МЕТОДЫ

Изложенный в статье анализ был выполнен на образцах следующих двух выборок базы данных dbGaP: мультицентрового исследования по изучению дистрофии Фукса (FECG Genetics Multi-center Study) [16] и исследования возрастных заболеваний глаз

(Age-Related Eye Disease Study AREDS, Refractive Error Substudy) [17–18] с использованием чипов Illumina HumanOmni2.5-4v1. Клинические проявления заболевания были классифицированы с использованием модифицированной оценочной шкалы Krachmer J., основанной на результатах биомикроскопии щелевой лампой [19].

Контроль качества данных осуществляли на уровне образцов и на уровне вариантов. Для подготовки данных генотипирования использовали программное обеспечение PLINK 1.9 [20], GRAF 2.4 [21–22], а также собственный программный код, написанный на языке R версии 4.1.0.

Для контроля качества вариантов на первом шаге были удалены генотипы с низкими показателями качества прочтения GenCall (GC) путем удаления вариантов с $GC < 0,3$. В анализ были включены маркеры, удовлетворяющие следующим критериям: доля пропущенных генотипов $< 10\%$, частота минорного аллеля $> 1\%$, число ошибок Менделевского наследования < 2 , p -value точного теста на равновесие Харди–Вайнберга $> 1 \times 10^{-10}$ для контрольных образцов и p -value $> 1 \times 10^{-15}$ — для пациентов с ДФ. Маркеры, являющиеся дубликатными, т. е. маркеры с различными ID, но идентичными геномными позициями и кодировкой аллелей, были детектированы и проанализированы отдельно. Пары дубликатных маркеров с дискордантными генотипами более чем в 10 образцах были полностью удалены из рассмотрения. Из каждой пары дубликатов с 10 отличиями и менее было исключено по одному маркеру с наименьшим показателем доли пропущенных генотипов (genotype missing rate). После применения всех фильтров в анализе осталось 1580746 SNP-маркеров.

На основе метаданных были сформулированы следующие критерии включения в группу пациентов с ДФ: возраст 47 лет и старше; кератопластика не менее, чем на одном глазу, или наличие 2-й или выше степени развития заболевания по модифицированной шкале Krachmer J., манифестировавшего не менее чем на одном глазу.

Критерии включения в группу контроля: возраст 60 лет и старше; нормальное развитие роговицы глаза (без зафиксированных в метаданных аномалий в состояниях эпителия, эндотелия и стромы), за исключением травм роговицы.

Критерии исключения: образцы с ошибками менделевского наследования; образцы с несоответствиями между заявленным полом и полом, установленным на основе генетических данных (по гетерозиготности X-хромосомы с одновременным обнаружением вариантов Y-хромосомы по данным PLINK); образцы с отсутствием более 5% генотипов; родственные образцы до второй линии родства включительно (по данным GRAF-rel).

С целью получения генетически однородной выборки в популяционном приближении с использованием GRAF-pop была проведена оценка популяции образцов. Были отфильтрованы образцы, являющиеся выбросами в пространстве генетических координат. Участники исследований были рекрутированы на базе нескольких научных центров, использующих разные критерии включения участников в выборку пациентов с ДФ и в контрольную выборку. Чтобы унифицировать данные критерии, мы дополнительно провели фильтрацию образцов по фенотипическим признакам, руководствуясь единой для всех образцов логикой формирования групп сравнения.

Далее разбиение пациентов на подгруппы с предпологаемым наличием или отсутствием экспансии повторов осуществляли в несколько этапов:

1-й этап: ранжирование и отсечение наиболее значимых вариантов локуса;

2-й этап: кластеризация участников исследования по гаплотипам/комбинациям отобранных вариантов;

3-й этап: проверка соответствия полученного результата доле вероятного носительства экспансии повторов по данным, известным из публикаций. Аллель повторов считали экспансированной, если число повторов составляло более 40 копий, и не экспансированной, если число повторов было менее 40.

Для реализации первого этапа был проведен полногеномный поиск ассоциаций с использованием модели логистической регрессии с добавлением пола и первых шести главных компонент в качестве ковариат. Значения p -value были скорректированы на множественное тестирование методом Бенджамини–Хохберга по FDR. Набор вариантов из 18 хромосомы (где расположен локус с повторами) сначала был отфильтрован по p -value $< 1 \times 10^{-15}$. Для сравнения с гаплотипным подходом из полученного набора вариантов были рассмотрены три SNP с самыми низкими значениями p -value в качестве возможных маркеров увеличенного числа повторов. Затем все варианты локуса были проведены через процедуру прунинга — сокращения числа анализируемых вариантов путем удаления высокоррелированных вариантов с попарной корреляцией: $r^2 > 0,6$. Матрица генотипов перед дальнейшим анализом была закодирована в соответствии с доминантной моделью наследования.

На втором этапе использовали предположение, что пациенты с экспансией повторов в гене TCF4 будут обладать определенным набором SNP. Мы ожидали, что образцы ДФ из анализируемого сета сгруппируются в локусе гена TCF4 на основании гаплотипов и комбинаций одиночных вариантов. При этом люди с фенокопией, обусловленной экспансией, на основании сходства комбинаций минорных вариантов попадут в общие кластеры. В эти же кластеры попадут люди из контрольной выборки, которые являются бессимптомными носителями повторов (2–10%), а также небольшая часть контрольной выборки, несущая минорные гаплотипы, не содержащие повторов. При этом комбинации мажорных вариантов и гаплотипы с преимущественно мажорными вариантами сформируют кластеры преимущественно из представителей контрольной выборки. Но сюда попадет и часть пациентов с ДФ с фенокопией без экспансии повторов, и часть пациентов с экспансией, более не сцепленной с минорными гаплотипами (у 7% пациентов с ДФ частично гаплотипы расцепляются с повторами, что ранее было показано для варианта rs613872 [23]). Поэтому доля пациентов с ДФ и участников без ДФ может быть использована как суррогатный маркер носительства повторов в конкретных кластерах.

Для разделения пациентов на группы к выборке был применен метод иерархической агломерационной кластеризации, реализованный в функции hclust пакета stats. Данный алгоритм выстраивает данные в древовидную структуру, последовательно объединяя пары кластеров с наименьшим расстоянием до тех пор, пока все элементы не будут объединены в один кластер. Алгоритм принимает на вход матрицу попарных расстояний между точками (образцами), на начальном этапе каждая точка представляет отдельный кластер. Из-за неидентичности гаплотипов мы ожидаем больше двух кластеров, а оптимальное их количество определяется метрикой Silhouette.

Таблица 1. Характеристика участников исследования

Выборка	Пациенты с ДФ	Контрольные образцы	
	По данным: Afshari et al, 2017	По данным: Afshari et al, 2017	По данным: ARED
Число участников	1287	2373	
		562	1811
Мужчины	408	989	
		245	744
Женщины	879	1384	
		317	1067
Медианный возраст	71	72	68
		69	

Для каждого кластера были определены доли пациентов и контрольных образцов. Кластеры с преобладанием доли пациентов мы считали ассоциированными с экспансией повторов. Дополнительно мы визуализировали входные данные с помощью тепловых карт, отображая распределение образцов в каждом кластере в качестве графических аннотаций.

Для подобранных трех SNP носительство минорного аллеля по группе контролей считали маркером носительства увеличенного числа повторов.

Для подтверждения правильности полученного разбиения было вычислено отношение шансов обнаружить экспансию в группе пациентов с ДФ по сравнению с группой контроля. Для отношения шансов был дополнительно рассчитан 95%-й доверительный интервал.

Для сравнения с экспериментальными данными нами был проведен подбор исследований по следующим критериям.

В исследовании методом фрагментного анализа или ПЦР с праймированием триплетов определяли число повторов в интроне гена *TCF4*.

Исследовали выборку людей европейского происхождения.

Число исследованных людей в группе больных ДФ или в группе контроля — не менее 50 человек.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

После реализации процедуры контроля качества была сформирована комбинированная рабочая выборка из 3660 образцов европейского происхождения (табл. 1).

После применения всех фильтров в анализе осталось 1 580 746 SNP-маркеров.

Так как GWAS был проведен на тех же наборах данных, что были использованы в исследовании Afshari et al, 2017 [10], его результаты (рис. 1) сопоставимы с результатами, описанными в статье. Коэффициент генетической инфляции составил 1.05, что свидетельствует о незначительном расслоении населения.

Для дальнейшего анализа рассматривали только локус 18-й хромосомы. В результате фильтрации вариантов по значениям p -value осталось 134 варианта, из которых три имеют наиболее низкие значения p -value: rs784257, rs72932578 и rs618869 (и частоты минорных аллелей С, Т и С в европейской популяции по данным проекта gnomAD v3.1.2, равны 0,17932, 0,05649 и 0,13451 соответственно) и были дополнительно проанализированы на предмет разделения пациентов на группы.

Размер гаплотипного блока составил 50 вариантов, оставшихся после процедуры прунинга. В результате кластеризации образцы были распределены по 10 подгруппам (рис. 2). Из них в трех подгруппах преобладали образцы пациентов с ДФ, в семи подгруппах преобладали образцы контрольной группы.

Результаты кластеризации показали, что кластеры с преобладанием людей контрольной группы гомогенны по их представленности. Однако три кластера с потенциально увеличенным числом повторов (с преобладанием пациентов с дистрофией Фукса) гетерогенны по гаплотипам, что отражено в неравномерности распределения людей с разными фенотипами в рамках каждого кластера. Это

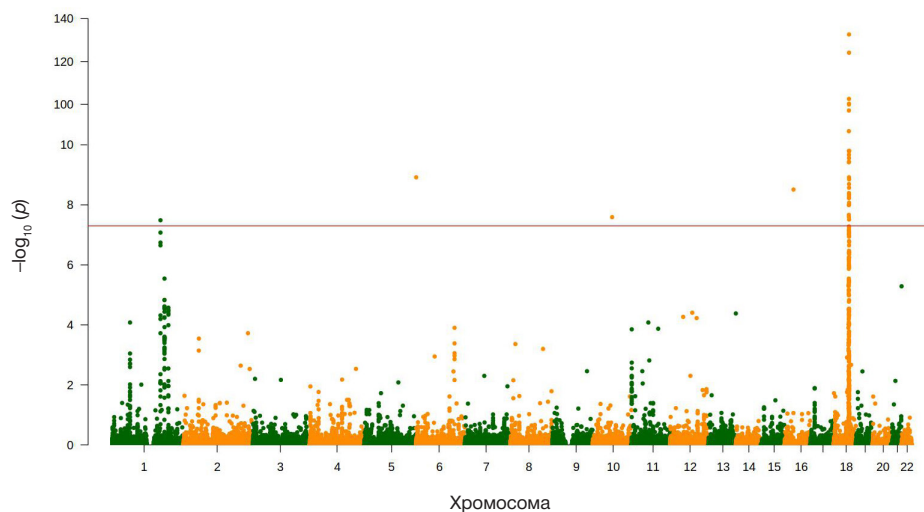


Рис. 1. Результаты полногеномного поиска ассоциаций, представленные с помощью Манхэттенского графика. Точки на графике представляют собой анализируемые варианты, их позиции по оси x определяются расположением вариантов в геноме, позиции по оси y отражают степень ассоциации с заболеванием, выраженную величиной $-\log_{10}(p\text{-value})$

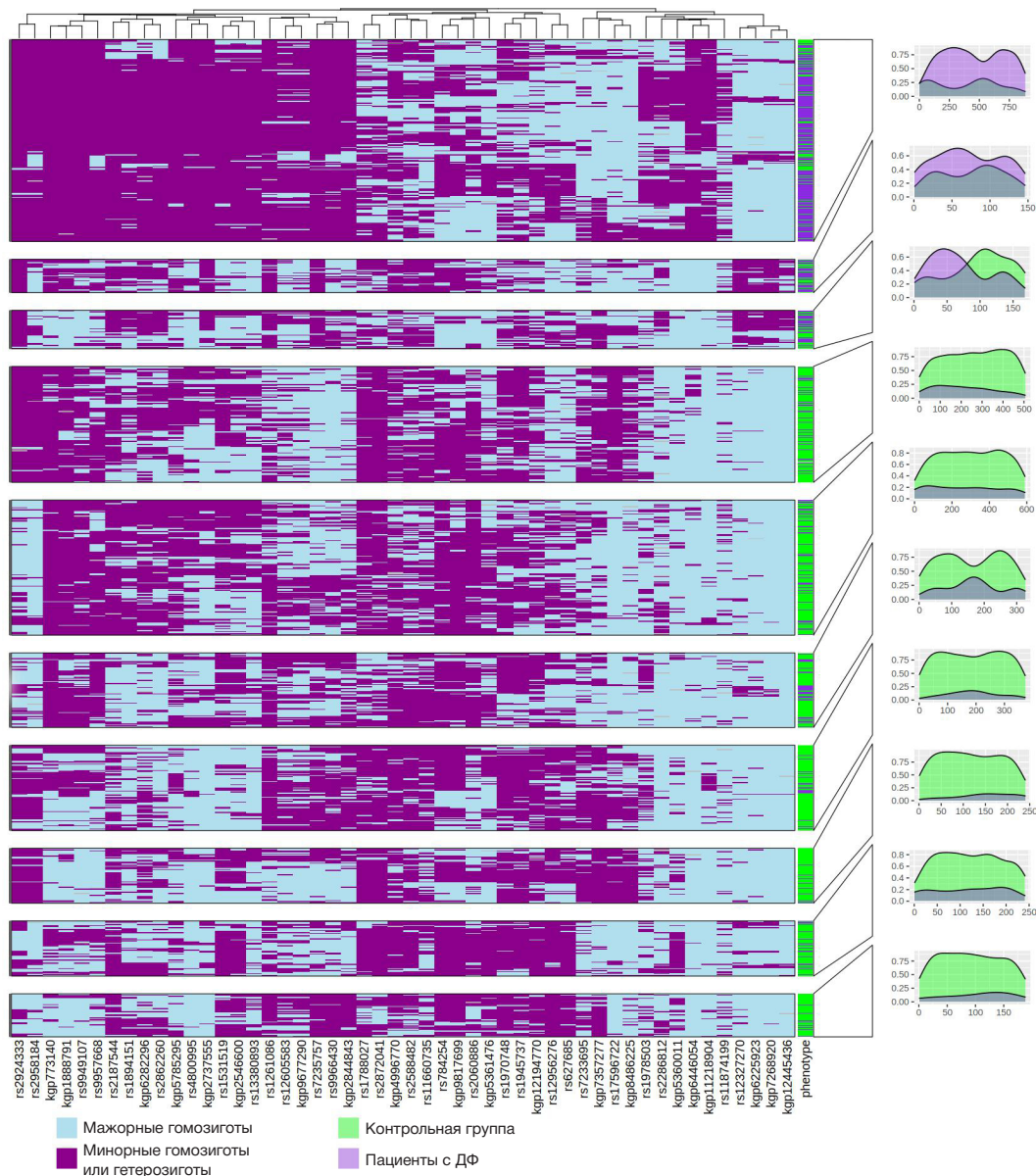


Рис. 2. Тепловая карта (хитмап) кластеризации пациентов. По вертикали обозначены пациенты, по горизонтали — различные геномные варианты. Аннотация справа отражает распределение образцов с ДФ и образцов контрольной группы в пределах каждого кластера

может быть связано как с бессимптомным носительством увеличенного числа повторов в данном локусе, так и с разрешающей способностью популяционных вариантов чипа, недостаточной для аккуратного разделения образцов по повтарам разной длины.

Наш анализ показал, что доля людей из кластеров с предположительным носительством экспансированных повторов в группе образцов с ДФ значительно выше, чем

в контрольной группе (табл. 2). При этом рассчитанная доля людей с предположительным носительством экспансии повторов сильно варьирует в зависимости от выбора метода — предсказание экспансии по гаплотипам/комбинациям вариантов или по генотипам отдельных вариантов с низкими показателями *p*-value).

Для проверки полученных результатов нами был проведен подбор исследований с результатами

Таблица 2. Распределение людей с предположенным статусом повторов по группам сравнения

Маркер увеличенного числа повторов	Пациенты с ДФ			Контрольная выборка		
	С предсказанной экспансией повторов	С предсказанным отсутствием экспансии повторов	Доля предсказанных носителей экспансии повторов, %	С предсказанной экспансией повторов	С предсказанным отсутствием экспансии повторов	Доля предсказанных носителей экспансии повторов, %
Группа кластеров	764	523	59,4	264	2109	11,1
rs784257	1046	237	81,5	765	1602	32,3
rs72932578	698	583	54,5	286	2082	12,1
rs618869	852	431	66,4	588	1780	24,8

Таблица 3. Результаты экспериментальных исследований доли людей с экспансией повторов в *TCF4* среди пациентов с ДФ и контрольных образцов европейского происхождения

	Страна	ДФ		Контрольная группа	
		Всего	С экспансией повторов	Всего	С экспансией повторов
Skorodumova et al, 2018 [15]	Россия	100	72	100	5
Viberg et al., 2022 [24]	Швеция	85	76	102	4
Foja et al., 2017 [25]	Германия	61	47	113	13
Kuot et al., 2017 [26]	Австралия	189	107	183	9

экспериментального определения экспансии повторов. Традиционно число повторов определяют каноническим фрагментным анализом или ПЦР с праймированием триплетов с последующим фрагментным анализом. Было подобрано пять публикаций с подходящими выборками (табл. 3).

Для сравнения предсказанной частоты носительства экспансии с частотой носительства экспансии в исследованиях было проведено объединение выборок из публикаций. Сравнительный анализ показал, что маркеры в разной степени воспроизводят частоту носительства экспансии повторов в группах сравнения (рис. 3).

Ни один из использованных подходов не отражает с достаточной точностью частоту повторов в группе пациентов и контрольной группе по сравнению с результатами прямого типирования из статей (табл. 4). Однако гаплотипный подход превзошел анализируемые одиночные варианты по параметру отношения шансов, перекрывая 95% доверительный интервал выборки двух исследований.

Интересно отметить, что рассмотренные нами одиночные варианты дают сильно дискордантные между собой результаты (рис. 4), т. е. во многом разные люди являются носителями минорных аллелей в данных вариантах, что делает неустойчивой используемую метрику. rs784257 сильнее всего отличается по носительству аллелей от гаплотипного подхода и одновременно является наиболее значимым по результатам GWAS. При этом он показывает максимальное несоответствие по долям потенциального носительства экспансии в группе контроля и не лучшее соответствие с группой ДФ. Скорее всего данный аллель менее сцеплен с носительством повторов, чем два других.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Молекулярно-генетическая классификация пациентов с полигенными заболеваниями является полезным инструментом для изучения генетической природы заболеваний. При этом в рамках одного фенотипа могут существовать пациенты с группами каузальных вариантов, сцепленных с разными гаплотипами. И хотя ген явно связан с заболеванием, *p*-value вариантов будет выше из-за обилия групп сцепленных вариантов, т. е. из-за особенностей генетической структуры заболевания варианты, значимые для каждой из групп в отдельности, не преодолевают общепринятый критерий значимости (*p*-value $< 5 \times 10^{-8}$). К тому же для многих локусов, приближающихся по значимости к общепринятому порогу, характерна заметная разреженность по вариантам в области низких значений *p*-value, при которой только единичные варианты показывают сильные ассоциации с заболеванием. Таким образом, без дополнительного анализа не представляется возможным выбрать между отнесением геномных вариантов к популяционным выбросам (значимость которых обусловлена случайным популяционным смещением частот) или к потенциально значимой группе вариантов. Поэтому методы структуризации генетических данных, включающие анализ взаимодействий как вариантов в блоках гаплотипов, так и гаплотипных блоков между собой, являются перспективным инструментом для уточнения генетики заболевания.

GWAS позволяет получить больше информации о генетической структуре заболевания, чем исключение вариантов по *p*-value и формирование локусов с привязкой к ближайшему гену и по сути переходом с

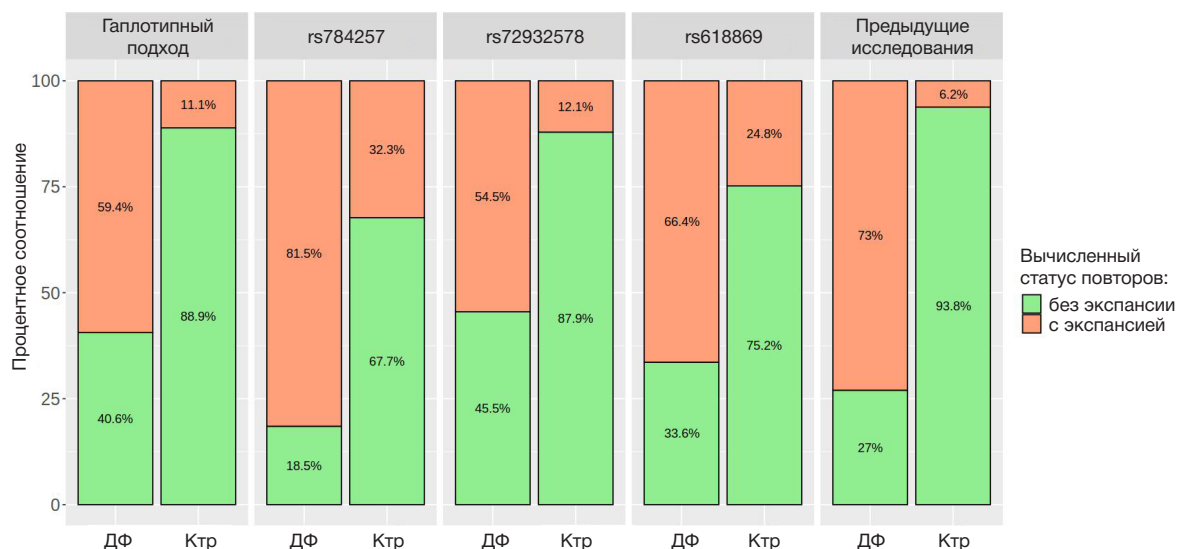


Рис. 3. Доля людей с экспансией и без экспансии повторов в интроне гена *TCF4* по данным нашего исследования (гаплотипный подход, носительство минорных аллелей вариантов rs784257, rs72932578, rs618869) и других исследований. ДФ — люди с дистрофией Фукса, Ктр — группа контроля

Таблица 4. Отношение шансов обнаружить экспансию в группе пациентов с ДФ по сравнению с группой контроля

	Отношение шансов	95%-й доверительный интервал
Гаплотипный подход (h-clust)	11,67	9,85–13,83
rs784257	9,24	7,83–10,90
rs72932578	8,72	7,39–10,28
rs618869	5,98	5,16–6,94
Skorodumova et al., 2018 [15]	48,86	17,98–132,76
Viberg et al., 2022 [24]	206,89	61,37–697,50
Foja et al., 2017 [25]	25,82	11,25–59,26
Kuot et al., 2017 [26]	25,23	12,17–52,31

уровня «вариант» на уровень «ген». При этом остаются вопросы о неравноценности вклада разных локусов в генетическую природу заболевания у конкретных групп людей с одним фенотипом. Это связано в том числе с отсутствием развитых подходов к формированию комбинаций вариантов, т. е. к работе на промежуточном уровне между уровнями «ген» и «вариант». Из-за неполной сцепленности вариантов между собой целесообразно рассматривать совокупности гаплотипов/комбинаций вариантов, определяющие дифференцированный риск заболевания, а не конкретные рисковые или протективные гаплотипы в отдельности, т. е. вариант с самым высоким популяционным атрибутивным риском (сочетание частоты аллелей и относительного риска), вероятно, и будет самым значимым в локусе.

До сих пор анализ групп гаплотипов, сцепленных с каузальными вариантами, является сложной задачей, однако все чаще превосходит методологию GWAS, даже несмотря на отсутствие высокопроизводительного единого подхода. В 2005 г. в результате GWAS было показано, что ген *CFH* ассоциирован с возрастной макулодистрофией сетчатки [27]. А позже появилось сообщение, что эта ассоциация не ограничивается одиночными вариантами, и видна в том числе в группах пациентов со структурными перестройками, в частности с частичными делециями генов *CFHR1-5* [27]. Более того, было выявлено, что большая часть вариабельности, которую приписывали одиночным вариантам, на самом деле являлась маркером гаплотипов с крупными структурными перестройками в этом регионе. И именно гаплотипные варианты структуры локуса, в том числе с разной популяционной представленностью, уже имеют намного более сильную связь с риском дегенерации сетчатки, чем большинство одиночных вариантов в этом локусе [28].

В данной работе мы реализовали кластеризацию образцов по вариантам региона с экспансией, опираясь на данные об ассоциации с повторами отдельных вариантов [14, 23, 29], в частности, аллеля G варианта rs613872, и гаплотипных блоков [29]. В результате кластеризации образцы из группы пациентов с ДФ и контрольной группы оказались неравномерно распределены по кластерам, что является косвенным свидетельством кластеризации по гаплотипам, сцепленным с экспансией. Все кластеры за исключением одного (кластера № 3), обладали однозначным статусом. Неопределенность при выявлении статуса была обусловлена равным соотношением в кластере долей пациентов и лиц контроля. В дальнейшем предстоит принять решение, что делать с такими кластерами — перекластеризовать людей из них отдельно или оставлять с неопределенным статусом. Необходимо также подобрать другую метрику качества кластеризации, а для этого нужны дополнительные данные, содержащие одновременно и результаты генотипирования образцов, и информацию о длине повторов. Несмотря на данные ограничения, результаты, полученные с использованием гаплотипного подхода, превзошли результаты, показанные одиночными вариантами, однако оказались недостаточно точными, чтобы считать использованную нами реализацию оптимальной.

Данная работа позволила реализовать две цели.

1. Первичное тестирование подхода, позволяющего на промежуточном уровне (не уровне одиночного варианта и не уровне гена, ближайшего к локусу) выполнить стратификацию пациентов и групп контроля без предварительного понимания структуры гаплотипов локуса. В качестве меры использована доля пациентов с ДФ и контрольных образцов в кластерах, что позволяет использовать данный подход для заболеваний, у которых

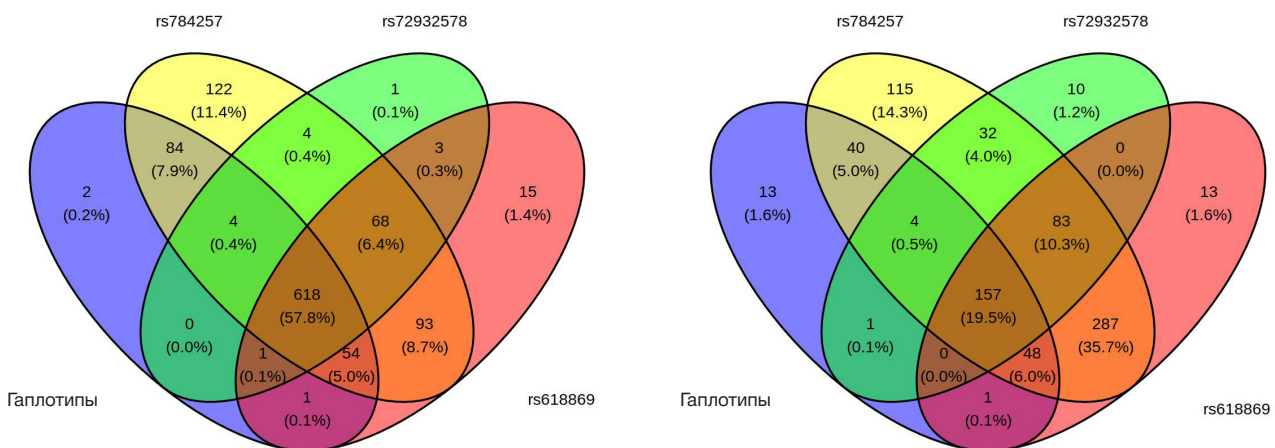


Рис. 4. Согласованность результатов определения статуса повторов с использованием разных подходов — по генотипам вариантов rs784257, rs72932578, rs618869 и с учетом гаплотипов. Слева — для носителей экспансии повторов в группе ДФ, справа — для носителей экспансии повторов в группе контроля

заведомо неизвестна ориентировочная доля людей с фенотипом, тесно связанная с данным локусом или обусловленная изменениями в нем.

2. Получить подвыборку пациентов с ДФ без носительства экспансии для прицельного переанализа GWAS с целью прояснения генетической структуры именно этой категории пациентов.

В будущем, кластеризация пациентов позволит не только выделять группы внутри фенотипа с сильным генетическим вкладом отдельных генетических вариантов, в том числе структурных, но и предлагать их обоснование и подходы к предсказанию ответа пациентов к разным видам терапии.

ВЫВОДЫ

В данном исследовании была продемонстрирована возможность применения подхода, основанного на гаплотипах, к молекулярно-генетической классификации

пациентов по причинам генетического заболевания, в данном случае, по наличию экспансии повторов. Результаты работы позволили нам сформулировать следующие выводы: 1) гаплотипный подход лучше, чем отдельные варианты, подходит для выявления связи локусов с отдельными группами пациентов; 2) для получения более точной картины следует пересмотреть подход к определению гаплотипного состава и моделированию матрицы данных для кластеризации. В частности планируется проанализировать ряд методов по вычислению генетического сходства (генетических расстояний) между образцами, а также применить более специфичные методы для начального отбора вариантов; 3) полученные результаты указывают на то, что кластеризация разделяет пациентов с ДФ и группу контроля с учетом групп гаплотипов/комбинаций вариантов, ассоциированных с экспансией повторов. Для дальнейших этапов тестирования подхода необходима дополнительная доказательная база, требующая подключения большего объема валидационных данных.

Литература

- Uffelmann E, Huang QQ, Munung NS, De Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. *Nature Reviews Methods Primers*. 2021; 1 (1): 59. Available from: <https://doi.org/10.1038/s43586-021-00056-9>.
- Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome medicine*. 2020; 12 (1): 1–1. Available from: <https://doi.org/10.1186/s13073-020-00742-5>.
- Stram DO. Multi-SNP haplotype analysis methods for association analysis. *Statistical Human Genetics: Methods and Protocols*. 2017: 485–504. Available from: https://doi.org/10.1007/978-1-4939-7274-6_24.
- Twesigomwe D, Wright GE, Drögemöller BI, da Rocha J, Lombard Z, Hazelhurst S. A systematic comparison of pharmacogene star allele calling bioinformatics algorithms: a focus on CYP2D6 genotyping. *NPJ genomic medicine*. 2020; 5 (1): 30. Available from: <https://doi.org/10.1038/s41525-020-0135-2>.
- Van Rheenen W, Van Der Spek RA, Bakker MK, Van Vugt JJ, Hop PJ, Zwamborn RA, et al. Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nature genetics*. 2021; 53 (12): 1636–48.
- Lee JM, Wheeler VC, Chao MJ, Vonsattel JP, Pinto RM, Lucente D, et al. Identification of genetic factors that modify clinical onset of Huntington's disease. *Cell*. 2015; 162 (3): 516–26.
- Fautsch MP, Wieben ED, Baratz KH, Bhattacharyya N, Sadan AN, Hafford-Tear NJ, et al. TCF4-mediated Fuchs endothelial corneal dystrophy: Insights into a common trinucleotide repeat-associated disease. *Progress in retinal and eye research*. 2021; 81: 100883.
- Biswas S, Munier FL, Yardley J, Hart-Holden N, Perveen R, Cousin P, et al. Missense mutations in COL8A2, the gene encoding the $\alpha 2$ chain of type VIII collagen, cause two forms of corneal endothelial dystrophy. *Human molecular genetics*. 2001; 10 (21): 2415–23.
- Wieben ED, Aleff RA, Tosakulwong N, Butz ML, Highsmith WE, Edwards AO, et al. A common trinucleotide repeat expansion within the transcription factor 4 (TCF4, E2-2) gene predicts Fuchs corneal dystrophy. *PLoS One*. 2012; 7 (11): e49083.
- Afshari NA, Igo Jr RP, Morris NJ, Stambolian D, Sharma S, Pulagam VL, et al. Genome-wide association study identifies three novel loci in Fuchs endothelial corneal dystrophy. *Nature communications*. 2017; 8 (1): 14898.
- Chung DW, Frausto RF, Ann LB, Jang MS, Aldave AJ. Functional impact of ZEB1 mutations associated with posterior polymorphous and Fuchs' endothelial corneal dystrophies. *Investigative Ophthalmology & Visual Science*. 2014; 55 (10): 6159–66.
- Chaurasia S, Ramappa M, Annapurna M, Kannabiran C. Coexistence of congenital hereditary endothelial dystrophy and Fuchs endothelial corneal dystrophy associated with SLC4A11 mutations in affected families. *Cornea*. 2020; 39 (3): 354–7.
- Riazuddin SA, Vasanth S, Katsanis N, Gottsch JD. Mutations in AGBL1 cause dominant late-onset Fuchs corneal dystrophy and alter protein-protein interaction with TCF4. *The American Journal of Human Genetics*. 2013; 93 (4): 758–64.
- Riazuddin SA, Parker DS, McGlumphy EJ, Oh EC, Liff BW, Schmedt T, et al. Mutations in LOXHD1, a recessive-deafness locus, cause dominant late-onset Fuchs corneal dystrophy. *The American Journal of Human Genetics*. 2012; 90 (3): 533–9.
- Skorodumova LO, Belodedova AV, Antonova OP, Sharova EI, Akopian TA, Selezneva OV, et al. CTG18. 1 expansion is the best classifier of late-onset Fuchs' corneal dystrophy among 10 biomarkers in a cohort from the European part of Russia. *Investigative Ophthalmology & Visual Science*. 2018; 59 (11): 4748–54.
- Louttit MD, Kopplin LJ, Igo Jr RP, Fondran JR, Tagliaferri A, Bardenstein D, et al. A multi-center study to map genes for Fuchs' endothelial corneal dystrophy: baseline characteristics and heritability. *Cornea*. 2012; 31 (1): 26.
- Age-Related Eye Disease Study Research Group. The age-related eye disease study (AREDS): design implications AREDS report no. 1. *Controlled clinical trials*. 1999; 20 (6): 573.
- Stambolian D, Wojciechowski R, Oexle K, Pirastu M, Li X, Raffel LJ, et al. Meta-analysis of genome-wide association studies in five cohorts reveals common variants in RBFOX1, a regulator of tissue-specific splicing, associated with refractive error. *Human molecular genetics*. 2013; 22 (13): 2754–64.
- Krachmer JH, Purcell JJ Jr, Young CW, Bucher KD. Corneal endothelial dystrophy. A study of 64 families. *Arch Ophthalmol*. 1978; 96 (11): 2036–9. DOI: 10.1001/archoph.1978.03910060424004.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015; 4: 7. DOI: 10.1186/s13742-015-0047-8.
- Jin Y, Schäffer AA, Sherry ST, Feolo M. Quickly identifying identical and closely related subjects in large databases using genotype data. *PLoS One*. 2017; 12 (6): e0179106. DOI: 10.1371/journal.pone.0179106.
- Jin Y, Schaffer AA, Feolo M, Holmes JB, Kattman BL. GRAF-pop: A Fast Distance-Based Method To Infer Subject Ancestry from Multiple Genotype Datasets Without Principal Components

- Analysis. *G3 (Bethesda)*. 2019; 9 (8): 2447–61. DOI: 10.1534/g3.118.200925.
23. Okumura N, Hayashi R, Nakano M, Tashiro K, Yoshii K, Aleff R. et al. Association of rs613872 and Trinucleotide Repeat Expansion in the TCF4 Gene of German Patients With Fuchs Endothelial Corneal Dystrophy. *Cornea*. 2019; 38 (7): 799–805. DOI: 10.1097/ICO.0000000000001952.
 24. Viberg A, Westin IM, Golovleva I, Byström B. TCF4 trinucleotide repeat expansion in Swedish cases with Fuchs' endothelial corneal dystrophy. *Acta Ophthalmol*. 2022; 100 (5): 541–8. DOI: 10.1111/aos.15032. Epub 2021 Oct 13.
 25. Foja S, Luther M, Hoffmann K, Rupprecht A, Gruenauer-Kloevekorn C. CTG18.1 repeat expansion may reduce TCF4 gene expression in corneal endothelial cells of German patients with Fuchs' dystrophy. *Graefes Arch Clin Exp Ophthalmol*. 2017; 255 (8): 1621–31. DOI: 10.1007/s00417-017-3697-7. Epub 2017 Jun 12.
 26. Kuot A, Hewitt AW, Snibson GR, Souzeau E, Mills R, Craig JE, et al. TGC repeat expansion in the TCF4 gene increases the risk of Fuchs' endothelial corneal dystrophy in Australian cases. *PLoS One*. 2017; 12 (8): e0183719. DOI: 10.1371/journal.pone.0183719.
 27. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. *Science*. 2005; 308 (5720): 385–9. DOI: 10.1126/science.1109557. Epub 2005 Mar 10.
 28. Spencer KL, Hauser MA, Olson LM, Schmidt S, Scott WK, Gallins P, et al. Deletion of CFHR3 and CFHR1 genes in age-related macular degeneration. *Hum Mol Genet*. 2008; 17 (7): 971–7. DOI: 10.1093/hmg/ddm369. Epub 2007 Dec 15.
 29. Baratz KH, Tosakulwong N, Ryu E, Brown WL, Branham K, Chen W, et al. E2-2 protein and Fuchs's corneal dystrophy. *N Engl J Med*. 2010; 363 (11): 1016–24. DOI: 10.1056/NEJMoa1007064. Epub 2010 Aug 25.

References

1. Uffelmann E, Huang QQ, Munung NS, De Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. *Nature Reviews Methods Primers*. 2021; 1 (1): 59. Available from: <https://doi.org/10.1038/s43586-021-00056-9>.
2. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome medicine*. 2020; 12 (1): 1–1. Available from: <https://doi.org/10.1186/s13073-020-00742-5>.
3. Stram DO. Multi-SNP haplotype analysis methods for association analysis. *Statistical Human Genetics: Methods and Protocols*. 2017: 485–504. Available from: https://doi.org/10.1007/978-1-4939-7274-6_24.
4. Twesigomwe D, Wright GE, Drögemöller BI, da Rocha J, Lombard Z, Hazelhurst S. A systematic comparison of pharmacogenes star allele calling bioinformatics algorithms: a focus on CYP2D6 genotyping. *NPJ genomic medicine*. 2020; 5 (1): 30. Available from: <https://doi.org/10.1038/s41525-020-0135-2>.
5. Van Rheenen W, Van Der Spek RA, Bakker MK, Van Vugt JJ, Hop PJ, Zwamborn RA, et al. Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nature genetics*. 2021; 53 (12): 1636–48.
6. Lee JM, Wheeler VC, Chao MJ, Vonsattel JP, Pinto RM, Lucente D, et al. Identification of genetic factors that modify clinical onset of Huntington's disease. *Cell*. 2015; 162 (3): 516–26.
7. Fautsch MP, Wieben ED, Baratz KH, Bhattacharyya N, Sadan AN, Hafford-Tear NJ, et al. TCF4-mediated Fuchs endothelial corneal dystrophy: Insights into a common trinucleotide repeat-associated disease. *Progress in retinal and eye research*. 2021; 81: 100883.
8. Biswas S, Munier FL, Yardley J, Hart-Holden N, Perveen R, Cousin P, et al. Missense mutations in COL8A2, the gene encoding the $\alpha 2$ chain of type VIII collagen, cause two forms of corneal endothelial dystrophy. *Human molecular genetics*. 2001; 10 (21): 2415–23.
9. Wieben ED, Aleff RA, Tosakulwong N, Butz ML, Highsmith WE, Edwards AO, et al. A common trinucleotide repeat expansion within the transcription factor 4 (TCF4, E2-2) gene predicts Fuchs corneal dystrophy. *PLoS One*. 2012; 7 (11): e49083.
10. Afshari NA, Igo Jr RP, Morris NJ, Stambolian D, Sharma S, Pulagam VL, et al. Genome-wide association study identifies three novel loci in Fuchs endothelial corneal dystrophy. *Nature communications*. 2017; 8 (1): 14898.
11. Chung DW, Frausto RF, Ann LB, Jang MS, Aldave AJ. Functional impact of ZEB1 mutations associated with posterior polymorphous and Fuchs' endothelial corneal dystrophies. *Investigative Ophthalmology & Visual Science*. 2014; 55 (10): 6159–66.
12. Chaurasia S, Ramappa M, Annapurna M, Kannabiran C. Coexistence of congenital hereditary endothelial dystrophy and Fuchs endothelial corneal dystrophy associated with SLC4A11 mutations in affected families. *Cornea*. 2020; 39 (3): 354–7.
13. Riazuddin SA, Vasanth S, Katsanis N, Gottsch JD. Mutations in AGBL1 cause dominant late-onset Fuchs corneal dystrophy and alter protein-protein interaction with TCF4. *The American Journal of Human Genetics*. 2013; 93 (4): 758–64.
14. Riazuddin SA, Parker DS, McGlumphy EJ, Oh EC, Iliff BW, Schmedt T, et al. Mutations in LOXHD1, a recessive-deafness locus, cause dominant late-onset Fuchs corneal dystrophy. *The American Journal of Human Genetics*. 2012; 90 (3): 533–9.
15. Skorodumova LO, Belodedova AV, Antonova OP, Sharova EI, Akopian TA, Selezneva OV, et al. CTG18.1 expansion is the best classifier of late-onset Fuchs' corneal dystrophy among 10 biomarkers in a cohort from the European part of Russia. *Investigative Ophthalmology & Visual Science*. 2018; 59 (11): 4748–54.
16. Louttit MD, Kopplin LJ, Igo Jr RP, Fondran JR, Tagliaferri A, Bardenstein D, et al. A multi-center study to map genes for Fuchs' endothelial corneal dystrophy: baseline characteristics and heritability. *Cornea*. 2012; 31 (1): 26.
17. Age-Related Eye Disease Study Research Group. The age-related eye disease study (AREDS): design implications AREDS report no. 1. *Controlled clinical trials*. 1999; 20 (6): 573.
18. Stambolian D, Wojciechowski R, Oexle K, Pirastu M, Li X, Raffel LJ, et al. Meta-analysis of genome-wide association studies in five cohorts reveals common variants in RBFox1, a regulator of tissue-specific splicing, associated with refractive error. *Human molecular genetics*. 2013; 22 (13): 2754–64.
19. Krachmer JH, Purcell JJ Jr, Young CW, Bucher KD. Corneal endothelial dystrophy. A study of 64 families. *Arch Ophthalmol*. 1978; 96 (11): 2036–9. DOI: 10.1001/archophth.1978.03910060424004.
20. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015; 4: 7. DOI: 10.1186/s13742-015-0047-8.
21. Jin Y, Schäffer AA, Sherry ST, Feolo M. Quickly identifying identical and closely related subjects in large databases using genotype data. *PLoS One*. 2017; 12 (6): e0179106. DOI: 10.1371/journal.pone.0179106.
22. Jin Y, Schaffer AA, Feolo M, Holmes JB, Kattman BL. GRAF-pop: A Fast Distance-Based Method To Infer Subject Ancestry from Multiple Genotype Datasets Without Principal Components Analysis. *G3 (Bethesda)*. 2019; 9 (8): 2447–61. DOI: 10.1534/g3.118.200925.
23. Okumura N, Hayashi R, Nakano M, Tashiro K, Yoshii K, Aleff R. et al. Association of rs613872 and Trinucleotide Repeat Expansion in the TCF4 Gene of German Patients With Fuchs Endothelial Corneal Dystrophy. *Cornea*. 2019; 38 (7): 799–805. DOI: 10.1097/ICO.0000000000001952.
24. Viberg A, Westin IM, Golovleva I, Byström B. TCF4 trinucleotide repeat expansion in Swedish cases with Fuchs' endothelial corneal dystrophy. *Acta Ophthalmol*. 2022; 100 (5): 541–8. DOI: 10.1111/aos.15032. Epub 2021 Oct 13.
25. Foja S, Luther M, Hoffmann K, Rupprecht A, Gruenauer-

- Kloevekorn C. CTG18.1 repeat expansion may reduce TCF4 gene expression in corneal endothelial cells of German patients with Fuchs' dystrophy. *Graefes Arch Clin Exp Ophthalmol*. 2017; 255 (8): 1621–31. DOI: 10.1007/s00417-017-3697-7. Epub 2017 Jun 12.
26. Kuot A, Hewitt AW, Snibson GR, Souzeau E, Mills R, Craig JE, et al. TGC repeat expansion in the TCF4 gene increases the risk of Fuchs' endothelial corneal dystrophy in Australian cases. *PLoS One*. 2017; 12 (8): e0183719. DOI: 10.1371/journal.pone.0183719.
27. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. *Science*. 2005; 308 (5720): 385–9. DOI: 10.1126/science.1109557. Epub 2005 Mar 10.
28. Spencer KL, Hauser MA, Olson LM, Schmidt S, Scott WK, Gallins P, et al. Deletion of CFHR3 and CFHR1 genes in age-related macular degeneration. *Hum Mol Genet*. 2008; 17 (7): 971–7. DOI: 10.1093/hmg/ddm369. Epub 2007 Dec 15.
29. Baratz KH, Tosakulwong N, Ryu E, Brown WL, Branham K, Chen W, et al. E2-2 protein and Fuchs's corneal dystrophy. *N Engl J Med*. 2010; 363 (11): 1016–24. DOI: 10.1056/NEJMoa1007064. Epub 2010 Aug 25.