

## Полнофеномные исследования ассоциаций как исследовательский инструмент для выявления новых патогенетических связей

Машкина Н. А., Ершова А. И., Копылова О. В., Драпкина О. М.

ФГБУ "Национальный медицинский исследовательский центр терапии и профилактической медицины" Минздрава России.  
Москва, Россия

Накопление биообразцов, связанных с большими массивами клинических данных (данные электронных медицинских карт, эпидемиологических и других крупномасштабных исследований), сделало возможным проведение исследований, направленных на изучение ассоциаций между генетическими вариантами и фенотипами, в частности полнофеномных исследований ассоциаций (PheWAS), изначально разработанных для проверки одного или нескольких генетических вариантов, связанных с заболеванием или клиническим симптомом, на предмет ассоциаций с другими фенотипами. PheWAS позволяют выявлять новые генетические и фенотипические ассоциации, дифференцировать истинную плейотропию и клиническую коморбидность, выделять новые подтипы заболеваний, а также находить новые мишени для лекарств. Будущие усилия, направленные на интеграцию широкого и надежного сбора фенотипических данных, а также усовершенствование инструментов для проведения PheWAS создадут ценный ресурс для более эффективного анализа генома и фенома, что приведет к новым открытиям в персонализированной медицине.

**Ключевые слова:** полнофеномные исследования ассоциаций, геном, фенотип, феном, вариант нуклеотидной последовательности.

**Отношения и деятельность:** нет.

**Поступила** 12/11-2024

**Рецензия получена** 20/11-2024

**Принята к публикации** 27/12-2024



**Для цитирования:** Машкина Н. А., Ершова А. И., Копылова О. В., Драпкина О. М. Полнофеномные исследования ассоциаций как исследовательский инструмент для выявления новых патогенетических связей. *Кардиоваскулярная терапия и профилактика*. 2024;23(12):4275. doi: 10.15829/1728-8800-2024-4275. EDN XMWWLS



### Phenome-wide association studies as a research tool for identifying new pathogenetic links

Mashkina N. A., Ershova A. I., Kopylova O. V., Drapkina O. M.

National Medical Research Center for Therapy and Preventive Medicine. Moscow, Russia

The accumulation of biospecimens associated with large-scale clinical data (data from electronic medical records, epidemiological and other large-scale studies) has made it possible to conduct studies aimed at studying associations between genetic variants and phenotypes, in particular, phenome-wide association studies (PheWAS). It was originally designed to test one or more genetic variants associated with a disease or clinical symptom for associations with other phenotypes. PheWAS can identify novel genetic and phenotypic associations, differentiate true pleiotropy and clinical comorbidity, identify new disease subtypes, and identify new drug targets. Future efforts to integrate broad and robust phenotypic data collection and improve PheWAS tools will provide a valuable resource for more efficient genome-phenome analysis, leading to new discoveries in personalized medicine.

**Keywords:** Phenome-wide association studies, genome, phenotype, phenome, sequence variant.

**Relationships and Activities:** none.

Mashkina N. A. ORCID: 0000-0003-1489-9670, Ershova A. I. ORCID: 0000-0001-7989-0760, Kopylova O. V.\* ORCID: 0000-0001-5397-5387, Drapkina O. M. ORCID: 0000-0002-4453-8430.

\*Corresponding author: sivoksana@yandex.ru

**Received:** 12/11-2024

**Revision Received:** 20/11-2024

**Accepted:** 27/12-2024

**For citation:** Mashkina N. A., Ershova A. I., Kopylova O. V., Drapkina O. M. Phenome-wide association studies as a research tool for identifying new pathogenetic links. *Cardiovascular Therapy and Prevention*. 2024;23(12):4275. doi: 10.15829/1728-8800-2024-4275. EDN XMWWLS

\*Автор, ответственный за переписку (Corresponding author):

e-mail: sivoksana@yandex.ru

[Машкина Н. А. — аспирант, ORCID: 0000-0003-1489-9670, Ершова А. И. — д.м.н., зам. директора по фундаментальной науке, ORCID: 0000-0001-7989-0760, Копылова О. В.\* — к.м.н., с.н.с. лаборатории клинической, ORCID: 0000-0001-5397-5387, Драпкина О. М. — д.м.н., профессор, директор, руководитель отдела фундаментальных прикладных аспектов ожирения, ORCID: 0000-0002-4453-8430].

АГ — артериальная гипертензия, ВНП — вариант нуклеотидной последовательности, МКБ — международная классификация болезней, МР — менделевская рандомизация, СД — сахарный диабет, ССЗ — сердечно-сосудистые заболевания, ШГР — шкала генетического риска, ФП — фибрилляция предсердий, ЭМК — электронная медицинская карта, CLARITE — Cleaning to Analysis: Reproducibility-based Interface for Traits and Exposures (От очистки до анализа: интерфейс для признаков и воздействий на основе воспроизводимости), GWAS — Genome-wide association studies (полногеномные исследования ассоциаций), PheWAS — Phenome-wide association studies (полнофеномные исследования ассоциаций), UKBiobank — биобанк Великобритании.

### Ключевые моменты

#### Что известно о предмете исследования?

- Метод полнофеномных исследований ассоциаций (Phenome-wide association studies, PheWAS) стал устоявшимся и активно развивающимся подходом для выявления связи между генетическими вариантами нуклеотидной последовательности и широким спектром фенотипических признаков.
- На сегодняшний день PheWAS считается одним из наиболее ценных методологических подходов в изучении патогенетических взаимосвязей.

#### Что добавляют результаты исследования?

- Представлены методология и основные этапы проведения полнофеномных исследований ассоциаций.
- Обобщен широкий спектр научно-практических задач, которые могут быть решены с использованием PheWAS: репликация ранее выявленных ассоциаций, выявление новых ассоциаций генотип-фенотип и плейотропности вариантов нуклеотидной последовательности, изучение новых патогенетических связей и коморбидности, поиск новых подходов к лечению и профилактике, изучение безопасности лекарственных средств.

### Key messages

#### What is already known about the subject?

- Phenome-wide association studies (PheWAS) have become an established and actively developing approach to identify associations between genetic sequence variants and a wide range of phenotypic traits.
- Today, PheWAS is considered one of the most valuable methodological approaches in studying pathogenetic relationships.

#### What might this study add?

- The methodology and main stages of conducting phenome-wide association studies are presented.
- There are following scientific and practical problems that can be solved using PheWAS: replication of previously identified associations; identification of new genotype-phenotype associations and pleiotropy of nucleotide sequence variants; study of new pathogenetic relationships and comorbidity; search for new approaches to treatment and prevention; study of drug safety.

## Введение

Секвенирование дезоксирибонуклеиновой кислоты (ДНК) в больших когортах людей с созданием баз данных, содержащих качественную и обширную клиническую аннотацию, позволило перейти к исследованиям, направленным на изучение причинно-следственных связей в патогенезе развития заболеваний на качественно новый уровень. С момента своего появления в 2010г технология полнофеномных исследований ассоциаций (Phenome-wide association studies, PheWAS) стала устоявшимся и активно развивающимся подходом для выявления связи между генетическими вариантами нуклеотидной последовательности (ВНП) и широким спектром фенотипических признаков (включая молекулярные, биохимические, клеточные показатели, а также клинические диагнозы, синдромы, симптомы и исходы) [1, 2]. Если полногеномные исследования ассоциаций (Genome-wide association studies, GWAS) начинаются с выбора интересующего фенотипа и систематически анализируют ВНП по всему геному на предмет ассоциации с фенотипом (GWAS может идентифицировать

множественные генетические ассоциации с фенотипом в сложных или полигенных признаках), то PheWAS начинаются с интересующего ВНП (или сочетаний ВНП) и систематически анализируют множество фенотипов на предмет ассоциации с генотипом [3]. Совокупность фенотипов организма представляет собой феном [1].

Обычно PheWAS проводят на больших выборках, включающих тысячи, десятки и даже сотни тысяч участников, что позволяет выявить статистически значимые ассоциации. Для проведения PheWAS, помимо генетических данных, используется информация из электронных медицинских карт (ЭМК) или баз данных крупномасштабных исследований, которая группируется в исследуемые фенотипы. Появлению технологии PheWAS во многом способствовало развитие биобанкирования в мире, основой для многих опубликованных качественно проведенных PheWAS стали базы данных Биобанка Великобритании (UKBiobank) (~500 тыс. чел.), биобанков Сети электронных медицинских карт и геномики (Electronic Medical Records & Genomics, eMERGE Network) (~100 тыс. чел.) и др. [4].

PheWAS позволяют реплицировать ранее выявленные ассоциации генотип-фенотип, выявлять новые ассоциации генотип-фенотип, идентифицировать плейотропные эффекты ВНП или множественные независимые фенотипы, связанные с одним ВНП [1]. Идентификация генетической плейотропии дает существенное представление об общей генетической этиологии нескольких заболеваний. Сопряженные с менделевской рандомизацией (MR) PheWAS служат полезным инструментом для выявления потенциальных причинно-следственных связей между фенотипами [5]. На сегодняшний день PheWAS считается одним из наиболее ценных методологических подходов в изучении патогенетических взаимосвязей. Исследования методом PheWAS способствуют разработке персонализированных подходов к лечению и профилактике заболеваний, основывающихся на индивидуальных особенностях генотипа пациента [4].

Цель настоящего обзора — представить краткую информацию о методологии проведения PheWAS и пользе PheWAS для клинической и экспериментальной медицины.

## Методология исследования

Проведен поиск в электронных научных библиотеках и базах данных за период 2009-2024гг:

1. "elibrary.ru" по словам, словосочетаниям и их комбинациям (в названиях и ключевых словах публикаций): "Полнофеномные исследования ассоциаций";
2. "pubmed.gov" по словам, словосочетаниям и их комбинациям (в названиях и ключевых словах публикаций): "Phenome-wide association studies", "PheWAS".

## Методология PheWAS

Методология PheWAS с момента возникновения претерпела ряд изменений, что обусловлено развитием подходов к фенотипированию и возможностью включения в анализ редких ВНП, выявляемых при полноэкзомном или полногеномном секвенировании [4].

Проведение PheWAS включает в себя несколько этапов (рисунок 1) [5].

**Этап 1:** Контроль качества фенома.

Как и GWAS, PheWAS требуют применения поправок на множественное тестирование и обладают низкой мощностью при исследовании редких вариантов [6]. Дополнительная сложность анализа наборов данных по всему феному заключается в широте и гетерогенности фенотипов, что указывает на необходимость тщательного контроля качества их формирования. Для получения фенотипической информации широко используются ЭМК, в этом случае контроль качества направлен на корректное объединение кодов Международной классификации болезней (МКБ) в отдельные фенотипы. Опу-

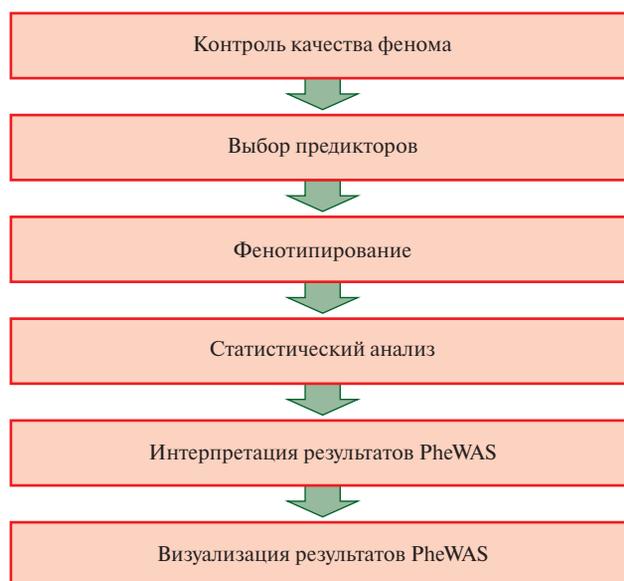


Рис. 1 Основные этапы проведения PheWAS.

Примечание: PheWAS — Phenome-wide association studies (полнофеномные исследования ассоциаций).

бликован ряд успешно выполненных PheWAS с использованием кодов МКБ [7, 8]. В то время как коды МКБ позволяют точно разделять фенотипы на случай-контроль, данные эпидемиологических исследований могут включать небинарные фенотипы. Такие данные должны быть глубже проанализированы, а фенотипы более четко идентифицированы по типу переменной, чтобы обеспечить корректность применения аналитических методов, таких как логистическая или линейная регрессия [9].

Корректность классификации фенотипов является лишь одним из аспектов контроля качества, другим важным аспектом формирования базы данных для проведения PheWAS является корректность и чистота данных. Выявление пропусков данных, поиск и исправление ошибок в первичном наборе информации (например, ошибки ручного ввода (опечатки) или некорректные значения с измерительных приборов из-за кратковременных сбоев) являются важными составляющими для формирования высококачественного фенома [9].

Базы данных для PheWAS имеют различные размеры выборок для количественных признаков и различное количество случаев и контролей для бинарных признаков по многим интересующим фенотипам, что может повлиять на статистическую мощность обнаружения ассоциаций [10]. Оценка минимального размера выборки для исследуемых фенотипов повышает статистическую мощность PheWAS, а исключение из анализа фенотипов с низкой изменчивостью уменьшает эффект множественных сравнений [9]. Для бинарных признаков Verma A, et al. (2018) показали, что общее количество включенных в исследование случаев более

важно с точки зрения мощности, чем соотношение между группами "случай" и "контроль". Для частых ВНП (с частотой минорного аллеля  $>0,05$ ) с пенетрантностью  $\geq 15\%$  достаточно 400 участников в группе контроля и 200 участников в группе "случай", для получения практически 100%-й мощности исследования. Отметим, что при более низких пенетрантности или частоте аллеля число участников резко возрастает [10].

Многомерность и большой размер феномных баз данных делают ручную проверку данных неэффективной. Кроме того, для повторения и воспроизведения результатов ранее проведенных PheWAS требуется прозрачность рабочих процессов контроля качества. Passero K, et al. в работе 2020г представили программное обеспечение CLARITE (CLEaning to Analysis: Reproducibility-based Interface for Traits and Exposures; От очистки до анализа: интерфейс для признаков и воздействий на основе воспроизводимости) для проведения воспроизводимого контроля качества фенотипов (как категориальных, так и количественных признаков), которое может быть адаптировано для применения к другим большим наборам данных с гетерогенными фенотипами [9]. CLARITE представляет собой инструмент для предварительной обработки гетерогенных и полученных из разных источников данных (биомаркеры, клинические параметры, результаты инструментальных исследований, прогноз и т.д.), в частности для очистки данных с широкими возможностями фильтрации и классификации фенотипов. Одним из достоинств CLARITE является хороший пользовательский интерфейс, предоставляющий возможности для установления ответственности и сохранения рабочих записей о принятии решений в отношении контроля качества [9].

#### Этап 2: Выбор предикторов

В качестве предикторов в PheWAS могут использоваться как отдельные ВНП, так и их сочетания. Используя ВНП, полученные из предыдущих GWAS, PheWAS позволяют обнаружить новые ассоциации или реплицируют известные ассоциации одного или нескольких ВНП с различными фенотипами. Хотя изначально PheWAS исследовали ассоциации между отдельными ВНП и множественными фенотипами, любая независимая переменная, включая лабораторные параметры, биомаркеры или даже заболевание или симптом, представляющие интерес, могут служить отправной точкой для проведения PheWAS. Например, хотя изменчивость экспрессии генов обычно не фиксируется в ЭМК или других больших наборах данных, можно использовать комбинации ВНП для прогнозирования экспрессии генов, а затем использовать PheWAS для поиска ассоциаций с прогнозируемой экспрессией генов по всем фенотипам человека [11].

#### Этап 3: Фенотипирование

Фенотипирование заключается в формировании набора исследуемых фенотипов, совокупность которых составляет феном индивидуума. Для PheWAS на основе ЭМК наиболее простым способом фенотипирования является использование системы кодирования МКБ. Для классификации кодов МКБ широко используются два подхода: организованное фенотипирование и целостное фенотипирование. Организованное фенотипирование представляет собой группирование кодов МКБ по общему этиологическому признаку в один фенотип, что приводит к меньшему количеству фенотипов с большим количеством случаев, а следовательно, к увеличению статистической мощности и снижению вероятности ложноположительных результатов [5]. Wu P, et al. (2019) приняли стратегию организованного фенотипирования и объединили  $>75\%$  кодов МКБ-10 в систему фекодов (PheCODE), состоящую из 1800 различных фекодов. Используя эту схему, они успешно воспроизвели несколько известных ассоциаций генотип-фенотип с сопоставимой статистической мощностью [8]. Напротив, целостное фенотипирование проверяет все коды МКБ и приводит к большему количеству фенотипов при относительно небольшом количестве случаев. Этот метод может привести к снижению мощности исследований и смещению результатов из-за корреляций между фенотипами [12].

Феномы также могут быть сформированы с использованием данных, собранных в рамках эпидемиологических исследований. С ростом числа крупномасштабных исследований и увеличением доступности больших массивов данных, в последние годы возросла возможность исследований фенотипов с помощью PheWAS за пределами бинарных клинических показателей. Подобные наборы данных могут включать лабораторные биомаркеры, антропометрические данные и многие другие фенотипы (например, социально-экономические факторы). Кроме того, указанные фенотипы включают в себя в основном количественные признаки, которые способствуют выявлению более ранних проявлений заболеваний (субклинических фенотипов) и повышению статистической мощности PheWAS. Первые PheWAS, основанные на эпидемиологических данных, были проведены с использованием сети PAGE (Population Architecture using Genomics and Epidemiology), аккумулирующей данные 8-и крупных популяционных исследований. В данном исследовании связанные фенотипы объединяли в единый класс, в итоге было идентифицировано 111 значимых ассоциаций генотип-фенотип [13]. Объединение данных из разных эпидемиологических исследований выгодно для получения большего размера выборки и увеличения статистической мощности. Однако этот подход имеет огра-

ничения, обусловленные сложностью согласования фенотипов между различными исследованиями из-за различий в определениях фенотипов [14].

В одном из исследований методом PheWAS изучали "фенотипическое влияние" генетических данных неандертальцев на современных людей с использованием данных ЭМК. Авторы анализировали ВНП, происходящие от неандертальцев, у ~28000 взрослых людей европейского происхождения и обнаружили ассоциации с неврологическими, психиатрическими, иммунологическими и дерматологическими фенотипами, а также гиперкоагуляцией и употреблением табака. В дополнение к проведению тестов на ассоциацию отдельных ВНП, в этом исследовании изучалось совокупное влияние аллелей неандертальцев на различные фенотипы, что показало связь с депрессией и активным кератозом. В этом исследовании подчеркивается важность использования качественных клинически аннотированных наборов данных для исследования вопросов фундаментальной науки — как в данном случае, эволюционной биологии [15].

#### Этап 4: Статистический анализ

PheWAS проводят с помощью линейной регрессии при непрерывных переменных и логистической регрессии при категориальных переменных с поправкой на ряд ковариат (возраст, пол и др.) и множественные сравнения (поправка Бонферрони, пермутационный тест). Разработано несколько стратегий для выполнения анализа PheWAS. Традиционный подход заключается в поиске ассоциаций с отдельными ВНП. Так, Thériault S, et al. (2019) сначала с помощью GWAS обнаружили локус (rs2069832) в гене *IL6*, ассоциированный с кальцинозом аортального клапана, а далее с помощью PheWAS выявили связь между данным локусом (отдельным ВНП) и повышенным уровнем эозинофилов, лейкоцитов, лимфоцитов, пульсовым давлением, систолическим артериальным давлением, астмой и оперативными вмешательствами на сонной артерии, что указывает на модуляцию путей, опосредуемых интерлейкином-6 [16]. Стратегия PheWAS, использующая один ВНП, обладает низкой мощностью для обнаружения фенотипических ассоциаций, поскольку один ВНП вносит небольшой вклад в несколько фенотипов. Другая стратегия заключается в поиске ассоциаций не с отдельными ВНП, а полигенным риском, рассчитанным с помощью шкал генетического риска (ШГР) для каждого участника исследования. Так, Leppert B, et al. (2020) провели PheWAS для изучения ассоциаций ШГР для 5-и психических расстройств (депрессия, биполярное расстройство, шизофрения, синдром дефицита внимания и гиперактивности и расстройство аутистического спектра) с 23004 фенотипами из базы данных UKBiobank. Анализ позволил выявить 294 значимые фенотипические ассо-

циации, и оказалось, что большинство из них связаны с факторами психического здоровья [17].

Третья стратегия PheWAS заключается в применении МР с использованием данных GWAS. МР — это способ исследования, в котором ВНП используется в качестве инструментальной переменной для установления и количественной оценки причинно-следственной связи. Метод МР позволяет контролировать влияние неучтенных факторов на связь между генетическими данными и фенотипом, что обеспечивает более надежные и точные выводы [18]. Так, Reay WR, et al. (2024) сначала с помощью GWAS с участием 22274 человек идентифицировали 8 распространенных ВНП, ассоциированных с уровнем ретинола, а затем провели PheWAS с МР, где исследовалась связь выявленных ВНП с 20 тыс. клинических фенотипов. Результаты PheWAS с применением МР указывают на то, что ретинол оказывает влияние на воспаление, ожирение, показатели зрения, микробиом и фенотипы состояния мозга, полученные с помощью магнитно-резонансной томографии. И, напротив, уровень циркулирующего ретинола может быть причинно обусловлен такими факторами, как липиды и сывороточный креатинин [19].

На сегодняшний день разработан ряд онлайн-ресурсов и инструментов для анализа PheWAS, имеющих свои сильные и слабые стороны: каталог PheWAS, PheKB (Phenotype Knowledge Base), SPAtest (Score Test and Meta-Analysis Based on Saddlepoint Approximation), SAIGE (Scalable and accurate implementation of generalized mixed model), BioBin (automated bioinformatics tool for the multi-level collapsing of rare variants), GenAMap (software system built on principled machine learning algorithms that detect genome and phenome wide associations), PHESANT (PHEnome Scan ANalysis Tool), MR-Base (database and analytical platform for Mendelian randomization), GeneATLAS (database of associations between hundreds of traits and millions of variants using the UKBiobank cohort) и др.

#### Этап 5: Интерпретация результатов PheWAS

Подобно GWAS, PheWAS можно рассматривать как анализ, генерирующий гипотезы. Когда обнаруживаются значимые ассоциации генотипа и фенотипа, необходимо повторение исследования, чтобы гарантировать, что положительные результаты представляют собой достоверные ассоциации и не являются случайными выводами или артефактами из-за неконтролируемых смещений. Соответствующее повторение для подтверждения результатов PheWAS должно быть проведено на независимых популяциях. В настоящее время для подтверждения результатов, полученных при PheWAS, получил распространение и показал свою эффективность метод МР: сначала выполняют анализ PheWAS, а затем применяют МР для подтверждения значимых результатов PheWAS [5].

Wang L, et al. в 2021г представили систематический обзор 195 исследований с применением PheWAS, из которых 17 были методологическими. Среди них 137 (77,0%) содержали  $\geq 10$  тыс. участников исследования, 164 (92,1%) использовали фенотипы из ЭМК, 14 (7,9%) — из эпидемиологических исследований, 140 (78,7%) в качестве предикторов использовали генетические варианты, 31 (17,4%) — ШГР, 88 (49,4%) — комбинации из нескольких ВНП, 10 (5,6%) — биомаркеры и 73 (41,0%) проводили репликационный анализ [5].

#### Этап 6: Визуализация результатов PheWAS

Для графического представления результатов PheWAS обычно используют манхэттенский график, где по оси X располагается множество фенотипов, и тепловая карта. Манхэттенский график используется для визуализации всех результатов PheWAS, а тепловой график наглядно демонстрирует коррелированные фенотипы и возможную плейотропию [5].

### Возможности PheWAS

#### Репликация ранее выявленных ассоциаций

Исследование с применением PheWAS, опубликованное в 2020г, было проведено на европейско-американской когорте из 6005 чел.: 4 из 7 ранее установленных ассоциаций генотип-фенотип было воспроизведено с  $p$ -значениями от  $2,8 \times 10^{-6}$  до 0,011 [1]. PheWAS с использованием данных UKBiobank подтвердил сильные ранее установленные связи генотипов *APOE* с деменцией (болезнь Альцгеймера), гиперлипидемией и ишемической болезнью сердца по сравнению с референсом  $\epsilon 3/\epsilon 3$  гена *APOE* [20]. С помощью PheWAS на когорте из 4268 детей удалось воспроизвести многие ранее известные ассоциации, полученные в GWAS, в частности, связи между отдельными ВНП и ювенильным ревматоидным артритом, астмой, аутизмом, общим расстройством психологического развития и сахарным диабетом (СД) 1 типа [21].

В 2013г было проведено исследование для демонстрации способности PheWAS воспроизводить GWAS. Было показано, что 210 (28%) из 751 ассоциаций, выявленных в GWAS из числа исследований, содержащихся в NHGRI (National Human Genome Research Institute, Национальный исследовательский институт генома человека) GWAS-каталоге, были воспроизведены с помощью PheWAS, включая 66% тех ассоциаций, которые имели достаточную мощность для обнаружения ассоциаций [22].

#### Выявление новых ассоциаций генотип-фенотип и плейотропности ВНП

Уникальным преимуществом PheWAS является то, что они позволяют одновременно исследовать ассоциации генетических предикторов с широким спектром фенотипов (заболеваний, синдромов, симптомов и пр.). Накапливаются данные о том,

что одни и те же ВНП могут быть связаны с  $\geq 2$  фенотипами, что указывает на способность PheWAS обнаруживать генетическую плейотропию. Идентификация генетической плейотропии дает новые представления об общей генетической этиологии нескольких заболеваний [5].

PheWAS с использованием данных UKBiobank позволил выявить потенциальные связи между наличием аллеля  $\epsilon 4$  *APOE* и меньшим риском наличия ожирения, СД 2 типа, хронической обструктивной болезни легких, желчнокаменной болезни и заболеваний печени по сравнению с носителями генотипа  $\epsilon 3/\epsilon 3$  гена *APOE*. Генотип *APOE*  $\epsilon 2/\epsilon 2$  был связан с большой частотой периферических тромбозов, аневризм, пептических язв, заболеваний шейки матки и вальгусной деформации большого пальца стопы по сравнению с носителями генотипа  $\epsilon 3/\epsilon 3$  гена *APOE*. Предполагается, что именно большой размер когорты (337484 чел.) и единообразный подход к генотипированию и фенотипированию, который может обеспечить UKBiobank, позволил выявить новые ассоциации [20].

Li X, et al. (2019), исследовав 339256 биообразцов неродственных белых британцев из UKBiobank, с помощью MR успешно воспроизвели выявленную в PheWAS связь генетически детерминированного уровня мочевой кислоты в сыворотке крови с подагрой, артериальной гипертензией (АГ), болезнями сердца и уровнями липидов в крови и продемонстрировали существование генетической плейотропии [23].

Zheutlin AB, et al. (2019) с помощью PheWAS определили, что ШГР шизофрении связаны с другими психическими расстройствами, такими как тревожность, неврологические и личностные расстройства, суицидальное поведение и потеря памяти, что указывает на общий генетический риск среди этих фенотипов [24].

Исследования с применением PheWAS могут как доказать взаимосвязь между генотип-фенотипом, так и опровергнуть ее. Так, исследование, проведенное в 2019г на когорте из 51700 европейцев с целью изучения плейотропного эффекта ВНП генов *PCSK9*, *APOB* и *LDLR*, ассоциированных с нарушениями липидного обмена [25], только для одного из исследуемых ВНП позволило обнаружить связь с нелипидным фенотипом, миопией, однако эта связь не была воспроизведена в репликационных когортах. Таким образом, предполагаемые потенциальные связи ВНП в генах *PCSK9*, *APOB* и *LDLR* с СД или какими-либо другими нелипидными фенотипами, включая нейрокогнитивные расстройства или катаракту, выявлены не были [25].

#### Изучение новых патогенетических связей и коморбидности

PheWAS служит полезным инструментом и для выявления потенциальных причинно-следственных

связей между фенотипами, в т.ч. генетически обоснованной коморбидности. Понимание того, какие пары заболеваний, вероятно, или маловероятно, будут встречаться одновременно, может помочь выяснить потенциальные связи между сложными заболеваниями. Сосредоточившись на вариантах, имеющих известную биологическую функцию и/или клиническое значение, PheWAS способны упростить интерпретацию новых результатов путем непосредственного применения известной информации о генетических вариантах и, таким образом, раскрытия потенциальной причинно-следственной связи. Понимание коморбидности необходимо для разработки индивидуальных программ профилактики, лечения и прогнозирования заболеваний [26].

Так, Prinzbach A, et al. (2018) при исследовании 433 больных целиакией vs 4330 человек из группы контроля с помощью PheWAS и данных ЭМК была продемонстрирована связь целиакии с рядом сопутствующих заболеваний (с 45-ю кодами МКБ-10), при этом впервые в контролируемом исследовании удалось выявить значимую ассоциацию между целиакией и эозинофильным эзофагитом [27]. Liu TY, et al. (2022), объединив данные о фенотипах из ЭМК 23349 чел. и данные таргетного секвенирования, провели GWAS и обнаружили 44 новых ВНП в 10 генах (*CTLA4*, *HCP5*, *HLA-B*, *POU5F1*, *CCHCR1*, *HLA-DRA*, *HLA-DRB9*, *TSHR*, *RPL17P3* и *CEP128*), ассоциированных с гипертиреозом, а далее с помощью PheWAS продемонстрировали, что выявленные ВНП могут предсказывать риск развития тиреотоксического криза и инсульта у больных с гипертиреозом [3].

Chen SY, et al. (2024) с помощью GWAS, проведенных на 75121 участниках, из которых 5694 являлись больными с фибрилляцией предсердий (ФП), выявили 47 ВНП, ассоциированных с ФП. Далее, объединив выявленные варианты в ШГР и превратив коды МКБ-10 в 1792 фекода, провели PheWAS, который продемонстрировал связь высокого полигенного риска ФП (верхний дециль ШГР) с рядом других фенотипов, среди которых трепетание предсердий, нарушения ритма сердца, застойная сердечная недостаточность, blastomикоз и гипертоническая болезнь сердца. Полученные результаты могут служить объяснением потенциального развития ряда сопутствующих ФП заболеваний [28].

Larsson SC, et al. (2022) провели изучение ассоциации уровня липопротеина(а) с широким спектром заболеваний у 385917 человек из UKBiobank. Была построена ШГР с использованием независимых ВНП в гене *LPA* с установленным влиянием на уровень липопротеина(а). PheWAS, в который был включен 1081 фенотип, выявил статистически значимые ассоциации повышенного уровня липопротеина(а) с более высоким риском значительного числа сердечно-сосудистых заболеваний

(ССЗ) (различные заболевания сердца, периферических сосудов, АГ, клапанные и цереброваскулярные патологии), а также некоторых эндокринных/метаболических заболеваний (нарушения липидного обмена и СД 2 типа), заболеваний мочеполовой системы (почечная недостаточность) и гематологических заболеваний (включая различные виды анемии) [29].

PheWAS для 15 сердечно-сосудистых фенотипов и 14 показателей уровней полиненасыщенных жирных кислот из базы данных исследования LURIC (Ludwigshafen Risk and Cardiovascular Health Study), включившей оставшихся после фильтрации по качеству 2824 человека и 577007 ВНП, позволили выявить 7 ВНП, значимо ассоциированных с дигомо-гамма-линоленовой кислотой, которая может быть преобразована в противовоспалительные метаболиты, играющие защитную роль в патогенезе ССЗ. С учетом потребления полиненасыщенных жирных кислот с пищей, полученный результат указывает на сложную взаимосвязь генетических особенностей и факторов внешней среды, управляющих ССЗ [9].

Lui J, et al. (2022) обнаружили прогностическое влияние ряда фосфолипидов на развитие АГ в будущем, таким образом уровень фосфолипидов может быть предиктором возникновения АГ [30].

Nam Y, et al. (2023), исследовав 427 фенотипов UKBiobank, представили математический подход для построения сети взаимодействия между заболеваниями, которая может учитывать не только значимую связь с заболеваниями, выявленную при PheWAS, но и направление (однонаправленное или разнонаправленное) эффектов ВНП, полученных в PheWAS. Кроме того, авторы статьи продемонстрировали алгоритм оценки коморбидности для прогнозирования того, будут ли заболевания встречаться одновременно при наличии конкретного основного заболевания [26].

#### Поиск новых подходов к лечению и профилактике

Выявление новых ассоциаций с помощью PheWAS может привести к открытию новых механизмов и взаимосвязей, детерминирующих здоровье человека, а также к разработке новых подходов к лечению и профилактике. PheWAS признаны перспективным подходом к созданию новых стратегий лечения посредством уточнения эффекта лекарственных препаратов. Millwood IY, et al. (2016) выполнили PheWAS для выявления ассоциаций ВНП V279F (rs76863441), который инактивирует активность липопротеин-ассоциированной фосфолипазы A2 (Lp-PLA2), с широким диапазоном исходов заболеваний. Они обнаружили, что низкая активность Lp-PLA2 на протяжении всей жизни не была связана с серьезными рисками сосудистых или несосудистых заболеваний [31]. Это открытие ставит под сомнение защитную роль ингибитора Lp-PLA2, та-

кого как дарапладиб, в предотвращении ССЗ и дает важные идеи для дальнейшей разработки лекарств.

Множество исследований продемонстрировали связи между уровнями витамина D и сотнями заболеваний. Однако доказательства причинно-следственной связи для большинства этих ассоциаций либо отсутствуют, либо не подтверждены рандомизированными клиническими испытаниями или исследованиями с МР. Meng X, et al. (2019) на выборке 339256 человек из UKBiobank исследовали предполагаемые причинно-следственные эффекты витамина D в отношении 920 фенотипов с помощью МР-PheWAS и не подтвердили связи ни для одного из протестированных фенотипов, что ставит под сомнение обоснованность широкого назначения витамина D [32].

#### Изучение безопасности лекарственных средств

Помимо перепрофилирования, подход PheWAS может быть использован для прогнозирования потенциальных побочных эффектов, связанных с использованием лекарств, в т.ч. на ранних стадиях клинических испытаний. Так, Jerome RN, et al. на когорте из 29722 европейцев реплицировали с помощью PheWAS данные о безопасности для 13 из 16 пар ген-лекарство [33].

Diogo D, et al. (2018), отобрав 25 ВНП, связанных по данным GWAS с 19 потенциальными лекарственными мишенями, провели PheWAS на когорте из 697815 человек, воспроизвели 75% известных ассоциаций GWAS и выявили 9 новых значимых ассоциаций, которые могут предсказывать побочные эффекты, такие как акне, повышение уровня холестерина, подагру, желчнокаменную болезнь и астму [34].

Используя PheWAS, Jerome RN, et al. (2018) проанализировали имеющиеся генотипические и клинические данные 29722 человек с целью поиска новых фенотипов (помимо влияния на уровень холестерина липопротеинов низкой плотности),

с которыми могут быть ассоциированы ВНП гена PCSK9. Был выявлен ВНП R46L, не только ассоциированный со значительным снижением риска гиперхолестеринемии (отношение шансов (OR — odds ratio) 0,68,  $p=7,6 \times 10^{-4}$ ) в связи с потерей функции белка PCSK9, но и связанный с повышенным риском врожденной спинно-мозговой грыжи. Результаты исследования, с одной стороны, предполагают, что дальнейшее изучение потенциальных связей между ингибированием PCSK9 и дефектами нервной трубки может быть оправдано, а с другой, — позволяют рассматривать PheWAS как инструмент для оценки безопасности и долгосрочной переносимости лекарственных средств, в частности, ингибиторов PCSK9 [35].

#### Заключение

Таким образом, PheWAS могут улучшить понимание того, какие заболевания могут или вряд ли будут сопутствовать друг другу. Определение ассоциаций заболеваний с учетом общих генетических предикторов позволит разработать генетические модели коморбидности, прогнозирующие риск сочетанного развития заболеваний еще до первых клинических проявлений, а также, вероятно, станет одной из основ в подходах к разработке новой медикаментозной терапии. Это свидетельствует о значимости и ценности дальнейшего использования генетических исследований методом PheWAS, в т.ч. для изучения, лечения, прогнозирования хронических неинфекционных заболеваний. Накопление данных в России — фенотипы и генотипы — предрасполагает к проведению PheWAS на когортах, в которые включены представители российской популяции.

**Отношения и деятельность:** все авторы заявляют об отсутствии потенциального конфликта интересов, требующего раскрытия в данной статье.

#### Литература/References

1. Denny JC, Ritchie MD, Basford MA, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics*. 2010;26(9):1205-10. doi:10.1093/BIOINFORMATICS/BTQ126.
2. Denny JC, Bastarache L, Roden DM. Phenome-Wide Association Studies as a Tool to Advance Precision Medicine. *Annu Rev Genomics Hum Genet*. 2016;17(1):353-73. doi:10.1146/annurev-genom-090314-024956.
3. Liu TY, Liao WL, Wang TY, et al. Genome-wide association study of hyperthyroidism based on electronic medical record from Taiwan. *Front Med*. 2022;9. doi:10.3389/fmed.2022.830621.
4. Liu S, Crawford DC. Maturation and application of phenome-wide association studies. *Trends Genet*. 2022;38(4):353-63. doi:10.1016/j.tig.2021.12.002.
5. Wang L, Zhang X, Meng X, et al. Methodology in phenome-wide association studies: a systematic review. *J Med Genet*. 2021;58(11):720-8. doi:10.1136/jmedgenet-2021-107696.
6. Roden DM. Phenome-wide association studies: a new method for functional genomics in humans. *J Physiol*. 2017;595(12):4109-15. doi:10.1113/JP273122.
7. Verma A, Verma SS, Pendergrass SA, et al. eMERGE Phenome-Wide Association Study (PheWAS) identifies clinical associations and pleiotropy for stop-gain variants. *BMC Med Genomics*. 2016; 9(S1):32. doi:10.1186/s12920-016-0191-8.
8. Wu P, Gifford A, Meng X, et al. Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation. *JMIR Med Informatics*. 2019;7(4):e14325. doi:10.2196/14325.
9. Passero K, He X, Zhou J, et al. Phenome-wide association studies on cardiovascular health and fatty acids considering phenotype quality control practices for epidemiological data. *Pac Symp Biocomput*. 2020;25:659-70.
10. Verma A, Bradford Y, Dudek S, et al. A simulation study investigating power estimates in phenome-wide association studies. *BMC Bioinformatics*. 2018;19(1):120. doi:10.1186/s12859-018-2135-0.

11. Bastarache L, Denny JC, Roden DM. Phenome-Wide Association Studies. *JAMA*. 2022;327(1):75. doi:10.1001/jama.2021.20356.
12. Hebring SJ. The challenges, advantages and future of phenome-wide association studies. *Immunology*. 2014;141(2):157-65. doi:10.1111/imm.12195.
13. Pendergrass SA, Brown-Gentry K, Dudek S, et al. Phenome-Wide Association Study (PheWAS) for Detection of Pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) Network. Gibson G, ed. *PLoS Genet*. 2013;9(1):e1003087. doi:10.1371/journal.pgen.1003087.
14. Evangelou E, Ioannidis JPA. Meta-analysis methods for genome-wide association studies and beyond. *Nat Rev Genet*. 2013;14(6):379-89. doi:10.1038/nrg3472.
15. Simonti CN, Vernot B, Bastarache L, et al. The phenotypic legacy of admixture between modern humans and Neandertals. *Science*. 2016;351(6274):737-41. doi:10.1126/science.aad2149.
16. Thériault S, Dina C, Messika-Zeitoun D, et al. Genetic Association Analyses Highlight IL6, ALPL, and NAV1 As 3 New Susceptibility Genes Underlying Calcific Aortic Valve Stenosis. *Circ Genomic Precis Med*. 2019;12(10):e002617. doi:10.1161/CIRCGEN.119.002617.
17. Leppert B, Millard LAC, Riglin L, et al. A cross-disorder PRS-pheWAS of 5 major psychiatric disorders in UK Biobank. Zhu X, ed. *PLOS Genet*. 2020;16(5):e1008185. doi:10.1371/journal.pgen.1008185.
18. Plotnikov DY. Mendelian randomization: using genetic information in epidemiological studies (review). *Research Results in Biomedicine*. 2023;9(2):158-72. (In Russ.) Плотников Д. Ю. Менделевская рандомизация: применение генетической информации в эпидемиологических исследованиях (обзор). Научные результаты биомедицинских исследований. 2023; 9(2):158-72. doi:10.18413/2658-6533-2023-9-2-0-1.
19. Reay WR, Kiltschewskij DJ, Di Biase MA, et al. Genetic influences on circulating retinol and its relationship to human health. *Nat Commun*. 2024;15(1):1490. doi:10.1038/s41467-024-45779-x.
20. Lumsden AL, Mulugeta A, Zhou A, Hyppönen E. Apolipoprotein E (APOE) genotype-associated disease risks: a phenome-wide, registry-based, case-control study utilising the UK Biobank. *eBioMedicine*. 2020;59:102954. doi:10.1016/j.ebiom.2020.102954.
21. Namjou B, Marsolo K, Carroll RJ, et al. Phenome-wide association study (PheWAS) in EMR-linked pediatric cohorts, genetically links PLCL1 to speech language development and IL5-IL13 to Eosinophilic Esophagitis. *Front Genet*. 2014;5. doi:10.3389/fgene.2014.00401.
22. Denny JC, Bastarache L, Ritchie MD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol*. 2013;31(12):1102-11. doi:10.1038/nbt.2749.
23. Li X, Meng X, He Y, et al. Genetically determined serum urate levels and cardiovascular and other diseases in UK Biobank cohort: A phenome-wide mendelian randomization study. Rahimi K, ed. *PLOS Med*. 2019;16(10):e1002937. doi:10.1371/journal.pmed.1002937.
24. Zheutlin AB, Dennis J, Karlsson Linnér R, et al. Penetrance and Pleiotropy of Polygenic Risk Scores for Schizophrenia in 106,160 Patients Across Four Health Care Systems. *Am J Psychiatry*. 2019;176(10):846-55. doi:10.1176/appi.ajp.2019.18091085.
25. Safarova MS, Satterfield BA, Fan X, et al. A phenome-wide association study to discover pleiotropic effects of PCSK9, APOB, and LDLR. *NPJ Genomic Med*. 2019;4(1):3. doi:10.1038/s41525-019-0078-7.
26. Nam Y, Jung SH, Yun JS, et al. Discovering comorbid diseases using an inter-disease interactivity network based on biobank-scale PheWAS data. Wren J, ed. *Bioinformatics*. 2023;39(1). doi:10.1093/bioinformatics/btac822.
27. Prinzbach A, Moosavinasab S, Rust S, et al. Comorbidities in Childhood Celiac Disease. *J Pediatr Gastroenterol Nutr*. 2018;67(4):488-93. doi:10.1097/MPG.0000000000002020.
28. Chen SY, Chen YC, Liu TY, et al. Novel Genes Associated With Atrial Fibrillation and the Predictive Models for AF Incorporating Polygenic Risk Score and PheWAS-Derived Risk Factors. *Can J Cardiol*. 2024;40(11):2117-27. doi:10.1016/j.cjca.2024.07.029.
29. Larsson SC, Wang L, Li X, Jiang F, Chen X, Mantzoros CS. Circulating lipoprotein(a) levels and health outcomes: Phenome-wide Mendelian randomization and disease-trajectory analyses. *Metabolism*. 2022;137:155347. doi:10.1016/j.metabol.2022.155347.
30. Liu J, de Vries PS, Del Greco MF, et al. A multi-omics study of circulating phospholipid markers of blood pressure. *Sci Rep*. 2022;12(1):574. doi:10.1038/s41598-021-04446-7.
31. Millwood IY, Bennett DA, Walters RG, et al. A phenome-wide association study of a lipoprotein-associated phospholipase A 2 loss-of-function variant in 90000 Chinese adults. *Int J Epidemiol*. 2016;45(5):1588-99. doi:10.1093/ije/dyw087.
32. Meng X, Li X, Timofeeva MN, et al. Phenome-wide Mendelian-randomization study of genetically determined vitamin D on multiple health outcomes using the UK Biobank study. *Int J Epidemiol*. 2019;48(5):1425-34. doi:10.1093/ije/dyz182.
33. Jerome RN, Joly MM, Kennedy N, et al. Leveraging Human Genetics to Identify Safety Signals Prior to Drug Marketing Approval and Clinical Use. *Drug Saf*. 2020;43(6):567-82. doi:10.1007/s40264-020-00915-6.
34. Diogo D, Tian C, Franklin CS, et al. Phenome-wide association studies across large population cohorts support drug target validation. *Nat Commun*. 2018;9(1):4285. doi:10.1038/s41467-018-06540-3.
35. Jerome RN, Pulley JM, Roden DM, et al. Using Human "Experiments of Nature" to Predict Drug Safety Issues: An Example with PCSK9 Inhibitors. *Drug Saf*. 2018;41(3):303-11. doi:10.1007/s40264-017-0616-0.