

Шкалы генетического риска для комплексных фенотипов и их применение к российской популяции

Зайченко М.^{1*}, Жарикова А.А.^{2,3}, Вяткин Ю.В.^{2,4}, Киселева А.В.², Ершова А.И.², Сотникова Е.А.², Мешков А.Н.^{2,5,6,7}, Раменский В.Е.^{2,3,8}, Драпкина О.М.²

¹ Московский физико-технический институт (национальный исследовательский университет), Долгопрудный, Московская область, Россия

² Национальный медицинский исследовательский центр терапии и профилактической медицины Минздрава России, Москва, Россия

³ Московский государственный университет им. М.В. Ломоносова, факультет биоинженерии и биоинформатики, Москва, Россия

⁴ Новосибирский государственный университет, Новосибирск, Россия

⁵ Национальный медицинский исследовательский центр кардиологии им. академика Е.И. Чазова Минздрава России, Москва, Россия

⁶ Медико-генетический научный центр им. академика Н.П. Бочкова, Москва, Россия

⁷ Российский национальный исследовательский медицинский университет им. Н.И. Пирогова Минздрава России, Москва, Россия

⁸ Институт перспективных исследований проблем искусственного интеллекта и интеллектуальных систем МГУ им. М.В. Ломоносова, Москва, Россия

* marija.zaichenoka@gmail.com

Ключевые слова: шкалы генетического риска; российская популяция; липидный профиль крови; семейная гиперхолестеринемия; атеросклероз

Мотивация и цель: Сердечно-сосудистые заболевания являются лидирующей причиной смертности в России [1]. Шкалы генетического риска (ШГР) позволяют оценивать индивидуальную предрасположенность развития комплексного (мультифакторного) заболевания до его клинических проявлений. Липидный профиль крови, включающий в себя уровень липопротеинов низкой (ЛПНП) и высокой плотности (ЛПВП), уровень триглицеридов (ТГ) и уровень общего холестерина (ОХС), тесно связан с развитием различных сердечно-сосудистых заболеваний – ишемической болезни сердца, атеросклероза и геморрагического инсульта. В последние годы было проведено множество разных полногеномных исследований, посвященных поиску ассоциаций вариантов генома с уровнями липидов крови [2, 3], на основе которых можно строить различные ШГР. Целью настоящей работы является оценка применимости опубликованных ранее ШГР для фенотипов липидного профиля к российской популяции, поиск оптимального метода применения данных ШГР и оценка применимости ШГР в клинической практике.

Методы и алгоритмы: В исследовании используются три выборки:

- популяционная выборка из Ивановской области, состоящая из 1673 человек (624 мужчины, 1049 женщин), собранная в ходе исследования ЭССЕ-РФ [4]. Медианный возраст выборки составляет 50 лет. Выборка была секвенирована с использованием таргетной панели, содержащей 242 гена и более 2000 участков генома с клинически важными однонуклеотидными вариантами, входящими в опубликованные ШГР;

- популяционная выборка из Ивановской области, состоящая из 817 человек (346 мужчин, 417 женщин), собранная во время исследования ЭССЕ-РФ [4]. Медианный возраст выборки составляет 47 лет. Выборка была секвенирована с использованием таргетной панели, содержащей 217 генов и более 18 000 участков генома с клинически важными однонуклеотидными вариантами;
- выборка больных семейной гиперхолестеринемией (СГХС), состоящая из 372 человек (107 мужчин, 265 женщин). Медианный возраст выборки составляет 57 лет. Выборка была секвенирована с помощью полногеномного секвенирования.

Для каждого участника были измерены показатели липидного профиля – ЛПНП, ЛПВП, ОХС и ТГ. В рамках субисследования АТЕРОГЕН-Иваново участникам исследования ЭССЕ-РФ по Ивановской области было проведено УЗИ сонных и бедренных артерий. Данные обследования имеются для 61.5 % участников выборки из Ивановской области. Имеются данные о стенозе, высоте атеросклеротических бляшек, их количество, толщина интима-медиа.

Для построения ШГР были использованы данные из трех различных статей [2, 3, 6], включающие в себя от 1308 до 2 432 587 вариантов для различных липидных фенотипов.

Рассматривались три способа применения ШГР:

- использование ШГР «как есть» на основе исходных оценок эффекта всех значимых вариантов;
- поиск оптимальной ШГР с помощью подхода C+T (clumping & thresholding), реализованного в методе PRSice-2 [7];
- расчет ШГР с использованием LDpred-2 [8].

Используя имеющиеся выборки, был проведен расчет R^2 (процент объясняемой дисперсии) полученных ШГР для фенотипов ЛПНП, ЛПВП, ТГ, ОХС.

Полученные шкалы были использованы для классификации больных/здоровых СГХС, а также для изучения связи наличия патогенного моногенного варианта и значения ШГР. Была произведена оценка возможности использования ШГР для липидного профиля для предсказания результатов исследования УЗИ бедренных и сонных артерий.

Результаты: В рамках работы было проанализировано по 20 ШГР для четырех показателей липидного профиля для Ивановской и Вологодской выборок соответственно: 12 шкал «как есть» и 8 разработанных методами PRSice-2 и LDpred-2. Результаты R^2 варьировали от 3.5 до 12 % (что совпадает с литературными данными) для различных фенотипов, наилучшие результаты были показаны при построении ШГР методом LDpred-2.

Показано, что распределения ШГР для ЛПНП и ОХС у больных СГХС и здорового контроля из Ивановской области значимо различаются. Более того, распределения значений ШГР для ОХС и ЛПНП значимо различаются между больными – носителями патогенных вариантов в генах *LDLR*, *APOB* и *PCSK9*, и больными неносителями.

Выводы: В рамках работы была оценена эффективность применения ШГР, построенных на основе данных для популяций, отличных от популяций Российской Федерации. Также была оценена возможность клинической применимости ШГР для различных заболеваний, в частности, СГХС и атеросклероза.

Polygenic risk scores for complex phenotypes and their application to the Russian population

Zaichenoka M.^{1*}, Zharikova A.A.^{2,3}, Vyatkin Yu.V.^{2,4}, Kiseleva A.V.², Ershova A.I.², Sotnikova E.A.², Meshkov A.N.^{2,5,6,7}, Ramensky V.E.^{2,3,8}, Drapkina O.M.²

¹ *Moscow Institute of Physics and Technology (National Research University), Dolgoprudny, Moscow region, Russia*

² *National Medical Research Center for Therapy and Preventive Medicine of the Ministry of Health of Russia, Moscow, Russia*

³ *Lomonosov Moscow State University, Faculty of Bioengineering and Bioinformatics, Moscow, Russia*

⁴ *Novosibirsk State University, Novosibirsk, Russia*

⁵ *National Medical Research Center of Cardiology named after academician E.I. Chazov of the Ministry of Health of Russia, Moscow, Russia*

⁶ *Medical Genetic Research Center named after academician N.P. Bochkov, Moscow, Russia*

⁷ *Russian National Research Medical University named after N.I. Pirogov of Ministry of Health of Russia, Moscow, Russia*

⁸ *Institute for Advanced Research on Artificial Intelligence and Intelligent Systems, Lomonosov Moscow State University, Moscow, Russia*

* *marija.zaichenoka@gmail.com*

Key words: genetic risk scales; Russian population; blood lipid profile; familial hypercholesterolemia; atherosclerosis

Motivation and Aim: Cardiovascular diseases are the leading cause of mortality in Russia [1]. Polygenic risk scores (PRS) allow one to assess an individual's susceptibility to the development of a complex (multifactorial) disease before its clinical manifestations. The blood lipid profile, including low-density lipoprotein (LDL) and high-density lipoprotein (HDL) levels, triglyceride (TG) levels and total cholesterol (TC) levels, is closely associated with the development of various cardiovascular diseases – coronary heart disease, atherosclerosis and hemorrhagic stroke. In recent years, many different genome-wide studies have been conducted to search for associations of genome variants with blood lipid levels [2, 3], on the basis of which various PRS can be constructed. The purpose of this work is to assess the applicability of previously published PRS for lipid profile phenotypes to the Russian population, to find the optimal method for using PRS data and to assess the applicability of PRs in clinical practice.

Methods and Algorithms: This study uses three samples:

- A population sample from the Ivanovo region, consisting of 1,673 people (624 men, 1,049 women), collected during the ESSE-RF study [4]. The median age of the sample is 50 years. The sample was sequenced using a targeted panel containing 242 genes and more than 2,000 genomic regions with clinically important single nucleotide variants included in the published PRS;
- A population sample from the Ivanovo region, consisting of 817 people (346 men, 471 women), collected during the ESSE-RF study [4]. The median age of the sample is 47 years. The sample was sequenced using a targeted panel containing 217 genes and more than 18,000 genomic regions with clinically important single nucleotide variants;

- A sample of patients with familial hypercholesterolemia (FH), consisting of 372 people (107 men, 265 women). The median age of the sample is 57 years. The sample was sequenced using whole genome sequencing.

For each participant, lipid profile indicators were measured – LDL, HDL, total cholesterol and TG. As part of the ATEROGEN-Ivanovo substudy, participants in the ESSE-RF study in the Ivanovo region underwent ultrasound of the carotid and femoral arteries. Survey data are available for 61.5 % of sample participants from the Ivanovo region. There is data on stenosis, the height of atherosclerotic plaques, their number, and intima-media thickness.

To construct the PRS, data from three different articles were used [2, 3, 6], including from 1,308 to 2,432,587 variants for various lipid phenotypes.

Three methods of using PRS were considered:

- Application of the PRS “as is” based on initial estimates of the effect of all significant variants;
- Finding the optimal PRS using the C+T (clumping & thresholding) approach implemented in the PRSice-2 method [7];
- Calculation of PRS using LDpred-2 [8].

Using the available samples, R^2 was calculated (percentage of explained variance) of the obtained PRS for the LDL, HDL, TG, and TC phenotypes.

The resulting scales were used to classify patients/healthy patients with FH, as well as to study the relationship between the presence of a pathogenic monogenic variant and the value of PRS. The feasibility of using lipid profile PRS to predict femoral and carotid artery ultrasound findings was assessed.

Results: As part of the work, 20 PRS were analyzed for 4 lipid profile indicators for the Ivanovo and Vologda samples, respectively: 12 “as is” scales and 8 developed by the PRSice-2 and LDpred-2 methods. R^2 of resulting PRS varied from 3.5 to 12 % (which coincides with the literature data) for various phenotypes, the best results were shown when constructing the PRS using the LDpred-2 method.

It was shown that the distributions of PRS for LDL and TC in patients with FH and healthy controls from the Ivanovo region are significantly different. Moreover, the distributions of PRS values for total cholesterol and LDL differ significantly between patients who are carriers of pathogenic variants in the genes *LDLR*, *APOB* and *PCSK9*, and sick non-carriers.

Conclusion: As part of the work, the effectiveness of using the PRS, built on the basis of data for populations other than the populations of the Russian Federation, was assessed. The possibility of clinical applicability of PRS for various diseases, in particular FH and atherosclerosis, was also assessed.

Список литературы/References

1. Боровкова Н.Ю., Токарева А.С., Савицкая Н.Н., Крисанова К.И., Курашин В.К., Одинцов Г.А. Современное состояние проблемы сердечно-сосудистых заболеваний в Нижегородском регионе: возможные пути снижения смертности. *Российский кардиологический журнал*. 2022;27(5):5024. doi 10.15829/1560-4071-2022-5024
2. Selvaraj M.S., Li X., Li Z. et al. Whole genome sequence analysis of blood lipid levels in >66,000 individuals. *Nat Commun*. 2022;13:5995. doi10.1038/s41467-022-33510-7
3. Xu Y., Ritchie S.C., Liang Y. et al. An atlas of genetic scores to predict multi-omic traits. *Nature*. 2023;616:123-131. doi 10.1038/s41586-023-05844-9
4. Бойцов С.А., Драпкина О.М., Шляхто Е.В., Конради А.О., Баланова Ю.А., Жернакова Ю.В., Метельская В.А., Ощепкова Е.В., Ротарь О.П., Шальнова С.А. Исследование ЭССЕ-РФ (Эпидемиология сердечно-сосудистых заболеваний и их факторов риска в регионах Российской

- Федерации). Десять лет спустя. *Кардиоваскулярная терапия и профилактика*. 2021;20(5):3007. doi 10.15829/1728-8800-2021-3007
5. Ершова А.И., Балахонова Т.В., Мешков А.Н., Куценко В.А., Яровая Е.Б., Шальнова С.А., Лищенко Н.Е., Новикова А.С., Александрова Е.Л., Шутемова Е.А., Белова О.А., Рачкова С.А., Бойцов С.А., Драпкина О.М. Распространенность атеросклероза сонных и бедренных артерий среди населения Ивановской области: исследование АТЕРОГЕН-Иваново. *Кардиоваскулярная терапия и профилактика*. 2021;20(5):2994. doi 10.15829/1728-8800-2021-2994
 6. Willer C.J., Schmidt E.M., Sengupta S. et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45(11):1274-1283. doi 10.1038/ng.2797
 7. Choi S.W., O'Reilly P.F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience*. 2019;8(7):giz082. doi 10.1093/gigascience/giz082
 8. Privé F., Arbel J., Vilhjálmsson B.J. LDpred2: better, faster, stronger. *Bioinformatics*. 2021;36(22-23):5424-5431. doi 10.1093/bioinformatics/btaa1029