

Валидация шкал генетического риска развития сахарного диабета 2 типа на выборке населения регионов России из биобанка "НМИЦ ТПМ"

Киселева А. В.¹, Сопленкова А. Г.¹, Куценко В. А.^{1,2}, Сотникова Е. А.¹, Вяткин Ю. В.^{1,2}, Жарикова А. А.^{1,2}, Ершова А. И.¹, Зайченко М.³, Раменский В. Е.^{1,2}, Скирко О. П.¹, Сметнев С. А.¹, Копылова О. В.¹, Лимонова А. С.¹, Блохина А. В.¹, Покровская М. С.¹, Шальнова С. А.¹, Мешков А. Н.¹, Драпкина О. М.¹

¹ФГБУ "Национальный медицинский исследовательский центр терапии и профилактической медицины" Минздрава России. Москва; ²ФГБОУ ВО "Московский государственный университет им. М. В. Ломоносова". Москва; ³ФГАОУ ВО "Московский физико-технический институт (национальный исследовательский университет)". Московская область, Долгопрудный, Россия

Цель. Провести валидацию и оценить точность 14 шкал генетического риска (ШГР) сахарного диабета 2 типа (СД2), созданных ранее в других странах, на выборке населения регионов России из биобанка ФГБУ "НМИЦ ТПМ" Минздрава России.

Материал и методы. Для генетического анализа были использованы данные секвенирования следующего поколения на выборке из российской популяции (n=1165) на основе коллекции биобанка. Исследование включало 14 ШГР, ассоциированных с СД2.

Результаты. В рамках исследования было продемонстрировано, что предсказательная сила 12 из 14 ШГР для СД2 воспроизводится на российской популяции. В качестве метрик качества была использована площадь под ROC-кривой, которая для моделей с включением только ШГР варьировалась от 54,49 до 59,46%, а для моделей, включающих ШГР, пол и возраст — от 77,56 до 78,75%.

Заключение. Впервые в России проведено исследование 14 ШГР СД2, разработанных на других популяциях. 12 ШГР прошли валидацию и в будущем могут быть применены для улучшения прогнозирования риска развития и профилактики СД2 в России.

Ключевые слова: сахарный диабет 2 типа, шкала генетического риска, валидация, популяционное исследование, биобанк.

Отношения и деятельность: нет.

Поступила 22/09-2023

Рецензия получена 18/10-2023

Принята к публикации 24/10-2023



Для цитирования: Киселева А. В., Сопленкова А. Г., Куценко В. А., Сотникова Е. А., Вяткин Ю. В., Жарикова А. А., Ершова А. И., Зайченко М., Раменский В. Е., Скирко О. П., Сметнев С. А., Копылова О. В., Лимонова А. С., Блохина А. В., Покровская М. С., Шальнова С. А., Мешков А. Н., Драпкина О. М. Валидация шкал генетического риска развития сахарного диабета 2 типа на выборке населения регионов России из биобанка "НМИЦ ТПМ". *Кардиоваскулярная терапия и профилактика*. 2023;22(11):3746. doi:10.15829/1728-8800-2023-3746. EDN VADXMO

*Автор, ответственный за переписку (Corresponding author):

e-mail: sanyutabe@gmail.com

[Киселева А. В.* — к.б.н., в.н.с., руководитель лаборатории молекулярной генетики Института персонализированной терапии и профилактики, ORCID: 0000-0003-4765-8021, Сопленкова А. Г. — лаборант отдела эпидемиологии хронических неинфекционных заболеваний, лаборатория биостатистики, ORCID: 0000-0003-0703-146X, Куценко В. А. — с.н.с. отдела эпидемиологии хронических неинфекционных заболеваний, лаборатория биостатистики, аспирант Механико-математического факультета, ORCID: 0000-0001-9844-3122, Сотникова Е. А. — с.н.с. лаборатории молекулярной генетики Института персонализированной терапии и профилактики, ORCID: 0000-0002-8395-4146, Вяткин Ю. В. — программист лаборатории геномной и медицинской биоинформатики Института персонализированной терапии и профилактики, с.н.с. Института перспективных исследований проблем искусственного интеллекта и интеллектуальных систем, ORCID: 0000-0002-9056-8796, Жарикова А. А. — к.б.н., в.н.с. лаборатории молекулярной генетики Института персонализированной терапии и профилактики, старший преподаватель Факультета биоинженерии и биоинформатики, ORCID: 0000-0003-0723-0493, Ершова А. И. — д.м.н., зам. директора по фундаментальной науке, руководитель лаборатории клинической, ORCID: 0000-0001-7989-0760, Зайченко М. — аспирант, ORCID: 0000-0002-2798-9811, Раменский В. Е. — к.ф.-м.н., в.н.с., руководитель лаборатории геномной и медицинской биоинформатики Института персонализированной терапии и профилактики, в.н.с. Института перспективных исследований проблем искусственного интеллекта и интеллектуальных систем, доцент Факультета биоинженерии и биоинформатики, ORCID: 0000-0001-7867-9509, Скирко О. П. — м.н.с. лаборатории молекулярной генетики Института персонализированной терапии и профилактики, ORCID: 0000-0003-3755-0279, Сметнев С. А. — м.н.с. отдела персонализированной диагностики, профилактики и терапии атеросклеротических сердечно-сосудистых заболеваний Института персонализированной терапии и профилактики, ORCID: 0000-0002-8493-4761, Копылова О. В. — к.м.н., с.н.с. лаборатории клинической, ORCID: 0000-0001-5397-5387, Лимонова А. С. — н.с. лаборатории клинической, ORCID: 0000-0003-1500-3696, Блохина А. В. — к.м.н., н.с. лаборатории клинической, ORCID: 0000-0002-3019-3961, Покровская М. С. — к.б.н., в.н.с., руководитель лаборатории "Банк биологического материала" Института персонализированной терапии и профилактики, ORCID: 0000-0001-6985-7131, Шальнова С. А. — д.м.н., профессор, руководитель отдела эпидемиологии хронических неинфекционных заболеваний, ORCID: 0000-0003-2087-6483, Мешков А. Н. — д.м.н., руководитель отдела персонализированной диагностики, профилактики и терапии атеросклеротических сердечно-сосудистых заболеваний, руководитель Института персонализированной терапии и профилактики, ORCID: 0000-0001-5989-6233, Драпкина О. М. — д.м.н., профессор, академик РАН, директор, ORCID: 0000-0002-4453-8430].

Validation of genetic risk scores for type 2 diabetes on a Russian population sample from the biobank of the National Medical Research Center for Therapy and Preventive Medicine

Kiseleva A. V.¹, Soplenkova A. G.¹, Kutsenko V. A.^{1,2}, Sotnikova E. A.¹, Vyatkin Yu. V.^{1,2}, Zharikova A. A.^{1,2}, Ershova A. I.¹, Zaichenoka M.³, Ramensky V. E.^{1,2}, Skirko O. P.¹, Smetnev S. A.¹, Kopylova O. V.¹, Limonova A. S.¹, Blokhina A. V.¹, Pokrovskaya M. S.¹, Shalnova S. A.¹, Meshkov A. N.¹, Drapkina O. M.¹

¹National Medical Research Center for Therapy and Preventive Medicine. Moscow; ²Lomonosov Moscow State University. Moscow; ³Moscow Institute of Physics and Technology. Moscow Oblast, Dolgoprudny, Russia

Aim. To validate and evaluate the accuracy of 14 genetic risk scores (GRSs) for type 2 diabetes (T2D), created earlier in other countries, using a Russian population sample from the biobank of the National Medical Research Center for Therapy and Preventive Medicine.

Material and methods. For genetic analysis, next generation sequencing data was used on a sample from the Russian population (n=1165) based on the biobank collection. The study included 14 GRSs associated with T2D.

Results. The study demonstrated that the predictive power of 12 out of 14 GRSs for T2D was replicated in the Russian population. As quality metrics, we used the area under the ROC curve, which for models including only GRS varied from 54,49 to 59,46%, and for models including GRS, sex and age — from 77,56 to 78,75%.

Conclusion. For the first time in Russia, a study of 14 T2D GRSs developed on other populations was conducted. Twelve GRSs have been validated and can be used in the future to improve risk prediction and prevention of T2D in Russia.

Keywords: type 2 diabetes, genetic risk score, validation, population study, biobank.

Relationships and Activities: none.

Kiseleva A. V.* ORCID: 0000-0003-4765-8021, Soplenkova A. G. ORCID: 0000-0003-0703-146X, Kutsenko V. A. ORCID: 0000-0001-9844-3122, Sotnikova E. A. ORCID: 0000-0002-8395-4146, Vyatkin Yu. V. ORCID: 0000-

0002-9056-8796, Zharikova A. A. ORCID: 0000-0003-0723-0493, Ershova A. I. ORCID: 0000-0001-7989-0760, Zaichenoka M. ORCID: 0000-0002-2798-9811, Ramensky V. E. ORCID: 0000-0001-7867-9509, Skirko O. P. ORCID: 0000-0003-3755-0279, Smetnev S. A. ORCID: 0000-0002-8493-4761, Kopylova O. V. ORCID: 0000-0001-5397-5387, Limonova A. S. ORCID: 0000-0003-1500-3696, Blokhina A. V. ORCID: 0000-0002-3019-3961, Pokrovskaya M. S. ORCID: 0000-0001-6985-7131, Shalnova S. A. ORCID: 0000-0003-2087-6483, Meshkov A. N. ORCID: 0000-0001-5989-6233, Drapkina O. M. ORCID: 0000-0002-4453-8430.

*Corresponding author: sanyutabe@gmail.com

Received: 22/09-2023

Revision Received: 18/10-2023

Accepted: 24/10-2023

For citation: Kiseleva A. V., Soplenkova A. G., Kutsenko V. A., Sotnikova E. A., Vyatkin Yu. V., Zharikova A. A., Ershova A. I., Zaichenoka M., Ramensky V. E., Skirko O. P., Smetnev S. A., Kopylova O. V., Limonova A. S., Blokhina A. V., Pokrovskaya M. S., Shalnova S. A., Meshkov A. N., Drapkina O. M. Validation of genetic risk scores for type 2 diabetes on a Russian population sample from the biobank of the National Medical Research Center for Therapy and Preventive Medicine. *Cardiovascular Therapy and Prevention*. 2023;22(11):3746. doi:10.15829/1728-8800-2023-3746. EDN VADXMO

ВНП — вариант нуклеотидной последовательности, СД2 — сахарный диабет 2 типа, ШГР — шкала генетического риска, ЗССЕ-РФ — Эпидемиология сердечно-сосудистых заболеваний и их факторов риска в регионах Российской Федерации, AUC — Area Under the Curve (площадь под ROC-кривой), GWAS — genome-wide association studies (полногеномный поиск ассоциаций), NGS — next-generation sequencing (секвенирование следующего поколения).

Ключевые моменты

Что известно о предмете исследования?

- Для оценки генетической предрасположенности человека к определенному признаку или заболеванию используют шкалы генетического риска (ШГР).
- Большинство ШГР объясняют лишь небольшую часть вариабельности признака, но они являются устойчивой мерой генетической предрасположенности человека к заболеваниям.

Что добавляют результаты исследования?

- Впервые в России проведена валидация 14 ШГР сахарного диабета (СД) 2 типа, разработанных на популяциях европейского происхождения.
- Из 14 ШГР СД 2 типа 12 были значимо ассоциированы с СД 2 типа в российской популяции.

Key messages

What is already known about the subject?

- To assess a person's genetic predisposition to a particular trait or disease, genetic risk scores (GRS) are used.
- Most GRSs explain only a small portion of trait variance, but they are a robust measure of a person's genetic susceptibility to disease.

What might this study add?

- For the first time in Russia, 14 GRSs for type 2 diabetes, developed on European populations, were validated.
- Of the 14 GRSs for type 2 diabetes, 12 were significantly associated with type 2 diabetes in the Russian population.

Введение

В 2019г сахарный диабет (СД) стал причиной ~2 млн смертей по всему миру. Более 95% от общего числа пациентов, страдающих СД, имеют СД 2 типа (СД2)¹. Кроме того, СД2 является фактором риска сердечно-сосудистых заболеваний [1]. Данные о наследуемости СД2 сильно варьируются от 26 до 72% согласно различным близнецовым исследованиям [2, 3].

За последние 20 лет, благодаря полногеномному поиску ассоциаций (genome-wide association studies — GWAS) показано, что большинство распространенных заболеваний человека объясняются не вариантами в одном гене, а большим количеством отдельных вариантов, распределенных по всему геному [4]. GWAS Catalog, база данных, агрегирующая информацию об исследованиях по полногеномному поиску ассоциаций, содержит информацию о 227 исследованиях, где одним из изучаемых признаков был СД2². На основании выявленных ассоциаций создаются шкалы генетического риска (ШГР), используемые для оценки генетической предрасположенности человека к определенному признаку или заболеванию. Значение ШГР для некоторого признака конкретного человека рассчитывается на основе генотипа человека и данных, полученных в GWAS для этого признака [5, 6]. Для создания ШГР могут быть использованы данные GWAS, проведенного в рамках того же исследования [7] или выполненного ранее на другой выборке [8, 9]. Число включенных в ШГР вариантов широко варьируется и может достигать нескольких млн³. Большинство ШГР объясняют лишь небольшую часть вариативности признака, однако они являются устойчивой мерой генетической предрасположенности человека к заболеваниям, что привело к рутинному применению ШГР в биомедицинских исследованиях [5].

Для проведения валидации уже существующих ШГР необходимы выборки, включающие тысячи участников. Поэтому наиболее оптимальным является использование данных биобанков, содержащих информацию об изучаемом признаке. Примером может выступать коллекция данных биобанка Соединенного Королевства (UK Biobank) [8, 10], которая позволяет проводить как повторные исследования, так и оценивать этнические особенности наследуемости заболевания или признака. На сегодняшний день в России такие исследования стали доступны благодаря развитию биобан-

ков. В частности, в настоящей работе использованы данные биобанка ФГБУ "НМИЦ ТПМ" Минздрава России [11] — одного из наиболее крупных биобанков РФ, члена Национальной ассоциации биобанков и специалистов по биобанкированию (НАСБИО) [12].

К настоящему времени созданы десятки ШГР СД2 [13]. В рамках ряда исследований было продемонстрировано, что при применении моделей оценки риска возникает необходимость принимать во внимание популяционный эффект даже для лиц из разных европейских популяций одинаковой этнической принадлежности [4, 14, 15]. Генетическая структура российской и европейской популяции различается еще более существенно [16, 17]. Кроме того, популяционные исследования, проведенные на выборках регионов европейской части России, показали достоверные отличия в аллельных частотах некоторых клинически значимых вариантов нуклеотидной последовательности (ВНП) от данных, полученных на европейской популяции (non-Finnish Europeans) [18, 19].

Целью настоящей работы: провести валидацию и оценить точность 14 шкал ШГР СД2, созданных ранее в других странах, на выборке населения регионов России из биобанка ФГБУ "НМИЦ ТПМ" Минздрава России.

Материал и методы

Выборка

В исследование была включена подвыборка из репрезентативной выборки населения Вологодской области (n=707) из исследования "Эпидемиология сердечно-сосудистых заболеваний и их факторов риска в регионах Российской Федерации" (ЭССЕ-Вологда) [20] и 460 пациентов, чья кровь была биобанкирована в Биобанке ФГБУ "НМИЦ ТПМ" Минздрава России, общая выборка составила 1167 участников, доля мужчин в выборке — 48,8%. Общая характеристика выборки представлена в таблице 1. В рамках исследования ЭССЕ-РФ проводился опрос по стандартному вопроснику, разработанному на основе адаптированных международных методик и включающему информацию о поведенческих привычках, анамнестических данных и т.д. [20]. Диагноз СД2 ставился при наличии хотя бы одного из следующих критериев: уровень глюкозы натощак >7 ммоль/л; положительный ответ на вопрос: "Говорил ли Вам когда-нибудь врач, что у Вас имеется/имелся сахарный диабет 2 типа?"; прием гипогликемической терапии. В общей выборке в анализ включены возраст, пол, уровень глюкозы натощак (ммоль/л) и/или диагноз СД2, и/или сахароснижающая терапия. Показатель, "диагноз СД2" был доступен для обеих выборок, а показатели "уровень глюкозы натощак (ммоль/л)" и "прием гипогликемической терапии" — только для ЭССЕ-Вологда (n=703). Количество участников с СД2 (поставленный диагноз или уровень глюкозы >7 ммоль/л) составило 286 из 1167. Обе выборки являются коллекцией биобанка ФГБУ "НМИЦ ТПМ" Минздрава России [11]. Исследование одобрено Независимым Этическим Комитетом ФГБУ "НМИЦ ТПМ"

¹ World Health Organization. <http://www.who.int/news-room/fact-sheets/detail/diabetes> (23 October 2023).

² GWAS Catalog. https://www.ebi.ac.uk/gwas/efotraits/MONDO_0005148 (23 October 2023).

³ PGS Catalog. https://www.pgscatalog.org/trait/MONDO_0005148/ (23 October 2023).

Таблица 1

Характеристика выборки

Показатель	Вся выборка, n=1167	Мужчины, n=569 (48,8%)	Женщины, n=598 (51,2%)	p
Возраст, лет (Me [Q25; 75], n=1167)	50 [38; 58]	48 [37; 56]	52,0 [39,3; 60,0]	<0,001
Уровень глюкозы, ммоль/л (Me [Q25; 75], n=703)	5,09 [4,7; 5,6]	5,25 [4,9; 5,8]	4,96 [4,6; 5,4]	<0,001
Диагноз СД2, % (n=1167)	286 (24,5%)	161 (28,3%)	125 (20,9%)	0,003
Рост, см (Me [Q25; 75], n=1164)	169,0 [161,5; 176,5]	176,55 [172,7; 181,0]	162,0 [158,0; 166,0]	<0,001
Вес, кг (Me [Q25; 75], n=1164)	83,0 [69,0; 100,3]	90,15 [77,4; 107,9]	74 [62,5; 92,0]	<0,001
ИМТ, кг/м ² (Me [Q25; 75], n=1165)	28,8 [24,2; 35,0]	29,0 [25,0; 34,0]	28,0 [23,4; 35,8]	0,064

Примечание: ИМТ — индекс массы тела, СД2 — сахарный диабет 2 типа, Me — медиана, Q — квартиль.

Минздрава России. Все участники дали письменное информированное согласие.

Выбор ШГР

В исследование были включены 14 ШГР [8, 9, 21–32], которые были доступны на момент начала исследования в 2021г, рассчитанные для европейской популяции (non-Finnish Europeans) [13]. Исходные данные по всем 14 ШГР были подробно описаны ранее [13].

Генетический анализ

Геномная дезоксирибонуклеиновая кислота (ДНК) из образцов цельной крови была выделена с помощью набора QIAamp DNA Blood Mini Kit (Qiagen, Германия). Концентрацию ДНК определяли на флуориметре Qubit 4.0 (Thermo Fisher Scientific, США) или спектрофотометре NanoDrop OneC (Thermo Fisher Scientific, США). Секвенирование следующего поколения (NGS — next-generation sequencing) с использованием таргетной панели, включавшей 14 ШГР СД2, было проведено для всех участников исследования. Библиотеки, приготовленные с помощью SeqCap EZ Prime Choice Library (Roche, Швейцария), были отсекуированы на приборе NextSeq 550 (Illumina, США). Все этапы NGS были выполнены в соответствии с протоколами производителей.

Биоинформатический анализ

Парноконцевые прочтения были получены в формате fastq и выровнены на референсный геном GRCh38. Дальнейшая обработка данных и оценка контроля качества были выполнены на основе ранее разработанного пайплайна [19] на базе GATK 3.8 [33]. ШГР были вычислены для генотипов каждого образца путем суммирования эффектов каждого ВНП из оригинальных исследований с учетом количества копий аллеля.

Статистический анализ

Для статистического анализа использовались библиотеки base, stats, dplyr и pROC языка R v. 4.1⁴. Непрерывные признаки описаны медианой и интерквартильным размахом Me [Q25; Q75]. Сравнение непрерывных параметров в независимых выборках выполнено с помощью критерия Манна-Уитни. Анализ ассоциаций ШГР с СД2 проведен при помощи логистической регрессии с поправкой на пол и возраст. В качестве метрики качества для моделей регрессии использована площадь под ROC-кривой (AUC). Уровень статистической значимости принят равным 0,05.

Результаты

Сила ассоциации между включенными в анализ ШГР и наличием СД2 оценена при помощи AUC, которые представлены в таблице 2. AUC для регрессии СД2 на пол и возраст без добавления ШГР составила 77,12%. AUC является стандартной метрикой качества и используется для определения общей производительности модели, благодаря чему можно сравнить результаты данной работы с полученными в исходных исследованиях. В таблице 2 приведены данные по количеству ВНП, включенных в каждую из 14 ШГР, а также количество ВНП, которое было использовано в данном исследовании. Кроме того, приведены данные AUC из оригинальных исследований, при условии их наличия. Результаты анализ AUC приведены и для ШГР, и для моделей, включающих ШГР, пол и возраст).

Из 14 ШГР, включенных в исследование, только две не показали достоверно значимых результатов на российской выборке [25, 30]. Наиболее высокий результат получен для ШГР Langenberg C, et al. (2014) [24] (49 ВНП), Liu SY, et al. (2015) [26] (39 ВНП) и Qi Q, et al. (2017) [29] (80 ВНП). Добавление к этим ШГР общепринятых факторов риска (пол, возраст) значимо увеличивало AUC с 59,05 до 78,75% (p<0,001), с 59,25 до 78,56% (p<0,001), с 59,46 до 78,41% (p<0,001), соответственно. AUC только для ШГР варьировался от 54,49 [23] до 59,46% [29], а AUC для моделей, включающих ШГР, пол и возраст — от 77,56 [23] до 78,75% [24].

К сожалению, данные AUC только для ШГР были указаны лишь в двух оригинальных работах [22, 27], а для моделей с включением ШГР, пола и возраста — в трех [8, 9, 21], что не позволило провести полноценное сравнение между исследованиями.

Обсуждение

В рамках настоящего исследования впервые проведена валидация на российской популяции 14 ШГР СД2, рассчитанных на выборках европейского происхождения и включавших от 296 до 303 053 участников и от 12 до 80 ВНП [8, 9, 21–32]. Включение лишь небольших по размеру ШГР объясня-

⁴ R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria, 2013. <https://www.R-project.org/> (23 October 2023).

Таблица 2

Анализ AUC для СД2 с участием ШГР, пола и возраста в исследуемой выборке (n=1167)

ШГР	Количество ВНП в оригинальном исследовании	Количество ВНП в этом исследовании	AUC в оригинальном исследовании (ШГР)/ШГР + пол + возраст), %	AUC в этом исследовании (ШГР)/ШГР + пол + возраст), %	p для ШГР в этом исследовании
Vassy JL, et al. (2014) [9]	62	62	-/72,6	58,96/78,64	$3,97 \times 10^{-06}$
Walford GA, et al. (2014) [21]	62	62	-/64,1	58,96/78,64	$3,97 \times 10^{-06}$
Cormier H, et al. (2014) [22]	36	34	54,78/-	57,3/77,88	$4,13 \times 10^{-04}$
Villegas R, et al. (2014) [23]	15	12	-/-	54,49/77,56	0,0158
Langenberg C, et al. (2014) [24]	49	49	-/-	59,05/78,75	$4,91 \times 10^{-06}$
Klimentidis YC, et al. (2015) [25]	31	31	-/-	50,65/77,12	0,94
Liu SY, et al. (2015) [26]	39	39	-/-	59,25/78,56	$7,85 \times 10^{-06}$
Talmud PJ, et al. (2015) [27]	65	65	60,2/-	58,74/78,07	$8,13 \times 10^{-06}$
Leong A, et al. (2016) [28]	38	36	-/-	56,27/77,73	$1,22 \times 10^{-03}$
Qi Q, et al. (2017) [29]	80	79	-/-	59,46/78,41	$1,38 \times 10^{-06}$
Liu T, et al. (2018) [30]	12	12	-/-	53,1/77,51	0,134
Szczerbiński L, et al. (2019) [31]	19	19	-/-	56,13/77,97	$2,93 \times 10^{-03}$
Kaplinski M, et al. (2019) [32]	44	44	-/-	57,28/78,04	$1,37 \times 10^{-04}$
Chen X, et al. (2021) [8]	60	59	-/68,1	58,58/78,25	$7,95 \times 10^{-06}$

Примечание: ВНП — вариант нуклеотидной последовательности, ШГР — шкала генетического риска, AUC — площадь под ROC-кривой.

ется тем, что в качестве метода генотипирования использовалось NGS с помощью таргетной панели отдельных участков генома, а не полногеномное секвенирование или высокоплотные микрочипы.

Невозможность валидации ШГР, включавших сотни тысяч и миллионы ВНП, была одним из ограничений этого исследования. Так, например, ШГР из 136795 ВНП была разработана в GWAS Mahajan A, et al. (2018) [10] на выборке UK Biobank (n=441894). С-статистика (аналог AUC) составила 66%, а доля объясненной вариативности СД2 — 18% [10]. В другой работе, выполненной Khera A, et al. (2018) (n=288978), на основе GWAS [34] была создана ШГР из 6 млн ВНП, объясняющая 2,9% вариативности СД2 [35]. В этих исследованиях использовался обобщенный коэффициент детерминации. Этот коэффициент в настоящем исследовании не оценивался, т.к. его сложно интерпретировать и сравнивать между исследованиями [5].

Пять ШГР из 6 исследований [8, 9, 21, 24, 27, 31], включенных в настоящее исследование, были построены на GWAS Morris AP, et al. (2012) [36]. ШГР, созданная на основе этого GWAS в работе Vassy JL, et al. (2014), включала 62 ВНП, AUC для моделей с включением ШГР, пола и возраста составила 72,6% [9]. В настоящем исследовании добавление ШГР к факторам риска (пол, возраст) значительно увеличивало AUC с 77,12 до 78,64% (p=0,009). Эта же шкала была использована в работе Walford GA, et al. (2014), где AUC для модели, включающей только ШГР, составила 64,1% [21]. Кроме того, еще в двух других исследованиях, в которых были разработаны ШГР на основе GWAS [34], были предоставлены данные по AUC: в исследовании Talmud PJ, et al. (2015) AUC

для модели, включающей только ШГР (65 ВНП), составила 60,2% [27], в работе Chen X, et al. (2021) AUC для модели с включением ШГР (60 ВНП), пола и возраста составила 68,1% [8].

Из 14 ШГР, разработанных на европейской популяции, 12 продемонстрировали достоверно значимые результаты на российской выборке. Одним из возможных объяснений отсутствия значимых результатов для двух ШГР является то, что их конструирование проводилось с нарушениями принятой методологии, описанной в работе Choi SW, et al. (2020) [5]. Эти результаты подчеркивают, что воспроизводимы только те ШГР, которые построены при помощи методов статистики, специфических для данной области количественной генетики.

В первой работе с нарушенной методикой, выполненной Klimentidis YC, et al. (2015), изучалась ассоциация генетического риска для повышенного уровня триглицеридов (ТГ) в плазме со случаями СД2 [25]. В частности, ассоциация ШГР для ТГ с СД2 подменена на ассоциацию взаимодействия ШГР для ТГ и СД2 при регрессии на ТГ. Во второй работе Liu T, et al. (2018) разработана ШГР из 12 ВНП, ассоциированных с процентным содержанием жира в организме [30]. В работе игнорировалась поправка на множественные сравнения и использовались невзвешенные ШГР.

Таким образом, представляется крайне важным обязательная валидация на выборках российских пациентов для ШГР, разработанных в других популяциях. При этом важно подчеркнуть, что проведение таких валидаций невозможно без использования данных о пациентах и их биообразцах из коллекций российских биобанков. Проведение

настоящего исследования стало возможным благодаря коллекциям биобанка ФГБУ "НМИЦ ТПМ" Минздрава России [11], содержащих информацию о наличии заболевания СД2, уровне глюкозы натощак и приеме сахароснижающей терапии.

Заключение

Впервые в России было проведено исследование 14 ШГР СД2, разработанных на других попу-

ляциях; 12 ШГР прошли валидацию и в будущем могут быть применены для улучшения прогнозирования риска развития и профилактики СД2 в России. Таким образом, Российские биобанки играют очень важную роль в валидации таких ШГР.

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

Литература/References

1. Kopylova OV, Ershova AI, Meshkov AN, et al. Lifelong prevention of cardiovascular disease. Part I: preconceptional, prenatal and infant periods of life. Cardiovascular Therapy and Prevention. 2020;19(6):2647. (In Russ.) Копылова О.В., Ершова А.И., Мешков А.Н. и др. Профилактика сердечно-сосудистых заболеваний на протяжении жизни. Часть I: преконцепционный, пренатальный и грудной периоды. Кардиоваскулярная терапия и профилактика. 2020;19(6):2647. doi:10.15829/1728-8800-2020-2647.
2. Poulsen P, Kyvik KO, Vaag A, et al. Heritability of Type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance — a population-based twin study. Diabetologia. 1999;42:139-45. doi:10.1007/s001250051131.
3. Willemsen G, Ward KJ, Bell CG, et al. The concordance and heritability of type 2 diabetes in 34,166 twin pairs from international twin registers: the discordant twin (DISCOTWIN) consortium. Twin Res Hum Genet. 2015;18(6):762-71. doi:10.1017/thg.2015.83.
4. Koch S, Schmidtke J, Krawczak M, et al. Clinical utility of polygenic risk scores: a critical 2023 appraisal. J Commun Genet. 2023;1-17. doi:10.1007/s12687-023-00645-z.
5. Choi SW, Mak TSH, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc. 2020;15(9):2759-72. doi:10.1038/s41596-020-0353-1.
6. Reddi HV, Wand H, Funke B, et al. Laboratory perspectives in the development of polygenic risk scores for disease: A points to consider statement of the American College of Medical Genetics and Genomics (ACMG). Genet Med. 2023;25(5):100804. doi:10.1016/j.gim.2023.100804.
7. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015; 518(7538):197-206. doi:10.1038/nature14177.
8. Chen X, Liu C, Si S, et al. Genomic risk score provides predictive performance for type 2 diabetes in the UK biobank. Acta Diabetol. 2021;58(4):467-74. doi:10.1007/s00592-020-01650-1.
9. Vassy JL, Hivert MF, Porneala B, et al. Polygenic type 2 diabetes prediction at the limit of common variant detection. Diabetes. 2014;63(6):2172-82. doi:10.2337/db13-1663.
10. Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet. 2018;50(11):1505-13. doi:10.1038/s41588-018-0241-6.
11. Kopylova OV, Ershova AI, Pokrovskaya MS, et al. Population-nosological research biobank of the National Medical Research Center for Therapy and Preventive Medicine: analysis of bio-samples, principles of collecting and storing information. Cardio-vascular Therapy and Prevention. 2021;20(8):3119. (In Russ.) Копылова О.В., Ершова А.И., Покровская М.С. и др. Популяционно-нозологический исследовательский биобанк "НМИЦ ТПМ": анализ коллекций биообразцов, принципы сбора и хранения информации. Кардиоваскулярная терапия и профилактика. 2021;20(8):3119. doi:10.15829/1728-8800-2021-3119.
12. Anisimov SV, Meshkov AN, Glotov AS, et al. National association of biobanks and biobanking specialists: new community for promoting biobanking ideas and projects in Russia. Biopreserv Biobank. 2021;19:73-82. doi:10.1089/bio.2020.0049.
13. Kurilova OV, Kiseleva AV, Meshkov AN, et al. Scales for assessing the genetic risk of developing type 2 diabetes mellitus. Preventive medicine. 2021;24(12):115-22. (In Russ.) Курилова О.В., Киселева А.В., Мешков А.Н. и др. Шкалы для оценки генетического риска развития сахарного диабета 2-го типа. Профилактическая медицина. 2021;24(12):115-22. doi:10.17116/profmed202124121115.
14. Gola D, Erdmann J, Läll K, et al. Population bias in polygenic risk prediction models for coronary artery disease. Circ-Genom Precis Me. 2020;13(6):e002932. doi:10.1161/CIRCGEN.120.002932.
15. Duncan L, Shen H, Gelaye B, et al. Analysis of polygenic risk score usage and performance in diverse human populations. Nat Commun. 2019;10(1):3328. doi:10.1038/s41467-019-11112-0.
16. Moffatt MF, Phil D, Gut IG, et al. A large-scale, consortium-based genomewide association study of asthma. New Engl J Med. 2010;363(13):1211-21. doi:10.1056/NEJMoa0906312.
17. Usoltsev D, Kolosov N, Rotar O, et al. Understanding complex trait susceptibilities and ethnical diversity in a sample of 4,145 Russians through analysis of clinical and genetic data. bioRxiv. 2023:2023.03.23.534000. doi:10.1101/2023.03.23.534000.
18. Sotnikova EA, Kiseleva AV, Kutsenko VA, et al. Identification of pathogenic variant burden and selection of optimal diagnostic method is a way to improve carrier screening for autosomal recessive diseases. J Pers Med. 2022;12(7):1132. doi:10.3390/jpm12071132.
19. Ramensky VE, Ershova AI, Zaichenok M, et al. Targeted sequencing of 242 clinically important genes in the Russian population from the Ivanovo region. Front Genet. 2021; 12:709419. doi:10.3389/fgene.2021.709419.
20. Boytsov SA, Chazov EI, Shlyakhto EV, et al. Epidemiology of cardiovascular diseases in different regions of Russia (ESSE-RF). The rationale for and design of the study. Preventive medicine. 2013;16(6):25-34. (In Russ.) Бойцов С.А., Чазов Е.И., Шляхто Е.В. и др. Эпидемиология сердечно-сосудистых заболеваний в различных регионах России (ЭССЕ-РФ). Обоснование и дизайн исследований. Профилактическая медицина. 2013;16(6):25-34.
21. Walford GA, Porneala BC, Dauriz M, et al. Metabolite traits and genetic risk provide complementary information for the prediction of future type 2 diabetes. Diabetes Care. 2014;37(9):2508-14. doi:10.2337/dc14-0560.
22. Cormier H, Vigneault J, Garneau V, et al. An explained variance-based genetic risk score associated with gestational diabetes

- antecedent and with progression to pre-diabetes and type 2 diabetes: a cohort study. *BJOG*. 2015;122(3):411-9. doi:10.1111/1471-0528.12937.
23. Villegas R, Goodloe RJ, McClellan BE, et al. Gene-carbohydrate and gene-fiber interactions and type 2 diabetes in diverse populations from the National Health and Nutrition Examination Surveys (NHANES) as part of the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study. *BMC Genet*. 2014;15:69. doi:10.1186/1471-2156-15-69.
24. Langenberg C, Sharp SJ, Franks PW, et al. Gene-lifestyle interaction and type 2 diabetes: the EPIC interact case-cohort study. *PLoS Med*. 2014;11(5):e1001647. doi:10.1371/journal.pmed.1001647.
25. Klimentidis YC, Chougule A, Arora A, et al. Triglyceride-Increasing Alleles Associated with Protection against Type-2 Diabetes. *PLoS Genet*. 2015;11(5):e1005204. doi:10.1371/journal.pgen.1005204.
26. Liu SY, Walter S, Marden J, et al. Genetic vulnerability to diabetes and obesity: does education offset the risk? *Soc Sci Med*. 2015;127:150-8. doi:10.1016/j.socscimed.2014.09.009.
27. Talmud PJ, Cooper JA, Morris RW, et al. Sixty-five common genetic variants and prediction of type 2 diabetes. *Diabetes*. 2015;64(5):1830-40. doi:10.2337/db14-1504.
28. Leong A, Porneala B, Dupuis J, et al. Type 2 diabetes genetic predisposition, obesity, and all-cause mortality risk in the U.S.: a multiethnic analysis. *Diabetes Care*. 2016;39(4):539-46. doi:10.2337/dc15-2080.
29. Qi Q, Stilp AM, Sofer T, et al. Genetics of type 2 diabetes in U.S. hispanic/latino individuals: results from the hispanic community health study/study of latinos (HCHS/SOL). *Diabetes*. 2017;66(5):1419-25. doi:10.2337/db16-1150.
30. Liu T, Li C, Shen L, et al. Heterogeneity in effects of genetically determined adiposity on insulin resistance and type 2 diabetes: The atherosclerosis risk in communities study. *J Diabetes Complicat*. 2018;32(3):330-4. doi:10.1016/j.jdiacomp.2017.12.012.
31. Szczerbiński Ł, Gościk J, Bauer W, et al. Efficacy of family history, genetic risk score, and physical activity in assessing the prevalence of type 2 diabetes. *Pol Arch Intern Med*. 2019;129(7-8):442-50. doi:10.20452/pamw.14866.
32. Kaplinski M, Taylor D, Mitchell LE, et al. The association of elevated maternal genetic risk scores for hypertension, type 2 diabetes and obesity and having a child with a congenital heart defect. *PloS one*. 2019;14(5):e0216477. doi:10.1371/journal.pone.0216477.
33. Van der Auwera, Geraldine A, O'Connor B. *Genomics in the cloud: using Docker, GATK, and WDL in Terra*. O'Reilly Media. 2020. ISBN: 9781491975190.
34. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50(9):1219-24. doi:10.1038/s41588-018-0183-z.
35. Scott RA, Scott LJ, Mägi R, et al. An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes*. 2017;66(11):2888-902. doi:10.2337/db16-1253.
36. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet*. 2012;44(9):981-90. doi:10.1038/ng.2383.