

# Геногеографические технологии популяционного биобанка как инструмент оценки эффектов отбора (на примере фармакогенетических биомаркеров сердечно-сосудистых заболеваний)

Балановская Е. В.<sup>1</sup>, Горин И. О.<sup>1</sup>, Пономарев Г. Ю.<sup>1</sup>, Пылев В. Ю.<sup>1,2</sup>, Белов Р. О.<sup>1</sup>, Почешхова Э. А.<sup>1,3</sup>, Абдуллаев Ш. П.<sup>4</sup>, Мирзаев К. Б.<sup>4</sup>, Сычев Д. А.<sup>4</sup>

<sup>1</sup>ФГБНУ "Медико-генетический научный центр им. акад. Н.П. Бочкова". Москва; <sup>2</sup>АНО "Биобанк Северной Евразии". Москва;

<sup>3</sup>ФГБОУ ВО "Кубанский государственный медицинский университет" Минздрава России. Краснодар; <sup>4</sup>ФГБОУ ДПО "Российская медицинская академия непрерывного профессионального образования" Минздрава России. Москва, Россия

Значительные различия между генофондами народов России требуют разработки этно-регионально адаптированных фармакогенетических тестов и выявления приоритетных регионов для их внедрения.

**Цель.** Разработать геногеографическую технологию для выявления эффектов отбора на примере биомаркеров, значимых для фармакотерапии пациентов с сердечно-сосудистыми заболеваниями (ССЗ), с помощью популяционного биобанка и базы данных "Фармакогенетика популяций России и сопредельных стран".

**Материал и методы.** Образцы дезоксирибонуклеиновой кислоты (ДНК) Биобанка Северной Евразии из 20 метапопуляций коренного населения Европейской части России изучены по двум массивам данных: по 24 фармакогенетическим маркерам ССЗ (3170 образцов); по 1276 191 полиморфным ДНК-маркерам аутосомного генома (1293 образца). По каждому массиву данных приведены оценки межпопуляционной изменчивости генофонда: разница между этими оценками характеризует давление отбора на каждый из 24 биомаркеров ССЗ. Создан геногеографический атлас, карты которого демонстрируют давление отбора на каждый биомаркер по степени отклонения от селективно-нейтральной изменчивости генофонда.

**Результаты.** 24 биомаркера ССЗ подразделены на три класса: близких к селективно-нейтральной изменчивости, подверженных стабилизирующему и дифференцирующему отбору. Для каждого из 24 биомаркеров ССЗ созданы геногеографические карты, выявляющие эффекты отбора в каждой из 20 метапопуляций. Большинство карт выявили популяции, находящиеся под давлением дифференцирующего отбора и поэтому являющиеся приоритетными для внедрения этно-регионально адаптированных фармакогенетических протоколов.

**Заключение.** Фармакогенетические маркеры и популяции, находящиеся под действием дифференцирующего отбора, требуют разработки этно-регионально адаптированных фармакогенети-

ческих тестов. Созданный картографический атлас отбора может служить основой для развития фармакогенетических исследований, проводимых методами геногеографии.

**Ключевые слова:** отбор, популяционный биобанк, база данных, фармакогенетика, ДНК-маркеры, сердечно-сосудистые заболевания, геногеография.

**Отношения и деятельность.** Исследование выполнено при поддержке гранта РНФ № 21-14-00363 (биоинформатический анализ, интерпретация результатов) и государственного задания Минобрнауки России для ФГБНУ "МГНЦ" (картографический и статистический анализ).

**Благодарности.** Авторы благодарят АНО "Биобанк Северной Евразии" за предоставление коллекций ДНК.

Поступила 11/10-2023

Рецензия получена 16/10-2023

Принята к публикации 25/10-2023



**Для цитирования:** Балановская Е. В., Горин И. О., Пономарев Г. Ю., Пылев В. Ю., Белов Р. О., Почешхова Э. А., Абдуллаев Ш. П., Мирзаев К. Б., Сычев Д. А. Геногеографические технологии популяционного биобанка как инструмент оценки эффектов отбора (на примере фармакогенетических биомаркеров сердечно-сосудистых заболеваний). *Кардиоваскулярная терапия и профилактика*. 2023; 22(11):3773. doi:10.15829/1728-8800-2023-3773. EDN XPMFOW

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

e-mail: balanovska@mail.ru

[Балановская Е. В. — д.б.н., профессор, г.н.с., зав. лабораторией популяционной генетики человека, ORCID: 0000-0002-3882-8300, Горин И. О. — н.с. лаборатории популяционной генетики человека, ORCID: 0000-0001-9532-8954, Пономарев Г. Ю. — н.с. лаборатории популяционной генетики человека, ORCID: 0000-0002-0643-4374, Пылев В. Ю. — н.с. лаборатории популяционной генетики человека, генеральный директор АНО "Биобанк Северной Евразии", ORCID: 0000-0001-9541-8319, Белов Р. О. — стажер сотрудник лаборатории популяционной генетики человека, ORCID: 0009-0009-4286-4791, Почешхова Э. А. — д.м.н., г.н.с. лаборатории популяционной генетики человека, доцент, профессор кафедры биологии с курсом медицинской генетики, ORCID: 0000-0002-8991-7194, Абдуллаев Ш. П. — к.б.н., зав. отделом предиктивных и прогностических биомаркеров НИИ молекулярной и персонализированной медицины, ORCID: 0000-0001-9001-1499, Мирзаев К. Б. — д.м.н., директор НИИ молекулярной и персонализированной медицины, проректор по научной работе и инновациям, ORCID: 0000-0002-9307-4994, Сычев Д. А. — академик РАН, д.м.н., профессор, профессор РАН, зав. кафедрой клинической фармакологии и терапии им. акад. Б. Е. Вотчала, ректор, ORCID: 0000-0002-4496-3680].

## Genogeographic technologies of a population biobank as a tool for assessing selection effects (using the example of pharmacogenetic biomarkers of cardiovascular diseases)

Balanovskaya E. V.<sup>1</sup>, Gorin I. O.<sup>1</sup>, Ponomarev G. Yu.<sup>1</sup>, Pylev V. Yu.<sup>1,2</sup>, Belov R. O.<sup>1</sup>, Pocheshkhova E. A.<sup>1,3</sup>, Abdullaev Sh. P.<sup>4</sup>, Mirzaev K. B.<sup>4</sup>, Sychev D. A.<sup>4</sup>

<sup>1</sup>Bochkov Medical Genetic Research Center. Moscow; <sup>2</sup>Biobank of Northern Eurasia. Moscow; <sup>3</sup>Kuban State Medical University. Krasnodar;

<sup>4</sup>Russian Medical Academy of Continuous Professional Education. Moscow, Russia

Significant differences between the gene pools of Russian peoples require the development of ethno-regional adapted pharmacogenetic tests and the identification of priority regions for their implementation.

**Aim.** To develop a genogeographic technology to identify selection effects using the example of biomarkers that are significant for pharmacotherapy of patients with cardiovascular diseases (CVD), using a population biobank and the Pharmacogenetics of Populations of Russia and Adjacent Countries database.

**Material and methods.** Deoxyribonucleic acid (DNA) samples from the Biobank of Northern Eurasia from 20 metapopulations of the indigenous population of the European Russia were studied using two following data sets: 24 pharmacogenetic markers of CVDs (3170 samples); 1 276 191 polymorphic DNA markers of the autosomal genome (1293 samples). For each data set, estimates of interpopulation variability in the gene pool are provided — the difference between these estimates characterizes the selection pressure on each of the 24 CVD biomarkers. A genogeographic atlas has been created, the maps of which demonstrate the selection pressure on each biomarker according to the degree of deviation from the selective-neutral variability of the gene pool.

**Results.** Twenty-four CVD biomarkers are divided into three following classes: those close to selective-neutral variability, those subject to stabilizing and differentiating selection. For each of the 24 CVD biomarkers, genogeographic maps were created that reveal selection effects in each of the 20 metapopulations. Most maps have identified populations that are under differential selection pressure and therefore a priority for the implementation of ethno-regionally adapted pharmacogenetic protocols.

**Conclusion.** Pharmacogenetic markers and populations under differential selection require the development of ethno-regionally adapted pharmacogenetic tests. The created cartographic atlas of selection can serve as the basis for pharmacogenetic studies carried out using genogeographic methods.

**Keywords:** selection, population biobank, database, pharmacogenetics, DNA markers, cardiovascular diseases, genogeography.

**Relationships and Activities.** The study was supported by the Russian Science Foundation grant № 21-14-00363 (bioinformatic analysis, interpretation of results) and the state assignment of the Ministry of Education and Science of Russia for the Bochkov Medical Genetic Research Center (cartographic and statistical analysis).

**Acknowledgments.** The authors are grateful to the Biobank of Northern Eurasia for providing DNA collections.

Balanovskaya E. V.\* ORCID: 0000-0002-3882-8300, Gorin I. O. ORCID: 0000-0001-9532-8954, Ponomarev G. Yu. ORCID: 0000-0002-0643-4374, Pylev V. Yu. ORCID: 0000-0001-9541-8319, Belov R. O. ORCID: 0009-0009-4286-4791, Pocheshkhova E. A. ORCID: 0000-0002-8991-7194, Abdullaev Sh. P. ORCID: 0000-0001-9001-1499, Mirzaev K. B. ORCID: 0000-0002-9307-4994, Sychev D. A. ORCID: 0000-0002-4496-3680.

\*Corresponding author: balanovska@mail.ru

**Received:** 11/10-2023

**Revision Received:** 16/10-2023

**Accepted:** 25/10-2023

**For citation:** Balanovskaya E. V., Gorin I. O., Ponomarev G. Yu., Pylev V. Yu., Belov R. O., Pocheshkhova E. A., Abdullaev Sh. P., Mirzaev K. B., Sychev D. A. Genogeographic technologies of a population biobank as a tool for assessing selection effects (using the example of pharmacogenetic biomarkers of cardiovascular diseases). *Cardiovascular Therapy and Prevention*. 2023;22(11):3773. doi:10.15829/1728-8800-2023-3773. EDN XPMFOW

ДНК — дезоксирибонуклеиновая кислота, ССЗ — сердечно-сосудистые заболевания, ФГТ — фармакогенетическое тестирование, SNP — Single Nucleotide Polymorphism (однонуклеотидные полиморфные сайты генома),  $F_{ST}$  — критерий Фишера для оценки межпопуляционных различий, IHS — integrated haplotype score (интегрированная оценка гаплотипов — метод на основе анализа расширенной гаплотипической гомозиготности),  $N_e$  — показатель генетически эффективного размера популяций,  $M_e$  — показатель генетически эффективных миграций.

## Введение

Получение генетической информации о пациенте для индивидуализации фармакотерапии — выбора лекарственных препаратов, подбора и коррекции их доз на основании молекулярно-генетических особенностей пациента — становится важным трендом в медицине. Особые успехи фармакогенетических исследований достигнуты в кардиологической практике. Примером могут служить такие пары ген-препарат: *CYP2C9*, *VKORC1*, *CYP4F2* [1] — варфарин (и другие антагонисты витамина К); *CYP2C19* и *ABCB1* — клопидогрел [2, 3]; *SLCO1B1* — статины [4]. Полиморфизмы *ABCB1* детерминируют также изменение фармакокинетики антикоагулянтов прямого действия (дабигатран, ривароксабан, апиксабан) [5].

Современная медицинская помощь основывается на национальных и международных клинических рекомендациях "усредненному пациенту" [6], характеристики которого определены в рандомизированных клинических исследованиях жителей Европы и США. Однако основанные на них фармакогенетические рекомендации могут быть не валидными для населения других регионов мира [7], что требует от национальных органов здравоохранения их адаптации к собственному народонаселению. Знание геногеографии клинически значимых биомаркеров позволяет выделять популяции, для которых проведение фармакогенетического тестирования (ФГТ) будет приоритетным и обоснованным; такой подход исключит необоснованное назначение ФГТ для всех пациентов.

**Ключевые моменты****Что известно о предмете исследования?**

- Значительные различия между генофондами народов России требуют этно-региональной адаптации фармакогенетических тестов и выявления приоритетных регионов для их внедрения.
- Наиболее важны для решения этой задачи те фармакогенетические ДНК-маркеры и популяции, которые находятся под давлением дифференцирующего отбора.

**Что добавляют результаты исследования?**

- Сочетание геногеографической технологии, популяционного биобанка и базы данных по фармакогенетике позволило создать картографический атлас эффектов отбора для 24 фармакогенетических биомаркеров сердечно-сосудистых заболеваний.
- Карты атласа показывают, в каких популяциях России и по каким фармакогенам действует дифференцирующий отбор, и демонстрируют приоритетные популяции для внедрения этно-регионально адаптированных фармакогенетических протоколов.

**Key messages****What is already known about the subject?**

- Significant differences between the gene pools of the Russian peoples require ethno-regional adaptation of pharmacogenetic tests and identification of priority regions for their implementation.
- The most important for solving this problem are those pharmacogenetic DNA markers and populations that are under differential selection pressure.

**What might this study add?**

- The combination of genogeographic technology, a population biobank and a pharmacogenetics database has produced a cartographic atlas of selection effects for 24 pharmacogenetic cardiovascular biomarkers.
- Atlas maps show in which Russia populations and for which pharmacogenes differential selection operates, and demonstrate priority populations for the implementation of ethno-regionally adapted pharmacogenetic protocols.

Во многих работах показано, что для народонаселения России характерна значительная этно-региональная изменчивость частот фармакогенетических маркеров [8-14], что приводит к необходимости адаптации фармакогенетических рекомендаций к конкретным популяциям и народам. Для выявления биомаркеров, требующих наибольшей адаптации ФГТ из-за значительных различий между популяциями, полезно использовать такой инструмент, как выявление эффектов отбора. Под давлением мощного стабилизирующего отбора межпопуляционные различия биомаркеров будут нивелироваться и тогда для большинства популяций региона будет приемлем стандартный протокол ФГТ. Но у той части биомаркеров, на которые действует дифференцирующий отбор, различия между популяциями становятся столь велики, что для них необходима разработка разных вариаций фармакогенетических протоколов. В этом случае целесообразно применение инструментария геногеографии, предлагающей легко читаемые карты, указывающие на действие разных типов отбора и его интенсивность в разных частях региона.

Наиболее простой способ выявления отбора — анализ отклонения межпопуляционных различий по данному биомаркеру от селективно-нейтрального уровня. Основные факторы микроэволюции популяций — дрейф генов и миграции — действуют на весь аутосомный геном одинаково и определяют единый для всех SNP (Single Nucleotide Poly-

morphism)-маркеров селективно-нейтральный уровень различий между популяциями. Поэтому увеличение таких межпопуляционных различий у конкретного биомаркера указывает на давление на него дифференцирующего отбора, а уменьшение различий — на действие стабилизирующего отбора [15-19].

Одним из наиболее применяемых методов для оценки межпопуляционных различий служит показатель  $F_{ST}$  (критерий Фишера для оценки межпопуляционных различий). Разработаны и описаны различные подходы к поиску значений  $F_{ST}$ , отличающихся от селективно-нейтрального уровня различий между популяциями [20]: ранжируя оценки  $F_{ST}$  по отдельным SNP-маркерам, приходят к выводу о влиянии отбора на маркеры с наибольшим отклонением от среднего уровня  $F_{ST}$  [21]. Этот метод применялся, например, для поиска маркеров, подверженных отбору, в популяциях афроамериканцев [22], в 7 популяциях североευропейского происхождения [23], для оценки функциональной значимости SNP-маркеров в массиве данных II фазы международного проекта НарМар (сокращенно от Narplotype Map — "карта гаплотипа" генома человека) [24].

Данный метод выявления эффектов отбора неоднократно использовался и для фармакогенетически значимых биомаркеров. Например, при анализе SNP-маркеров гена *NOS1* (играющего решающую роль в иммунорегуляции и в неспецифической защите хозяина при таких инфекциях, как

малярия, токсоплазмоз, лейшманиоз, трипаносомоз, шистосомоз) сравнение шести популяций Средиземноморья выявило отклонение от селективной нейтральности в марокканской популяции [25]. При оценке межпопуляционной вариабельности частот SNP-маркеров генов иммунного ответа (*IL4*, *IL10*, *IL13*, *IL12A*) показано действие дифференцирующего отбора на варианты rs2070874 и rs6441286 про- и противовоспалительных цитокинов в разных популяциях [26]. При анализе 1044 SNP-маркеров 243 фармакогенов с помощью двух независимых методов — IHS (integrated haplotype score, интегрированная оценка гаплотипов) и  $F_{ST}$  — обнаружено действие отбора [27] на фармакогены различных классов: *ABCC6* (ген-переносчик многих противоопухолевых препаратов); *UGT2B17* (ген, кодирующий фермент переноса глюкуроновой кислоты из уридиндифосфоглюкуроновой кислоты в разнообразные субстраты, включая стероидные гормоны и жирорастворимые лекарственные средства); *SLC25A20* (ген, ассоциированный с дефицитом карнитин-ацилкарнитинтранслоказы, мутации в котором могут вызывать различные патологические состояния, такие как гипогликемия, остановка сердца, гепатомегалия, печеночная дисфункция и мышечная слабость).

Эти и другие исследования указывают, что выявление эффектов отбора через анализ межпопуляционных различий служит эффективным инструментом для обнаружения генов-кандидатов, имеющих наибольшее функциональное и медицинское значение для ряда популяций. Выявление фармакогенетических маркеров и популяций, наиболее подверженных дифференцирующему отбору, может стать важным инструментом в рамках стратегии перехода к персонализированной медицине при разработке региональных алгоритмов персонализации терапии различных нозологий, учитывающих генетические особенности местного населения, и для выделения отдельных регионов или популяций приоритетных для внедрения соответствующих протоколов ФГТ.

Цель работы — с помощью геногеографических технологий, опирающихся на популяционный биобанк и базу данных по фармакогенетике, оценить давление отбора на биомаркеры, значимые для персонализации фармакотерапии пациентов с сердечно-сосудистыми заболеваниями (ССЗ).

## Материал и методы

Исследование основано на биообразцах популяционного Биобанка Северной Евразии (Москва) и информации базы данных "Фармакогенетика популяций России и сопредельных стран". Биобанк Северной Евразии предоставил 3170 образцов дезоксирибонуклеиновой кислоты (ДНК), которые были изучены по 24 биомаркерам ССЗ: *ABCB1* (rs1045642), *ABCB1* (rs4148738), *ABCG2* (rs2231137), *ABCG2* (rs2231142), *ADRB1* (rs1801252), *CES1* (rs2244613),

*CYP2C19* (rs12248560), *CYP2C19* (rs4244285), *CYP2C19* (rs4986893), *CYP2C8* (rs10509681), *CYP2C9* (rs1057910), *CYP2C9* (rs1799853), *CYP2D6* (rs1065852), *CYP2D6* (rs28371725), *CYP2D6* (rs3892097), *CYP3A4* (rs2242480), *CYP3A5* (rs776746), *CYP4F2* (rs2108622), *P2RY12* (rs2046934), *PON1* (rs662), *SLCO1B1* (rs4149056), *VKORC1* (rs7294), *VKORC1* (rs9923231), *VKORC1* (rs9934438). Данные об их частотах в 61 популяции Европейской части России и сопредельных стран объединены с дополнительной информацией базы данных "Фармакогенетика популяций России и сопредельных стран".

Для анализа селективно-нейтральной изменчивости Биобанк Северной Евразии предоставил данные о 1401 образце, изученном по большому массиву SNP-маркеров аутосомного генома с помощью биочипа Infinium OmniExome BeadChip Kit (Illumina; США) на приборе iScan (Illumina; США). Первичный анализ и оценка качества этих данных проведены в программе GenomeStudio v2011.1, показатель CallRate составлял не <0,99. При фильтрации с помощью PLINK 1.9 [27] исключены SNP-маркеры, прочитанные менее чем у 90% образцов (geno 0,1), с частотой минорного варианта <1% (maf 0,01), тесно сцепленные SNP ( $R^2 > 0,5$ ); а также исключены образцы, у которых прочитано <95% SNP-маркеров (mind 0,05). В результате для анализа использованы 1276 191 SNP-маркеров для 1293 образцов.

Оба массива данных Биобанка Северной Евразии ДНК репрезентативно представляют генофонды коренного населения европейской части России (таблица 1). Коллекции ДНК сформированы по единой технологии (детально описанной в [28]) в ходе многолетних экспедиционных обследований коренного населения: в выборки включены образцы неродственных между собой индивидов, предки которых на глубину не <3-х поколений относили себя к данному этносу и данной популяции [28]. Обследование проведено на добровольной основе с письменным информированным согласием, одобренным Этическим комитетом ФГБНУ "МГНЦ" (Заключение от 29.06.2020г).

Для корректного сравнения и формирования репрезентативных выборок оба массива популяционных данных объединены в 20 метапопуляций в соответствии с общностью этногенеза или региональной принадлежности (таблица 1). Генетические различия  $F_{ST}$  между каждой парой из 20 метапопуляций рассчитаны с помощью программы smartpca пакета eigensoft [29]. Геногеографические карты построены с помощью программного пакета GeneGeo [30] методом средневзвешенной интерполяции со степенью весовой функции 2 и радиусом влияния 1200 км. В специальном окне легенды каждой карты указаны ее статистические характеристики: K — число исходных популяций для построения карты; min — минимальное значение признака; max — максимальное значение признака; avg — среднее значение признака; GST — межпопуляционные различия (аналог  $F_{ST}$  для дуаллельных маркеров).

## Результаты

Чтобы выявить давление отбора на биомаркер, необходимо получить две оценки: селективно-нейтрального уровня межпопуляционной изменчивости для данного региона и межпопуляционной изменчивости данного биомаркера в этом же регионе.



Таблица 1

Состав и численность популяций, изученных по биомаркерам ССЗ, и метапопуляций, по которым дана оценка селективно-нейтральной изменчивости

Популяции ССЗ			Метапопуляции	
Номера популяций на карте частоты ССЗ	Состав популяций	Выборка ССЗ, n (человек)	Краткое условное название метапопуляции на карте	Выборка в метапопуляциях, m (человек)
1-4	Русский Север (Архангельская область)	213	Поморы	84
5-7	русские северных областей	205	Русс-север	95
8-10	русские центральных областей	142	Русс-центр	84
11, 12, 17	русские южных областей, казаки	166	Русс-юг	88
13-16	украинцы	206	Украинцы	78
18	белорусы	50	—	—
19, 23	вепсы, ижора, водь	87	Вепсы	30
20-22	карелы, саамы	157	Карелы	30
24	финны-ингерманландцы	37	—	—
25	ненцы	21	—	—
26	удмурты, бесермяне	90	Удмурты	28
27-29	коми-пермяки	218	Коми-пермяки	47
30, 31	марийцы	129	Марийцы	31
32-34	чуваши	134	Чуваши	33
35, 36	татары Поволжья	156	Татары	54
37, 38	Мордовия (мокша, шокша, эрзя)	155	Мордовия	47
39, 40	башкиры	100	Башкиры	42
41, 42	угры (манси, ханты)	96	Угры	41
43-46	народы Западного Кавказа	167	Запад. Кавказ	86
47-48	народы европейской степи	95	Степь	83
49-50, 55-58	народы Центрального Кавказа и Закавказья	265	Центр. Кавказ	193
51-54	народы Восточного Кавказа	155	Вост. Кавказ	88
59-60	народы Зарубежной Европы	90	Заруб. Европа	31
61	калмыки	36	—	—
	Суммарно	3170	Суммарно	1293

Примечание: ССЗ — сердечно-сосудистые заболевания.

### 1. Оценка селективно-нейтрального уровня межпопуляционных различий $F_{ST}$

Для генофонда каждого региона характерен свой селективно-нейтральный уровень различий между популяциями:  $F_{ST(NEUTR)}$ . Его значение одинаково для любых генетических маркеров данного региона, поскольку он задается не свойствами того или иного гена, а исключительно демографическими факторами — средним генетически эффективным размером популяций ( $N_e$ ) и миграционными связями между популяциями региона ( $M_e$ ). Эти два демографических фактора и являются двумя основными, но противодействующими, силами микроэволюции популяций, которые в равной степени действуют на любые гены аутосомного генома. Чем меньше генетически эффективный размер популяции ( $N_e$ ), тем сильнее дрейф генов, увеличивающий различия между популяциями. Чем интенсивнее генетически эффективные миграции ( $M_e$ ), тем больше нивелируются различия между по-

пуляциями, возникающие в результате дрейфа генов. Их взаимодействие и определяет селективно-нейтральный уровень межпопуляционных различий  $F_{ST(NEUTR)}=1/(4N_eM_e)$ .

Есть два способа определения уровня селективно-нейтральной изменчивости  $F_{ST(NEUTR)}$  [15-19]. Прямой способ — через исключительно демографические оценки генетически эффективных размера популяций  $N_e$  и миграций  $M_e$ . Но корректные оценки  $N_e$  и  $M_e$  даже для небольшого региона получить крайне сложно: для их получения используются разные модели популяционной генетики, вызывающие вопросы об их применимости в каждом конкретном случае. Поэтому более надежным оказывается непрямой способ оценки  $F_{ST(NEUTR)}$  через среднюю межпопуляционную изменчивость совокупности генов ( $\bar{F}_{ST}$ ). Непрямой метод оценки селективно-нейтрального уровня различий между популяциями исходит из гипотезы, что каждый из совокупности анализируемых гене-

Таблица 2

Оценка типа давления отбора для 24 биомаркеров ССЗ  
по параметру отклонения от селективно-нейтрального уровня  $F_{ST}$

Ген	Маркер	Диапазон частоты	Средняя частота	Средний размер метапопуляции	Межпопуляционные различия	Средняя оценка отбора	Оценка отбора по карте	Тип давления отбора	
		$\Delta q$	$\bar{q}$	$n_{(i)}$	$F_{ST(i)}$	$\Delta = F_{ST(i)} - \bar{F}_{ST}$	$\Delta = F_{ST(i)} - \bar{F}_{ST(i)}$	Класс отбора	Характеристика класса отбора
<i>VKORC1</i>	rs9923231	0,73	0,53	146	3,45	+2,21	+2,16	DIFF	ДИФФЕРЕНЦИРУЮЩИЙ отбор: изменчивость больше, чем различия между популяциями при селективно-нейтральной изменчивости: $F_{ST(i)} > \bar{F}_{ST}$
<i>CES1</i>	rs2244613	0,41	0,29	148	3,39	+2,15	+2,39	DIFF	
<i>VKORC1</i>	rs9934438	0,73	0,53	148	3,16	+1,92	+1,88	DIFF	
<i>ABCG2</i>	rs2231137	0,49	0,13	147	2,75	+1,51	+1,72	DIFF	
<i>CYP4F2</i>	rs2108622	0,36	0,23	146	2,28	+1,04	+1,13	DIFF	
<i>CYP2D6</i>	rs1065852	0,33	0,15	79	2,20	+0,96	+0,91	DIFF	
<i>SLCO1B1</i>	rs4149056	0,28	0,19	148	1,86	+0,62	+0,71	DIFF	
<i>VKORC1</i>	rs7294	0,43	0,32	138	1,82	+0,58	+0,53	DIFF	
<i>CYP2D6</i>	rs3892097	0,28	0,13	83	1,58	+0,34	+0,13	NEUTR	Условно СЕЛЕКТИВНО НЕЙТРАЛЬНЫЕ: межпопуляционная изменчивость слабо отличается от селективно-нейтрального уровня: $F_{ST(i)} \approx \bar{F}_{ST}$
<i>ADRB1</i>	rs1801252	0,44	0,17	52	1,48	+0,24	+0,48	NEUTR	
<i>CYP2C19</i>	rs4986893	0,12	0,01	149	0,97	-0,27	-0,25	NEUTR	
<i>P2RY12</i>	rs2046934	0,33	0,14	75	0,88	-0,36	-0,49	NEUTR	
<i>ABCG2</i>	rs2231142	0,15	0,08	148	0,88	-0,36	-0,49	NEUTR	
<i>ABCB1</i>	rs4148738	0,29	0,49	142	0,79	-0,45	-0,49	NEUTR	
<i>CYP2C19</i>	rs12248560	0,29	0,24	69	0,77	-0,47	-0,47	NEUTR	
<i>CYP2C9</i>	rs1057910	0,26	0,08	142	0,74	-0,50	-0,55	NEUTR	
<i>ABCB1</i>	rs1045642	0,33	0,43	150	0,63	-0,61	-0,59	STAB	СТАБИЛИЗИРУЮЩИЙ отбор: изменчивость меньше селективно-нейтрального уровня: $F_{ST(i)} < \bar{F}_{ST}$
<i>CYP2D6</i>	rs28371725	0,21	0,06	112	0,62	-0,62	-0,69	STAB	
<i>CYP2C9</i>	rs1799853	0,50	0,09	83	0,61	-0,63	-0,61	STAB	
<i>CYP2C8</i>	rs10509681	0,21	0,09	118	0,58	-0,66	-0,71	STAB	
<i>CYP3A4</i>	rs2242480	0,22	0,12	143	0,47	-0,77	-0,82	STAB	
<i>CYP3A5</i>	rs776746	0,20	0,11	141	0,41	-0,83	-0,86	STAB	
<i>PON1</i>	rs662	0,46	0,31	149	0,28	-0,96	-1,01	STAB	
<i>CYP2C19</i>	rs4244285	0,25	0,15	147	0,28	-0,96	-1,00	STAB	

Примечание: ССЗ — сердечно-сосудистые заболевания, DIFF — биомаркеры, отнесенные к классу действия дифференцирующего отбора, NEUTR — условно селективно нейтральные биомаркеры, STAB — биомаркеры, отнесенные к классу действия стабилизирующего отбора. Цветное изображение доступно в электронной версии журнала.

тических маркеров находится под действием своего собственного вектора отбора, в результате которого для каждого  $i$ -того маркера характерен свой собственный показатель межпопуляционных различий  $F_{ST(i)}$ . Поэтому усреднение оценок межпопуляционных различий по большой совокупности генов (среднее  $\bar{F}_{ST}$ ) усредняет эти разнонаправленные векторы, нивелируя эффекты отбора до нуля и давая оценку селективно-нейтральной изменчивости:  $\bar{F}_{ST} = F_{ST(NEUTR)}$ . Эта гипотеза неоднократно проверялась в регионах, для которых демографические оценки  $N_e$  и  $M_e$  считались корректными: было показано, что прямые (через  $N_e$  и  $M_e$ ) и не прямые (через  $\bar{F}_{ST}$ ) оценки практически совпадают. Поэтому обычно для оценки  $F_{ST(NEUTR)}$  используется непрямой метод (среднее  $\bar{F}_{ST}$ ), но при этом важно,

чтобы выборка маркеров была случайной по отношению к отбору и достаточно большой.

Для классических генетических маркеров научное сообщество пришло к выводу, что достаточно случайной выборки, включающей 50 генетических маркеров [15-19], но современные технологии позволяют значительно увеличивать выборки ДНК-маркеров в исследованиях, увеличивая тем самым надежность оценки селективно-нейтрального уровня различий между популяциями. В данной работе средняя оценка межпопуляционной изменчивости генофонда  $\bar{F}_{ST}$  получена по 1276 191 SNP-маркеру аутосомного генома, что надежно обеспечивает ее нейтральность по отношению к отбору. Массив 1293 образцов представителей 57 популяций Восточной Европы Биобанка Северной Евразии (изу-

Таблица 3

Характеристики 24 биомаркеров ССЗ: эффект отбора, лекарственные препараты и нозологии, для фармакотерапии которых могут использоваться соответствующие лекарственные препараты

Биомаркеры и характеристики эффектов отбора					Группа лекарственных препаратов						Нозологии, для фармакотерапии которых могут использоваться соответствующие лекарственные препараты							
Хромосома	Ген	Маркер	Интенсивность отбора $\Delta = F_{ST(0)} - F_{ST}$	Тип отбора	Антикоагулянтные средства	Гиполипидемические средства	Антиагрегантные средства	$\beta$ -адреноблокаторы	Ингибиторы ангиотензин-превращающего фермента	Антиаритмические средства	Тромбозы различной этиологии	Гиперхолестеринемия	Артериальная гипертензия	Острый коронарный синдром	Ишемическая болезнь сердца	Инфаркт миокарда	Аритмии, стенокардия	Хроническая сердечная недостаточность
20	VKORC1	rs9923231	+2,21	DIFF	+						+							
10	CES1	rs2244613	+2,15	DIFF			+				+	+	+	+		+		+
21	VKORC1	rs9934438	+1,92	DIFF	+							+						
23	ABCG2	rs2231137	+1,51	DIFF				+					+		+	+	+	+
8	CYP4F2	rs2108622	+1,04	DIFF			+				+						+	
3	CYP2D6	rs1065852	+0,96	DIFF				+						+	+			
14	SLCO1B1	rs4149056	+0,62	DIFF		+								+		+		
18	VKORC1	rs7294	+0,58	DIFF	+									+		+		
12	CYP2D6	rs3892097	+0,34	NEUTR	+							+						
6	ADRB1	rs1801252	+0,24	NEUTR	+							+						
16	CYP2C19	rs4986893	-0,27	NEUTR	+	+			+					+		+		
7	P2RY12	rs2046934	-0,36	NEUTR		+					+	+	+					
9	ABCG2	rs2231142	-0,36	NEUTR		+							+		+		+	
13	ABCB1	rs4148738	-0,45	NEUTR	+								+		+		+	
4	CYP2C19	rs12248560	-0,47	NEUTR			+						+		+		+	
2	CYP2C9	rs1057910	-0,50	NEUTR				+		+	+	+	+					
1	ABCB1	rs1045642	-0,61	STAB	+	+			+		+							
11	CYP2D6	rs28371725	-0,62	STAB	+						+							
5	CYP2C9	rs1799853	-0,63	STAB				+			+							
24	CYP2C8	rs10509681	-0,66	STAB	+	+			+			+		+	+			
22	CYP3A4	rs2242480	-0,77	STAB			+							+		+		
19	CYP3A5	rs776746	-0,83	STAB		+					+							
17	PON1	rs662	-0,96	STAB		+	+					+						
15	CYP2C19	rs4244285	-0,96	STAB			+				+							
ИТОГО					10	8	6	4	3	1	10	8	7	7	6	6	5	2

Примечание: ССЗ — сердечно-сосудистые заболевания, DIFF — биомаркеры, отнесенные к классу действия дифференцирующего отбора, NEUTR — условно селективно нейтральные биомаркеры, STAB — биомаркеры, отнесенные к классу действия стабилизирующего отбора, Цветное изображение доступно в электронной версии журнала,

ченный по 1276 191 SNP-маркеру) максимально полно охватывает разнообразие коренного населения региона, включающего представителей трех языковых семей — алтайской, индоевропейской, уральской. Для более репрезентативной оценки  $F_{ST}$  и увеличения выборки, популяции объединены в 20 метапопуляций (таблица 1) в соответствии с критериями генетического сходства и этногенетического родства. Итоговая средняя оценка селективно-нейтральной дифференциации популя-

ций  $F_{ST}$ , полученная по столь большому массиву анализируемых данных, составила  $F_{ST}=1,24$ .

## 2. Средняя оценка давления отбора для 24 биомаркеров ССЗ

Выборка, изученная по 24 биомаркерам ССЗ (детально описанная в [12]), в 2,5 раза больше ( $n_{CCZ}=3170$ ), чем выборка для оценки селективно-нейтральной изменчивости ( $n_{NEUTR}=2193$ , таблица 1). Однако некоторые биомаркеры изучены по меньшему числу образцов (таблица 2) и выборки



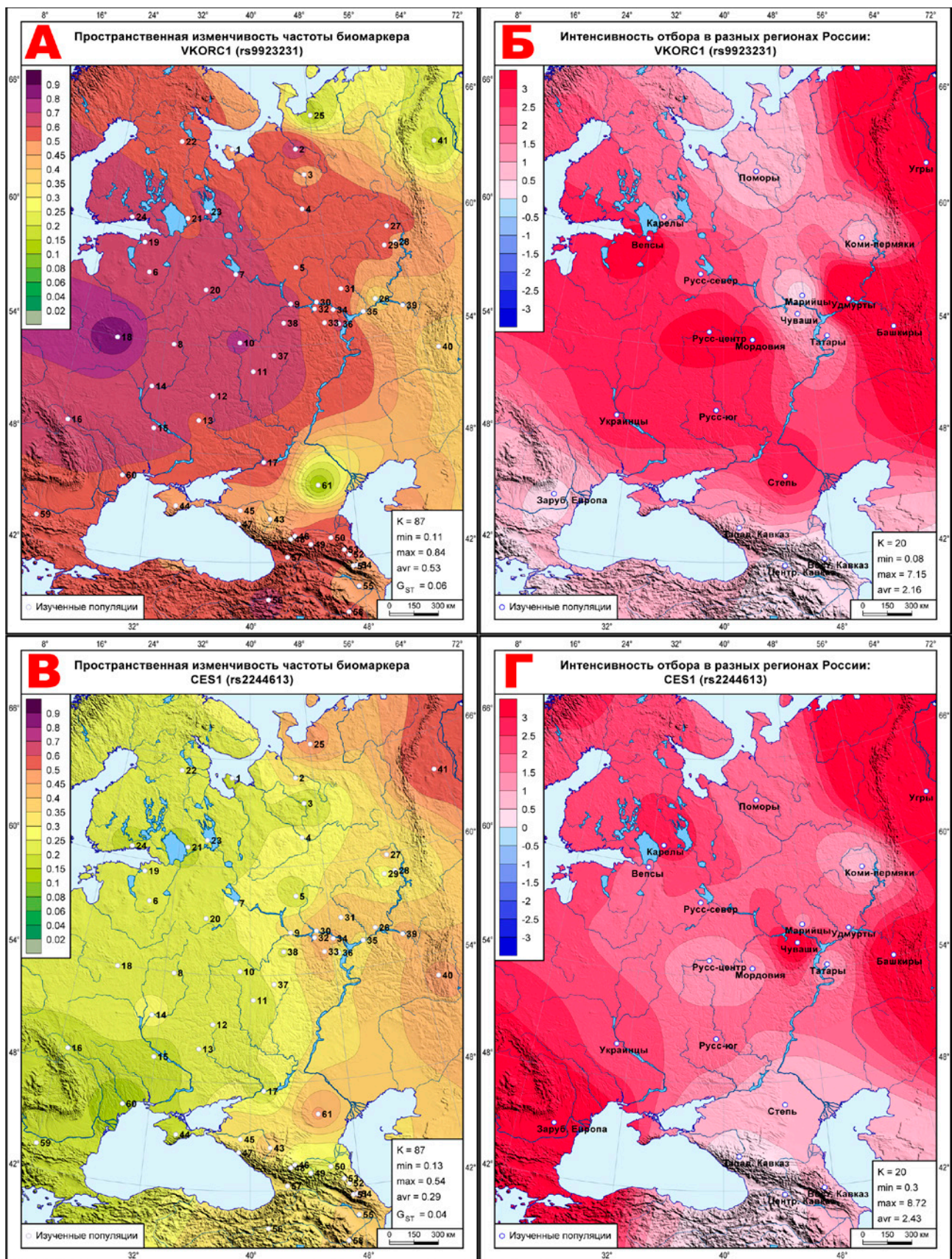


Рис. 1 Геногеография частоты (слева) и эффектов отбора (справа) для фармакогенетических биомаркеров CC3, подверженных дифференцирующему отбору: А, Б: *VKORC1* (rs9923231); В, Г: *CES1* (rs2244613); Д, Е: *VKORC1* (rs9934438); Ж, И: *ABCG2* (rs2231137); К, Л: *CYP4F2* (rs2108622).

Примечание: CC3 — сердечно-сосудистые заболевания.



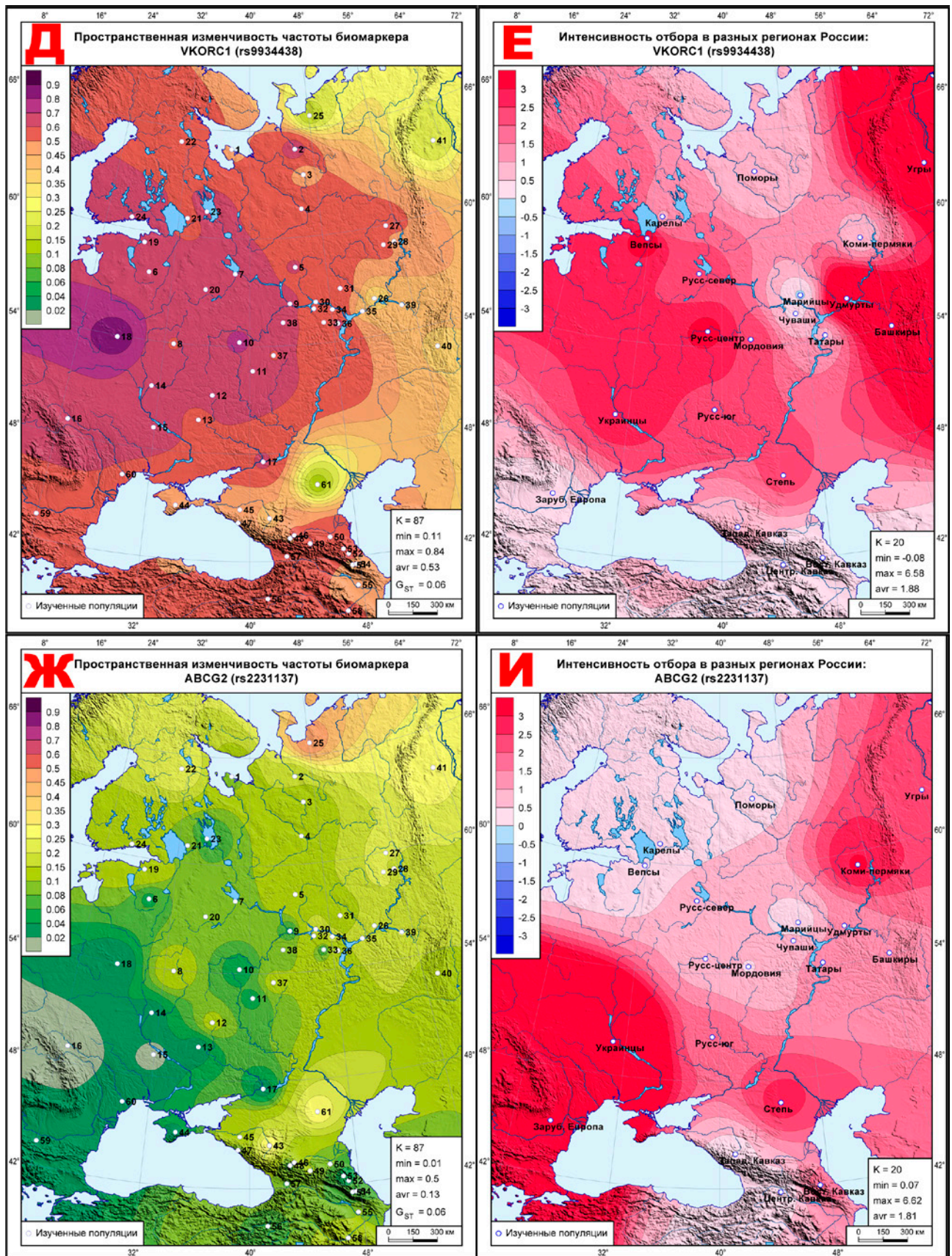


Рис. 1 (Продолжение) Геногеография частоты (слева) и эффектов отбора (справа) для фармакогенетических биомаркеров CC3, подверженных дифференцирующему отбору: А, Б: VKORC1 (rs9923231); В, Г: CES1 (rs2244613); Д, Е: VKORC1 (rs9934438); Ж, И: ABCG2 (rs2231137); К, Л: CYP4F2 (rs2108622).

Примечание: CC3 — сердечно-сосудистые заболевания.



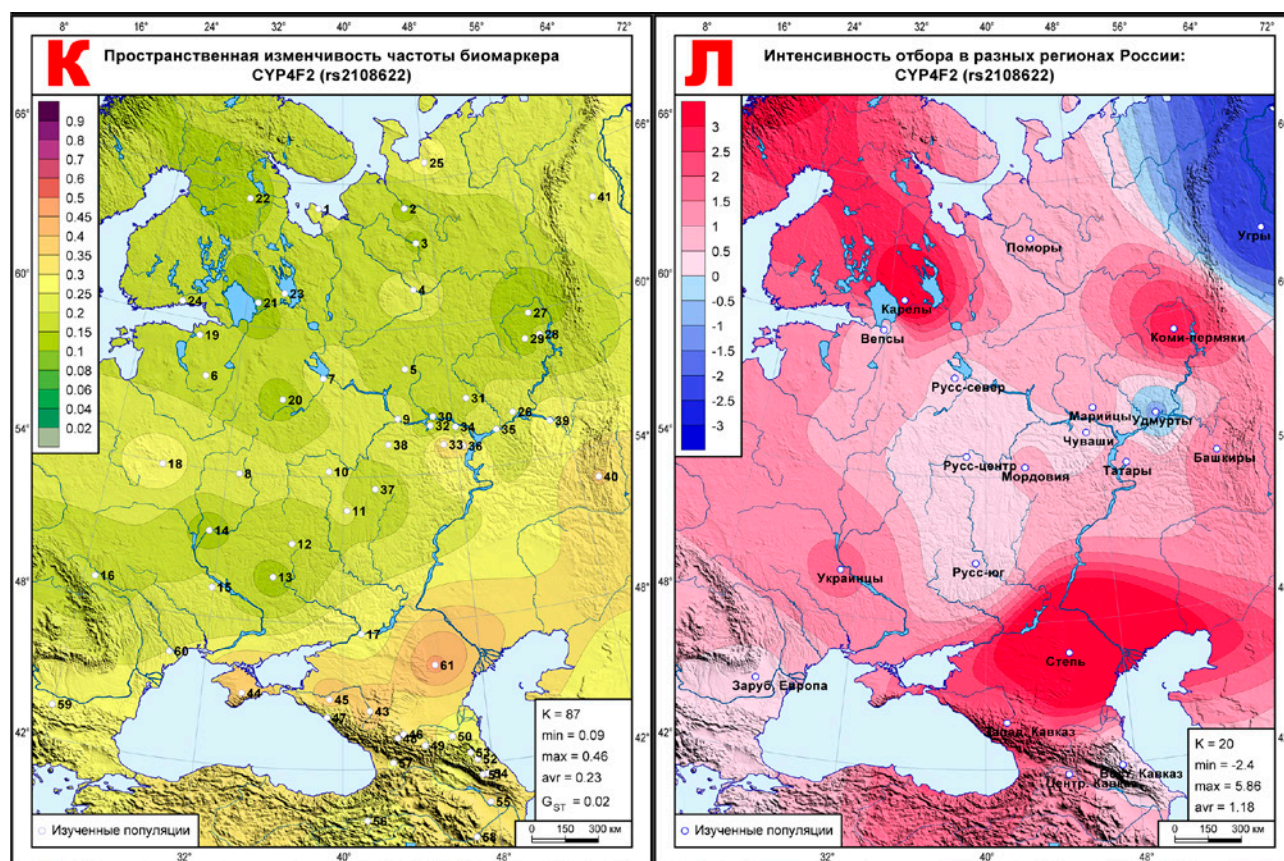


Рис. 1 (Продолжение) Геногеография частоты (слева) и эффектов отбора (справа) для фармакогенетических биомаркеров CC3, подверженных дифференцирующему отбору: А, Б: *VKORC1* (rs9923231); В, Г: *CES1* (rs2244613); Д, Е: *VKORC1* (rs9934438); Ж, И: *ABCG2* (rs2231137); К, Л: *CYP4F2* (rs2108622).

Примечание: CC3 — сердечно-сосудистые заболевания.

малы для некоторых популяций. Поэтому для увеличения репрезентативности данных выборок показатели межпопуляционной изменчивости  $F_{ST(i)}$  для каждого  $i$ -того биомаркера CC3 рассчитаны как средние не по 61 исходной популяции [12], а по 20 крупным метапопуляциям (таблица 2). Это позволяет провести корректное сравнение с теми же 20 метапопуляциями, по которым рассчитан уровень селективно-нейтральной изменчивости. Показатель  $F_{ST(i)}$  складывается из двух величин: а) общей для всех маркеров селективно-нейтральной величины  $\bar{F}_{ST}=1,24$ ; б) тех различий между популяциями, которые возникают в результате действия отбора, индивидуальны для каждого маркера и накладываются на селективно-нейтральную изменчивость. В результате действия дифференцирующего отбора различия между популяциями  $F_{ST(i)}$  становятся больше, чем без отбора:  $F_{ST(i)} > \bar{F}_{ST}$ , а воздействие стабилизирующего отбора уменьшает межпопуляционные различия:  $F_{ST(i)} < \bar{F}_{ST}$ .

С помощью такого теста 24 биомаркера CC3, расположенные по убыванию значения  $F_{ST(i)}$ , разделены на три условных класса: маркеры дифференцирующего отбора, селективно-нейтральные

маркеры и маркеры стабилизирующего отбора (таблица 2). В настоящей работе мы не касаемся проблемы статистически значимого выделения границ между классами [16-19], поскольку на первом этапе решаем более простую задачу — создание рабочей технологии обнаружения эффектов отбора. Поэтому 8 биомаркеров с положительным знаком  $\Delta > 0,5$  отклонения от селективно-нейтрального уровня условно отнесены к классу ДНК-маркеров, подверженных дифференцирующему отбору ( $F_{ST(i)} > \bar{F}_{ST}$ ). Следующие 8 биомаркеров, мало отличающихся от  $\bar{F}_{ST}$  ( $-0,5 < \Delta < 0,5$ ), отнесены к классу условно селективно-нейтральных. Последние 8 ДНК-маркеров с наибольшим отрицательным отклонением от  $\bar{F}_{ST}$  ( $-1 < \Delta < -0,5$ ) условно отнесены к классу маркеров, подверженных стабилизирующему отбору (таблица 2). В таблице 3 для каждого из 24 биомаркеров CC3 указаны характеристики их связи с лекарственными препаратами и заболеваниями.

### 3. Географическая изменчивость давления отбора

Полученные средние оценки давления отбора на 24 биомаркера CC3 относятся только к населению Европейской части России и сопредельных территорий (таблица 1). В других регионах



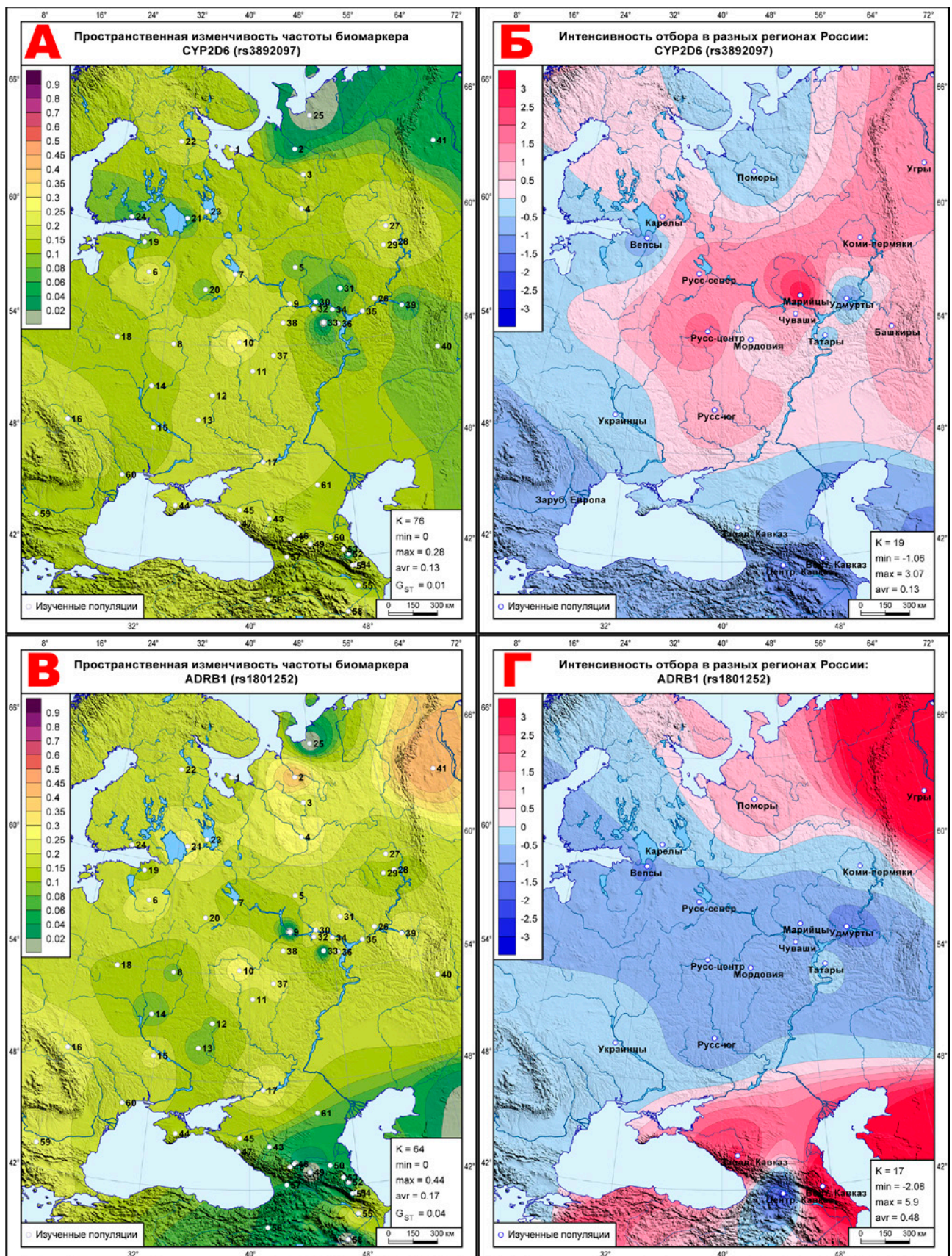


Рис. 2 Геногеография частоты (слева) и эффектов отбора (справа) для фармакогенетических биомаркеров CC3, условно селективно-нейтральных: А, Б: CYP2D6 (rs3892097); В, Г: ADRB1 (rs1801252); Д, Е: CYP2C19 (rs4986893); Ж, И: P2RY12 (rs2046934); К, Л: ABCG2 (rs2231142).

Примечание: CC3 — сердечно-сосудистые заболевания.



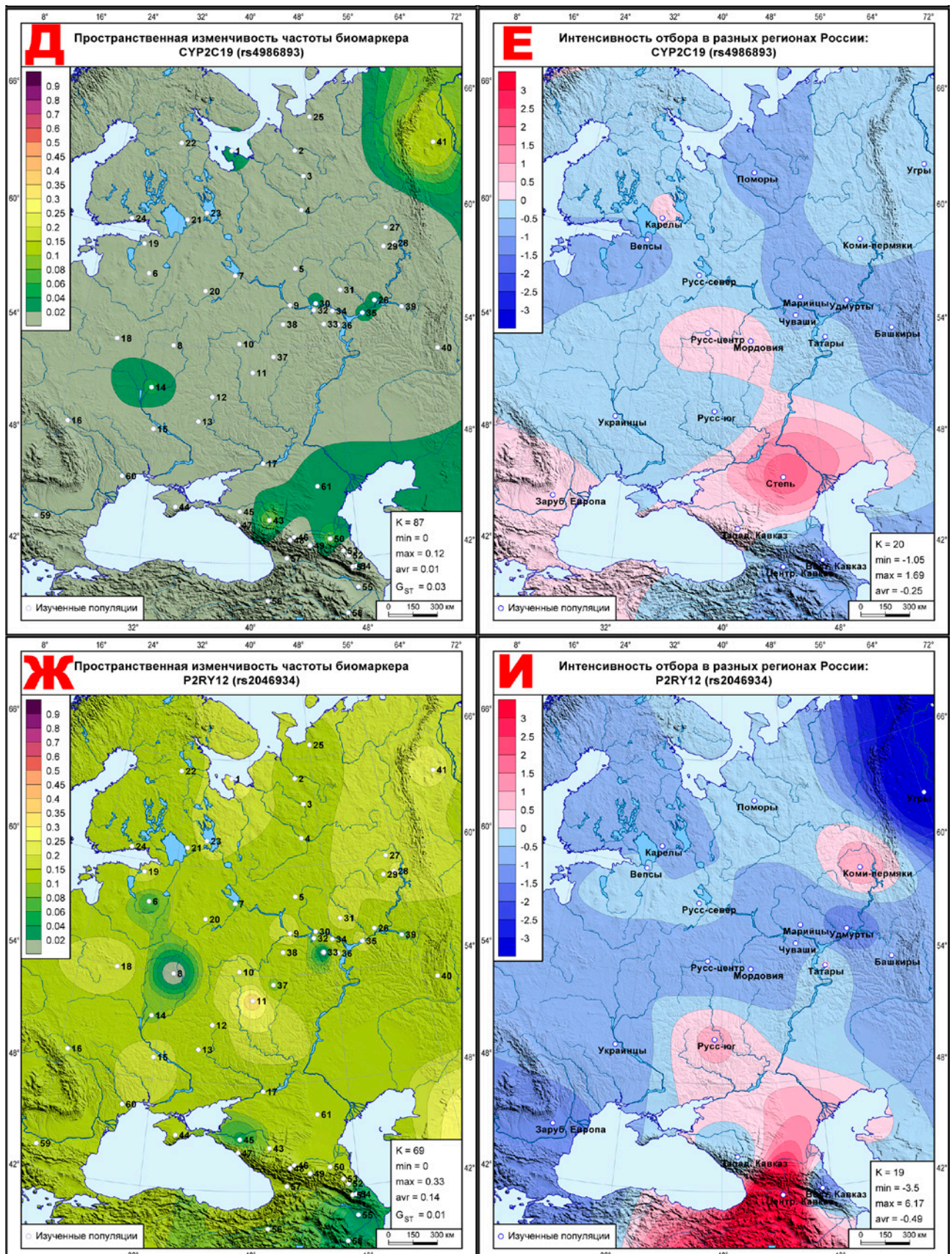


Рис. 2 (Продолжение) Геногеография частоты (слева) и эффектов отбора (справа) для фармакогенетических биомаркеров CC3, условно селективно-нейтральных: А, Б: CYP2D6 (rs3892097); В, Г: ADRB1 (rs1801252); Д, Е: CYP2C19 (rs4986893); Ж, И: P2RY12 (rs2046934); К, Л: ABCG2 (rs2231142).

Примечание: CC3 — сердечно-сосудистые заболевания.



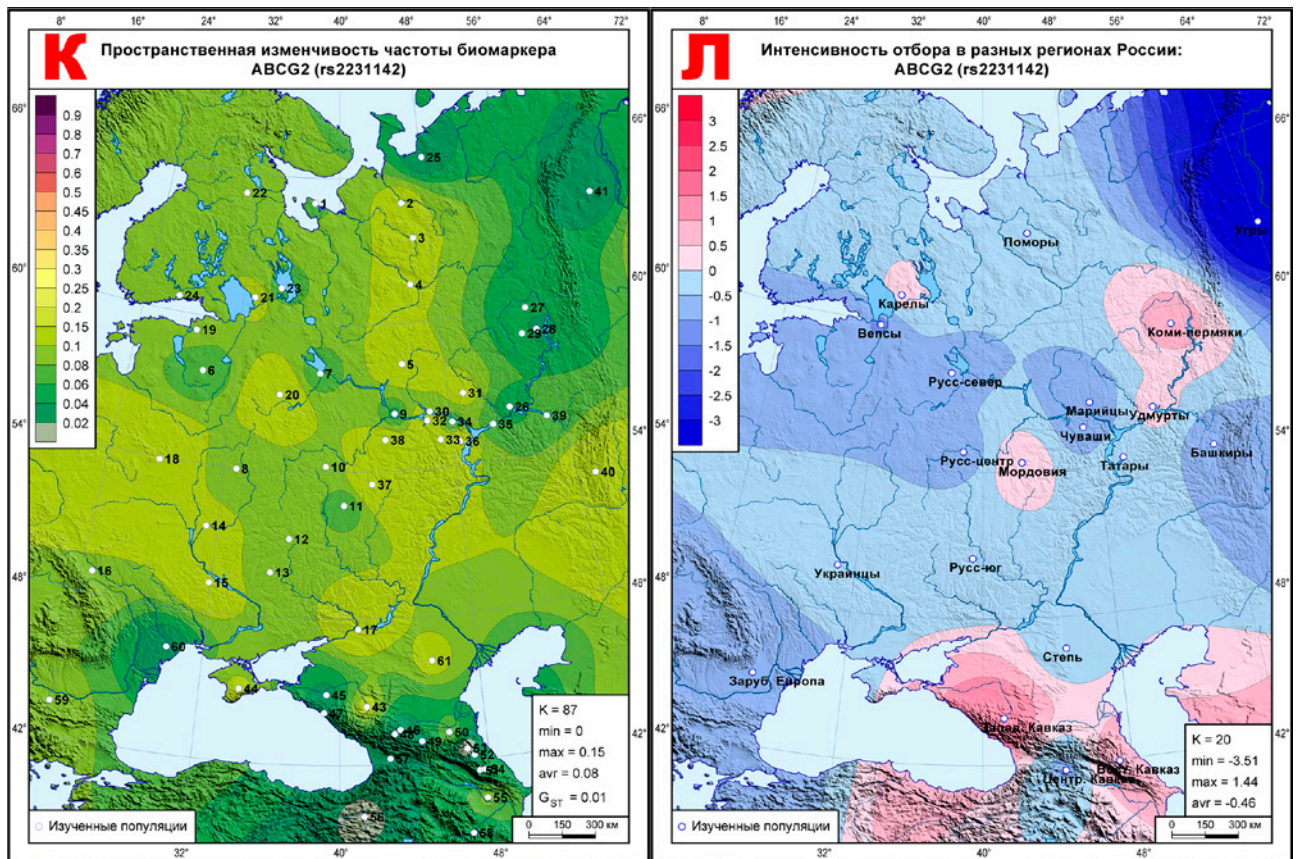


Рис. 2 (Продолжение) Геногеография частоты (слева) и эффектов отбора (справа) для фармакогенетических биомаркеров ССЗ, условно селективно-нейтральных: А, Б: *CYP2D6* (rs3892097); В, Г: *ADRB1* (rs1801252); Д, Е: *CYP2C19* (rs4986893); Ж, И: *P2RY12* (rs2046934); К, Л: *ABCG2* (rs2231142).

Примечание: ССЗ — сердечно-сосудистые заболевания.

действие отбора может быть иным, что требует отдельного исследования. Более того, действие отбора может быть неодинаковым в разных частях анализируемого региона. Поэтому средние оценки отбора (таблица 2) являются вполне достоверным указанием на то, какие именно биомаркеры требуют наиболее пристального изучения. Если биомаркеры подвержены действию сильного стабилизирующего отбора, то их межпопуляционные различия сглажены по сравнению с селективно-нейтральной изменчивостью. Но даже для этих биомаркеров геногеографический подход помогает выявить те части региона, где межпопуляционные различия все же велики. Дифференцирующий отбор, напротив, указывает на существенные отличия между популяциями региона по данному биомаркеру и на то, что для разных частей региона может потребоваться выделение отдельных групп популяций, в которых использование разных фармакогенетических протоколов будет клинически рациональным. Однако и здесь геногеография может помочь выделить те части региона, где дифференцирующий отбор ослаблен и стандартные подходы к лечению будут все же пригодны. Поэтому для биомаркеров, под-

верженных любым типам и интенсивности отбора, важно оценить, как отбор действует в разных частях ареала анализируемого региона.

Обнаружить особенности разных популяций региона в отношении действия отбора помогает атлас геногеографических карт отбора. Для его создания проведены следующие операции: а) для каждой (j) из 20 метапопуляций (их краткие названия приведены на картах отбора, а описание входящих в них популяций — в таблице 1) дана оценка межпопуляционной изменчивости  $F_{ST(j)}$  для каждого i-того биомаркера ССЗ; б) для каждой (j) из 20 метапопуляций выявлен показатель селективно-нейтральных различий  $\bar{F}_{ST(j)}$ ; в) дана оценка отбора по i-тому биомаркеру для каждой j-той метапопуляции через отклонение от селективно-нейтрального уровня изменчивости ( $\Delta_{ij} = F_{ST(ij)} - \bar{F}_{ST(j)}$ ); г) проведена интерполяция полученных оценок отбора  $\Delta_{ij}$  по ареалу карты. Подчеркнем, что и для карт, и для расчета средних оценок отбора  $\bar{F}_{ST(i)}$  оценки получены для одних и тех же 20 метапопуляций и потому полностью сопоставимы (таблица 2). Отличия средних оценок по 20 метапопуляциям от оценок, полученных по картам (таблица 2), связаны с тем, что



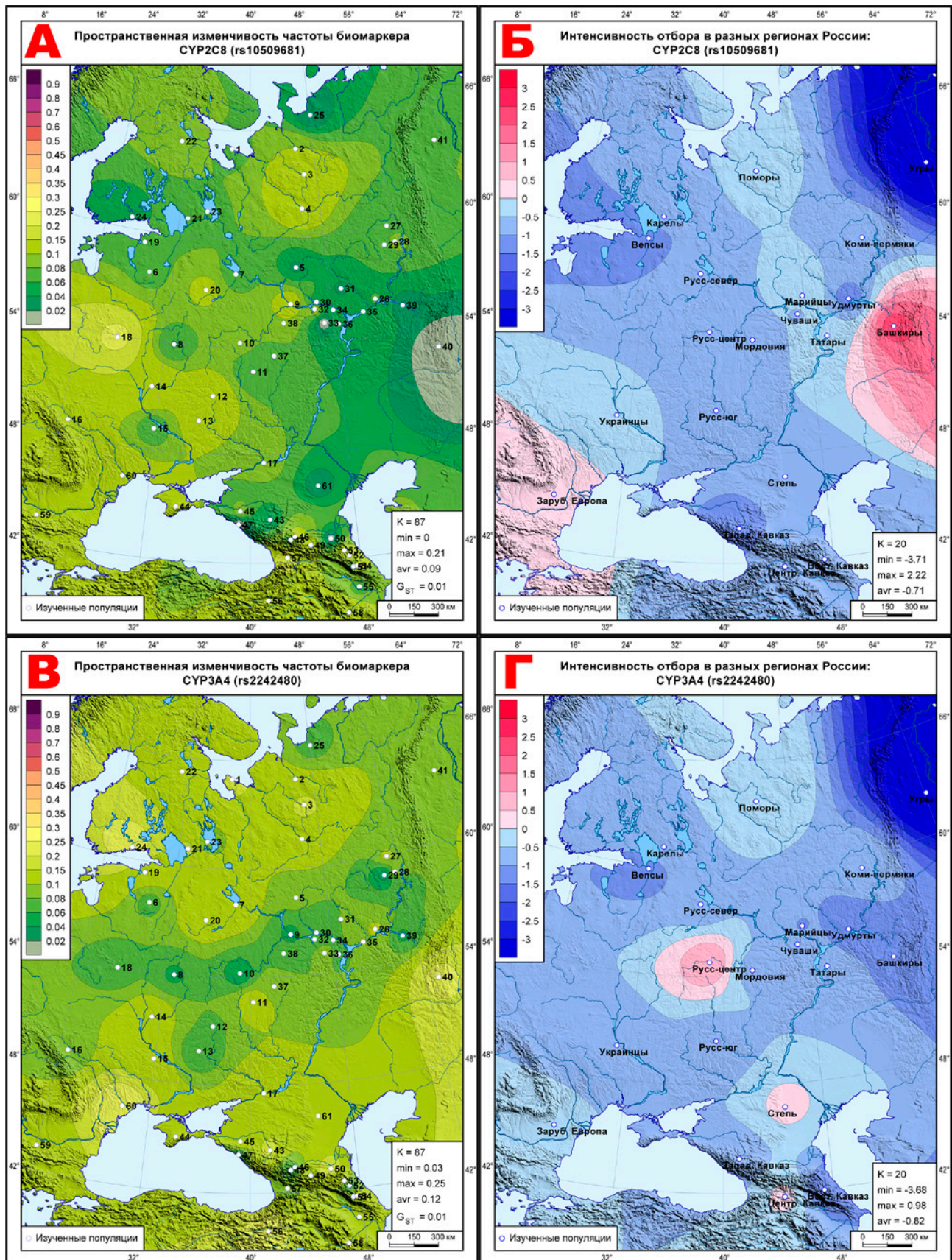


Рис. 3 Геногеография частоты (слева) и эффектов отбора (справа) для фармакогенетических биомаркеров CC3, подверженных стабилизирующему отбору: А, Б: CYP2C8 (rs10509681); В, Г: CYP3A4 (rs2242480); Д, Е: CYP3A5 (rs776746); Ж, И: PON1 (rs662); К, Л: CYP2C19 (rs4244285).

Примечание: CC3 — сердечно-сосудистые заболевания.



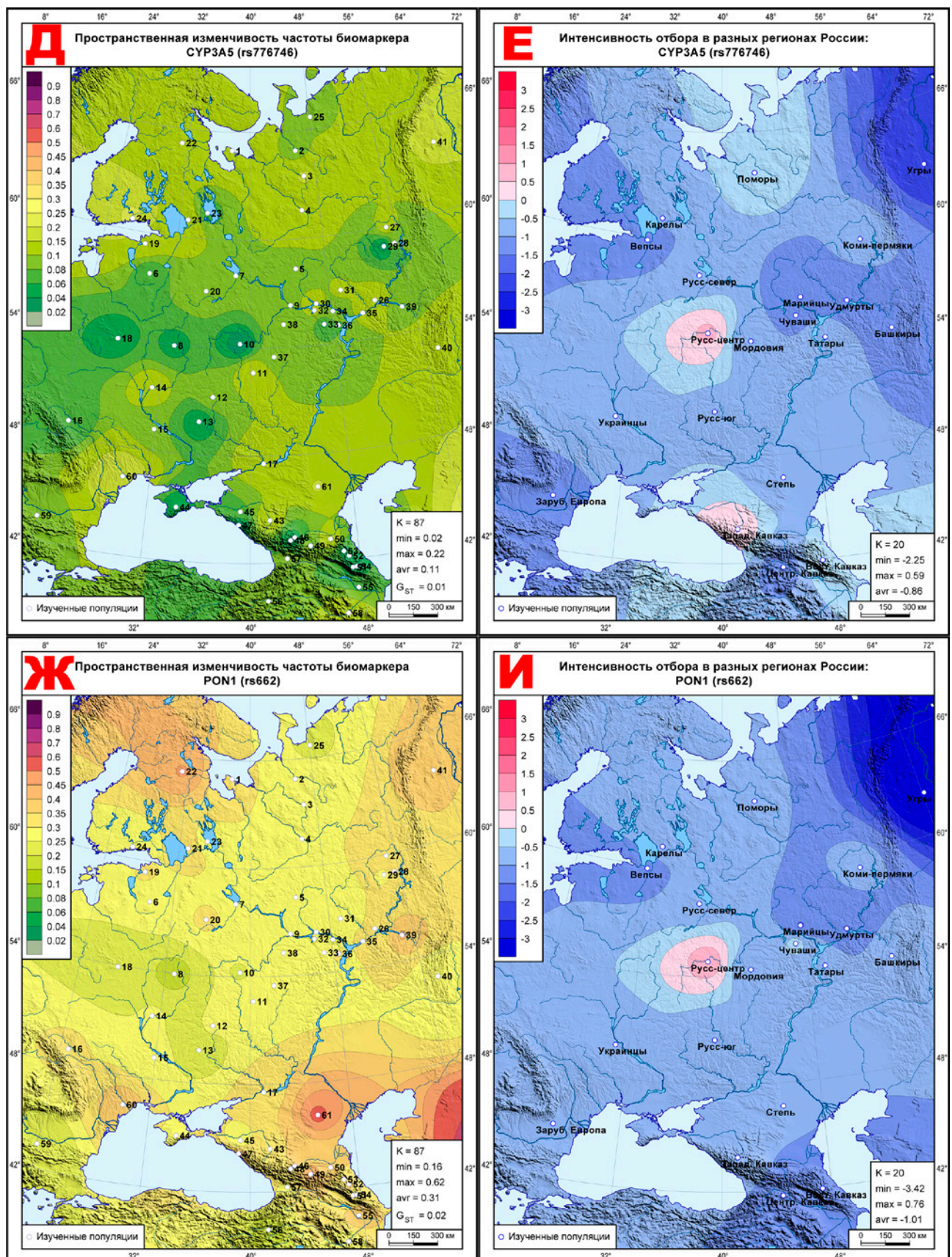


Рис. 3 (Продолжение) Геногеография частоты (слева) и эффектов отбора (справа) для фармакогенетических биомаркеров ССЗ, подверженных стабилизирующему отбору: А, Б: CYP2C8 (rs10509681); В, Г: CYP3A4 (rs2242480); Д, Е: CYP3A5 (rs776746); Ж, И: PON1 (rs662); К, Л: CYP2C19 (rs4244285).

Примечание: ССЗ — сердечно-сосудистые заболевания.



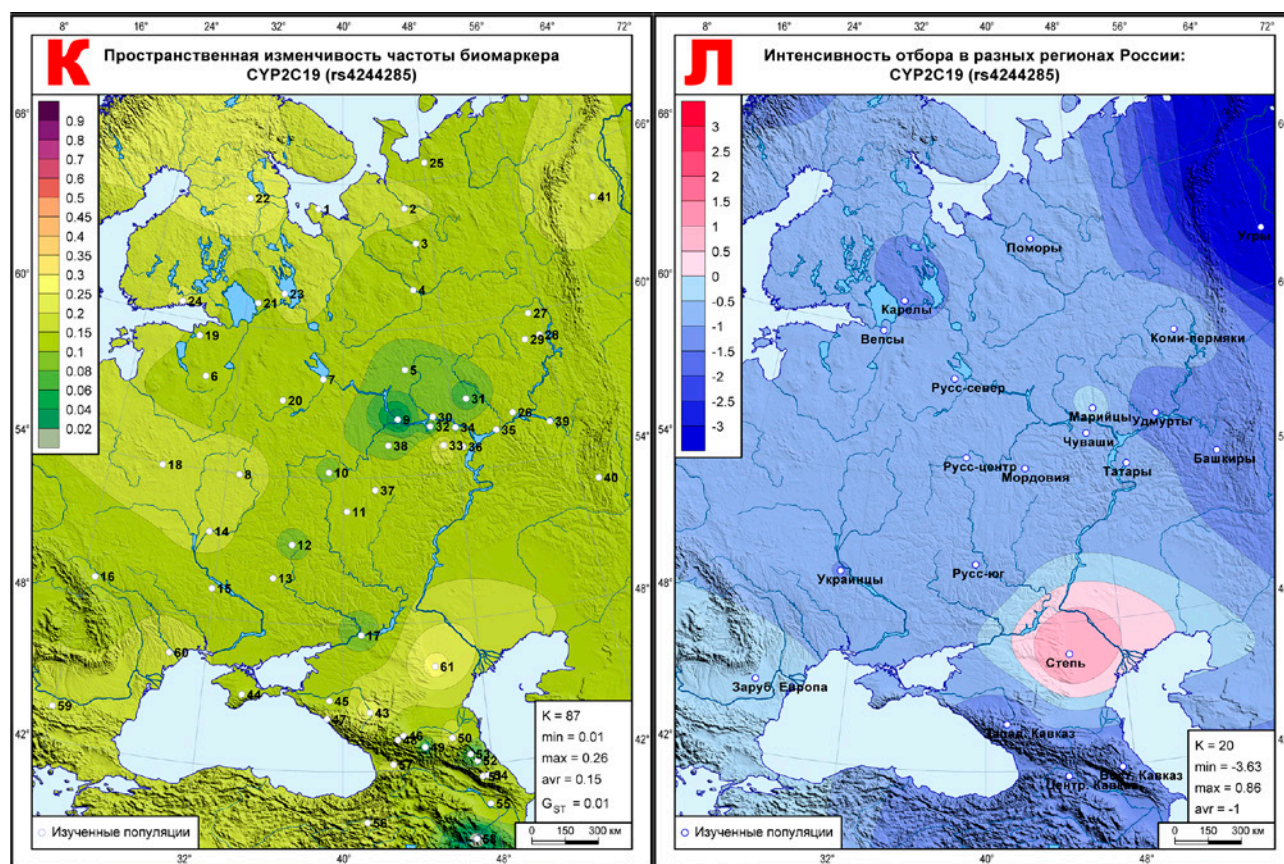


Рис. 3 (Продолжение) Геногеография частоты (слева) и эффектов отбора (справа) для фармакогенетических биомаркеров CC3, подверженных стабилизирующему отбору: А, Б: *CYP2C8* (rs10509681); В, Г: *CYP3A4* (rs2242480); Д, Е: *CYP3A5* (rs776746); Ж, И: *PON1* (rs662); К, Л: *CYP2C19* (rs4244285).

Примечание: CC3 — сердечно-сосудистые заболевания.

геногеографическое картографирование позволяет учесть важнейшее свойство популяции — ее ареал. Поэтому карта, учитывая различия в размерах ареала метапопуляций и пространственные соотношения ареалов, дает более корректную оценку изменчивости отбора в географическом пространстве.

Для демонстрации пространственной изменчивости отбора из атласа выбрано по 5 карт каждого класса отбора, наиболее ярко его демонстрирующих: 5 карт биомаркеров, наиболее подверженных дифференцирующему отбору (рисунок 1); 5 карт биомаркеров, наиболее близких к селективно-нейтральной изменчивости (рисунок 2); 5 карт биомаркеров, наиболее подверженных стабилизирующему отбору (рисунок 3). Шкала на всех картах отбора одинакова (от -3 до +3 с интервалом 0,5), и чем интенсивнее отбор, тем интенсивнее окраска. Области с положительным знаком  $\Delta F_{ST}$ , указывающие на действие дифференцирующего отбора, окрашены в красные тона, а территории стабилизирующего отбора с отрицательным знаком  $\Delta F_{ST}$  показаны в синей гамме оттенков. Каждая из 15 карт отбора сопровождается картой геногеографической изменчивости частоты этого маркера в 61 исходной популяции (рисунки 1-3),

где каждая исходная популяция обозначена номером, позволяющим найти ее описание в таблице 1. Для карт частот биомаркеров использована иная шкала: от 0 до 0,1 — с шагом 0,02; от 0,1 до 0,5 — с шагом 0,05; от 0,5 до 1 — с шагом 0,1.

#### Геногеография дифференцирующего отбора

Весь ареал всех 5-и карт окрашен в красные тона дифференцирующего отбора, указывая на его действие во всех метапопуляциях. Исключение составляет карта (рисунок 1Л) *CYP4F2\*3* (rs2108622), где среднее давление дифференцирующего отбора ( $\bar{\Delta}=1,04$ ) меньше, чем у остальных 4-х биомаркеров ( $1,51 < \bar{\Delta} < 2,21$ ) этого класса (рисунок 1). И поэтому на карте этого маркера (рисунок 1Л) две метапопуляции обозначены синим цветом, указывающим на действие в этих метапопуляциях стабилизирующего отбора: для удмуртов  $\Delta=-0,61$ , для угров  $\Delta=-2,40$ . На остальных картах рисунка 1 отбор варьирует в рамках дифференцирующего отбора, но в некоторых метапопуляциях настолько смягчается, что уже близок к селективной нейтральности. Например, на рисунках 1Б и 1Е — такие метапопуляции обнаруживаются на юге и в Приуралье, на рисунке 1Г — на Западном и Центральном Кавказе, а на рисунке 1И, кроме

Кавказа, уменьшение давления отбора уже охватывает метапопуляции северо-запада (карелы, вепсы, поморы), и далее снижение дифференцирующего отбора тянется до правобережья Волги. На рисунке 1Л отбор снижается до селективной нейтральности в центре ареала — во всех трех русских популяциях и у чувашей, а также в Зарубежной Европе и на Восточном Кавказе.

#### **Геногеография условно селективно-нейтральной изменчивости**

Карты этого класса окрашены в тона обоих типов отбора — дифференцирующего и стабилизирующего. Кроме двух интервалов ( $-0,5 < \Delta < 0,5$ ) условной селективной нейтральности, видим и усиление отбора в ряде метапопуляций: на карте *CYP2D6* (rs3892097) (рисунок 2Б) дифференцирующий отбор (красные тона) проявляется у центральных русских и марийцев, а стабилизирующий отбор — на юге ареала. На карте *ADRB1* (rs1801252) (рисунок 2Г) дифференцирующий отбор ярко проявился на северо-востоке и на юге, а стабилизирующий — у вепсов, удмуртов и на Центральном Кавказе. Карта отбора *CYP2C19* (rs4986893) (рисунок 2Е) уже почти вся окрашена в синие тона близости к стабилизирующему отбору, а дифференцирующий обнаружен только в степной метапопуляции. Схожую картину мы видим и на карте *P2RY12* (rs2046934) (рисунок 2И), где выраженный дифференцирующий отбор сосредоточен на Кавказе, с невысокими значениями обнаруживается у коми и южных русских, а остальной ареал принадлежит синим тонам стабилизирующего отбора. Здесь уже нет выраженного дифференцирующего отбора — за пределы селективной нейтральности незначительно выходят лишь метапопуляции коми-пермяков и Кавказа. Но и стабилизирующий отбор ярко выражен лишь у угров, а в целом почти все метапопуляции близки к селективной нейтральности.

#### **Геногеография стабилизирующего отбора**

Карты этого блока продолжают тенденцию предыдущего: постепенное нарастание синих тонов стабилизирующего отбора с выявлением тех метапопуляций, где действует дифференцирующий отбор. Так, на карте *CYP2C8* (rs10509681) (рисунок 3Б) коридор нейтральных значений тянется от поморов к татарам, переходя у башкир к ярко выраженному дифференцирующему отбору. Селективно-нейтральные значения на этой карте занимают юго-запад (украинцы, Зарубежная Европа). Полоса невысоких значений стабилизирующего отбора находится между этими двумя зонами нейтральных значений, простираясь от северо-запада до Кавказа. При этом ядро максимального стабилизирующего отбора (угры,  $\Delta = -3,71$ ) находится вблизи от ядра максимального дифференцирующего отбора (башкиры,  $\Delta = +2,22$ ). Остальные карты этого блока в целом схожи — везде максимальные значения

стабилизирующего отбора приходится на угров, ареалы селективно-нейтральных показателей почти исчезают, уступая ровному ландшафту стабилизирующего отбора. Он нарушается небольшими всплесками дифференцирующего отбора в двух метапопуляциях: у центральных русских на картах трех биомаркеров ( $\Delta = +0,98$ , *CYP3A4* (rs2242480), рисунок 3Г;  $\Delta = +0,59$ , *CYP3A5* (rs776746), рисунок 3Е;  $\Delta = +0,76$ , *PON1* (rs662), рисунок 3И) и у степных народов на карте *CYP2C19* (rs4244285), где  $\Delta = +0,66$  (рисунок 3Л).

### **Обсуждение**

Предложенная геногеографическая технология позволяет выделять те биомаркеры и популяции, на которые стоит обратить внимание при разработке общей стратегии внедрения персонализированных подходов с использованием знаний фармакогенетики пациентов в том или ином регионе или в группе популяций.

Биомаркеры, находящиеся под давлением стабилизирующего отбора, близки по своим показателям у разных этнических групп региона. Селективно-нейтральные маркеры более изменчивы у разных народов. Биомаркеры, подверженные дифференцирующему отбору, значительно различаются в разных популяциях. Знание этих закономерностей позволяет выделять регионы и отдельные популяции, где внедрение и использование соответствующих фармакогенетических протоколов будет клинически и клинико-экономически обоснованным.

Различия по межпопуляционной изменчивости биомаркеров отчасти отражаются и в их частоте. Так, средняя частота биомаркеров (таблица 2) в классе дифференцирующего отбора ( $\bar{q} = 0,30$ ) в ~2 раза выше средней частоты биомаркеров в двух других классах — и селективно-нейтральных ( $\bar{q} = 0,17$ ), и подверженных стабилизирующему отбору ( $\bar{q} = 0,17$ ). Но это лишь общая тенденция: ориентироваться только на показатель частоты нельзя. И в классе дифференцирующего отбора есть SNP-маркеры с невысокими частотами:  $\bar{q} = 0,13$  *ABCG2* (rs2231137),  $\bar{q} = 0,15$  *CYP2D6* (rs1065852),  $\bar{q} = 0,19$  *SLCO1B1* (rs4149056). И в классах селективной нейтральности и стабилизирующего отбора встречаются биомаркеры с высокой средней частотой:  $\bar{q} = 0,49$  *ABCB1* (rs4148738),  $\bar{q} = 0,43$  *ABCB1* (rs1045642),  $\bar{q} = 0,31$  *PON1* (rs662).

То же можно сказать и о размахе изменчивости ( $q_{\text{MAX}} - q_{\text{MIN}}$ ): в среднем он в 2 раза больше (таблица 2) в классе дифференцирующего отбора (0,55), чем в других классах — и селективно-нейтральном (0,26), и стабилизирующего отбора (0,27). Но и из этой общей тенденции есть исключения. В классе дифференцирующего отбора есть SNP-маркеры с небольшим диапазоном вариации частоты ( $q_{\text{MAX}} - q_{\text{MIN}}$ ): 0,33 *CYP2D6* (rs1065852), 0,28 *SLCO1B1* (rs4149056), а в других классах — мар-



керы с большим диапазоном частот: 0,44 *ADRB1* (rs1801252), 0,46 *PON1* (rs662).

Поэтому наиболее полезным инструментом оказывается не "средняя температура по больнице", а геногеографические карты: они дают не только общерегиональные, но и конкретные показатели отбора в конкретных популяциях. Приведенные карты геногеографического атласа отбора (рисунки 1-3) показывают, что даже на территориально близкие популяции могут действовать разные типы отбора (например, рисунки 1Л, 2И, 3Б). Здесь наиболее показателен Кавказ: например, для *P2RY12* (rs2046934) на Западном Кавказе обнаружена селективно-нейтральная изменчивость, на Центральном Кавказе — дифференцирующий отбор, на Восточном Кавказе — стабилизирующий (рисунок 2И). Для маркера *ADRB1* (rs1801252) наблюдается не меньшее разнообразие: на Центральном Кавказе действует стабилизирующий отбор, а на Западном и Восточном — дифференцирующий (рисунок 2Г). Если учесть, что для региона Восточной Европы в целом эти маркеры отнесены к селективно-нейтральным, становится очевидной необходимость использования геногеографических карт для анализа эффектов отбора в конкретных метапопуляциях. Чтение этих карт не требует специальной подготовки: популяции, требующие осторожности в применении стандартных фармакотерапевтических протоколов, на карте отмечены интенсивным красным цветом.

**Ограничения исследования.** Определенные ограничения исследования связаны с малым количеством исследованных биообразцов в отдельных регионах (средние объемы выборок указаны в таблице 2) по 6 биомаркерам: *ADRB1* (rs1801252), *CYP2D6* (rs1065852), *CYP2D6* (rs3892097), *CYP2C19* (rs12248560), *CYP2C9* (rs1799853), *P2RY12* (rs2046934).

## Заключение

Геногеографические технологии отбора могут служить дополнительным инструментом при разработке этно-регионально адаптированных фармакогенетических протоколов и выявления приоритетных популяций для их внедрения.

Основой использования геногеографических технологий служит популяционный биобанк, содержащий биообразцы из основных популяций интересующего региона.

Для оценки отбора необходимо иметь базу данных фармакогенетических биомаркеров о данном

регионе и базу данных с большой совокупностью других аутосомных SNP-маркеров, изученных в тех же популяциях и позволяющих оценить для них уровень селективно-нейтральных различий.

Данные этих двух баз необходимо организовать таким образом, чтобы они содержали данные об одних и тех же популяциях с репрезентативными выборками из каждой популяции.

Оценку отбора по данному биомаркеру можно получить сравнением оценки его межпопуляционных различий и оценки различий между этими же популяциями по большой совокупности аутосомных SNP-маркеров, которая дает точку отсчета — уровень селективно-нейтральной межпопуляционной изменчивости.

Те биомаркеры, изменчивость которых мало отличается от селективно-нейтральной, можно отнести к классу условно селективно-нейтральных. Если различия между популяциями по данному биомаркеру значительно меньше селективно-нейтральной изменчивости популяций, то этот биомаркер находится под действием стабилизирующего отбора. Если же межпопуляционные различия значительно превосходят селективно-нейтральный уровень, то данный биомаркер находится под давлением дифференцирующего отбора.

Именно биомаркеры, находящиеся под давлением дифференцирующего отбора, требуют разработки этно-регионально адаптированных подходов к использованию фармакогенетических протоколов.

Поскольку изменчивость и остальных биомаркеров обнаруживает отдельные популяции, находящиеся под действием дифференцирующего отбора, при выявлении приоритетных популяций для внедрения этно-регионально адаптированных подходов к использованию фармакогенетических протоколов необходимо использовать геногеографические карты отбора, указывающие на такие приоритетные популяции.

**Благодарности.** Авторы благодарят АНО "Биобанк Северной Евразии" за предоставление коллекций ДНК.

**Отношения и деятельность.** Исследование выполнено при поддержке гранта РНФ № 21-14-00363 (биоинформатический анализ, интерпретация результатов) и государственного задания Минобрнауки России для ФГБНУ "МГНЦ" (картографический и статистический анализ).

## Литература/References

- Johnson JA, Caudle KE, Gong L, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for pharmacogenetics-guided warfarin dosing: 2017 update. *Clin Pharmacol Ther.* 2017;102(3):397-404. doi:10.1002/cpt.668.
- Lee CR, Luzum JA, Sangkuhl K, et al. Clinical pharmacogenetics implementation consortium guideline for CYP2C19 genotype and clopidogrel therapy: 2022 update. *Clin Pharmacol Ther.* 2022;112(5):959-67. doi:10.1002/cpt.2526.
- Mega JL, Close SL, Wiviott SD, et al. Genetic variants in ABCB1 and CYP2C19 and cardiovascular outcomes after treatment with clopidogrel and prasugrel in the TRITON-TIMI 38 trial: a pharmacogenetic analysis. *Lancet.* 2010;376(9749):1312-9. doi:10.1016/S0140-6736(10)61273-1.
- Cooper-DeHoff RM, Niemi M, Ramsey LB, et al. The clinical pharmacogenetics implementation consortium guideline for SLCO1B1, ABCG2, and CYP2C9 genotypes and statin-associated musculoskeletal symptoms. *Clin Pharmacol Ther.* 2022;111(5):1007-21. doi:10.1038/clpt.2014.125.
- Kanuri SH, Kreutz RP. Pharmacogenomics of novel direct oral anticoagulants: newly identified genes and genetic variants. *J Pers Med.* 2019;9(1):7. doi:10.3390/jpm9010007.
- Kukes VG, Sychyov DA, Al Akhmad Faisal, et al. The impact of unique patient characteristics on the risk of adverse drug reaction. *Bulletin of Roszdravnadzor.* 2011;(6):59-63. (In Russ.) Кукес В. Г., Сычев Д. А., Фейсал А. А. и др. Влияние индивидуальных особенностей пациентов на риск развития нежелательных лекарственных реакций. *Вестник Росздравнадзора.* 2011;(6):59-63.
- Stepanov VA. Genomes, populations, diseases: ethnic genomics and personalised medicine. *Acta Naturae.* 2010;2(4):18-34. (In Russ.) Степанов В. А. Геномы, популяции, болезни: этническая геномика и персонализированная медицина. *Acta Naturae.* 2010;2(4):18-34.
- Mirzaev KB, Fedorinov DS, Ivashchenko DV, et al. ADME pharmacogenetics: future outlook for Russia. *Pharmacogenomics.* 2019;20(11):847-65. doi:10.2217/pgs-2019-0013.
- Mirzaev KB, Fedorinov DS, Ivashchenko DV, et al. Multi-Ethnic Analysis of Cardiac Pharmacogenetic Markers of Cytochrome p450 and Membrane Transporters Genes in the Russian Population. *Rational Pharmacotherapy in Cardiology.* 2019;15(3):393-406. (In Russ.) Мирзаев К. Б., Федоринов Д. С., Иващенко Д. В. и др. Мультиэтнический анализ кардиологических фармакогенетических маркеров генов цитохрома P450 и мембранных транспортеров в российской популяции. *Рациональная фармакотерапия в кардиологии.* 2019;15(3):393-406. doi:10.20996/1819-6446-2019-15-3-393-406.
- Gorin IO, Petrushenko VS, Zapisetskaya Yu S. Population-based biobank for analyzing the frequencies of clinically relevant DNA markers in the Russian population: bioinformatic aspects. *Cardiovascular Therapy and Prevention.* 2020;19(6):2732. (In Russ.) Горин И. О., Петрушенко В. С., Записецкая Ю. С. и др. Применение популяционного биобанка для анализа частот клинически значимых ДНК-маркеров у населения России: биоинформатические аспекты. *Кардиоваскулярная терапия и профилактика.* 2020;19(6):2732. doi:10.15829/1728-8800-2020-2732.
- Balanovska EV, Petrushenko VS, Koshel SM, et al. Cartographic atlas of frequency variation for 45 pharmacogenetic markers in populations of Russia and its neighbor states. *Bulletin of RSMU.* 2020(6):39-52. (In Russ.) Балановская Е. В., Петрушенко В. С., Кошель С. М. и др. Картографический атлас распространения 45 фармакогенетических маркеров в народонаселении России и сопредельных стран. *Вестник Российского государственного медицинского университета.* 2020(6):39-52. doi:10.24075/vrgmu.2020.080.
- Pylev VYu, Agdzhoyan AT, Gorin IO, et al. Population biobank as a basis for determining spatial variation of clinically relevant pharmacogenetic biomarkers of cardiovascular diseases. *Cardiovascular Therapy and Prevention.* 2022;21(11):3430. (In Russ.) Пылев В. Ю., Агджоян А. Т., Горин И. О. и др. Популяционный биобанк как основа для выявления пространственной изменчивости клинически значимых фармакогенетических биомаркеров сердечно-сосудистых заболеваний. *Кардиоваскулярная терапия и профилактика.* 2022;21(11):3430. doi:10.15829/1728-8800-2022-3430.
- Balanovsky O, Petrushenko V, Mirzaev K, et al. Variation of genomic sites associated with severe Covid-19 across populations: global and national patterns. *Pharmacogenomics Pers Med.* 2021;4:1391-402. doi:10.2147/PGPM.S320609.
- Balanovska EV, Gorin IO, Petrushenko VS, et al. Geographic distribution of the LZTFL1 SNP markers associated with severe COVID-19 in Russia and worldwide. *Bulletin of RSMU.* 2022;(5):31-40. (In Russ.) Балановская Е. В., Горин И. О., Петрушенко В. С. и др. Геногеография в России и мире SNP-маркеров гена LZTFL1, ассоциированных с тяжелым течением COVID-19. *Вестник Российского государственного медицинского университета.* 2022;(5):31-40. doi:10.24075/vrgmu.2022.047.
- Balanovska EV, Rychkov Ju G. Ethnic genetics: ethnogeographical diversity of the gene pool of the world's peoples. *Genetics.* 1990;26(1):114-21. (In Russ.) Балановская Е. В., Рычков Ю. Г. Этническая генетика: этногеографическое разнообразие генофонда народов мира. *Генетика.* 1990;26(1):114-21.
- Balanovska EV, Nurbaev SD. The selective structure of the gene pool. I. Possibilities of studying. *Genetics.* 1997;33(11):1572. (In Russ.) Балановская Е. В., Нурбаев С. Д. Селективная структура генофонда. I. Возможности изучения. *Генетика.* 1997;33(11):1572.
- Balanovska EV, Nurbaev SD. Selective structure of the gene pool. II. Standard method using F (ST) statistics. *Genetika.* 1998;34(9):1307-21. (In Russ.) Балановская Е. В., Нурбаев С. Д. Селективная структура генофонда. II. Стандартная технология определения через FST статистики. *Генетика.* 1998;34(9):1307-21.
- Balanovska EV, Nurbaev SD. Selective structure of the gene pool. III. Technology of determination through FST-statistics by means of numerical resampling. *Genetics.* 1998;34(10):1434-46. (In Russ.) Балановская Е. В., Нурбаев С. Д. Селективная структура генофонда. III. Технология определения через FST-статистики с помощью численного ресэмплинга. *Генетика.* 1998;34(10):1434-46.
- Piras I, Falchi A, Moral P, et al. Selective neutrality analysis of 17 STRs in Mediterranean populations. *J Hum Genet.* 2010;55(4):207-14. doi:10.1038/jhg.2010.7.
- Lotterhos KE, Whitlock MC. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Mol Ecol.* 2014;23(9):2178-92. doi:10.1111/mec.12725.
- Bhatia G, Patterson N, Sankararaman S, et al. Estimating and interpreting FST: the impact of rare variants. *Genome Res.* 2013;23(9):1514-21. doi:10.1101/gr.154831.113.
- Jin W, Xu S, Wang H, et al. Genome-wide detection of natural selection in African Americans pre- and post-admixture. *Genome Res.* 2012;22(3):519-27. doi:10.1101/gr.124784.111.

23. McEvoy BP, Montgomery GW, McRae AF, et al. Geographical structure and differential natural selection among North European populations. *Genome Res.* 2009;19(5):804-14. doi:10.1101/gr.083394.108.
24. Barreiro LB, Laval G, Quach H, et al. Natural selection has driven population differentiation in modern humans. *Nat Genet.* 2008;40(3):340-5. doi:10.1038/ng.78.
25. Cherednichenko AA, Trifonova EA, Vagaitseva KV, et al. Association of the genetic polymorphism of cytokines and their receptors with climate and geographic factors in human populations. *Russ J Genet.* 2014;50:1112-6. doi:10.1134/S1022795414100020.
26. Sivadas A, Salleh MZ, Teh LK, et al. Genetic epidemiology of pharmacogenetic variants in South East Asian Malays using whole-genome sequences. *Pharmacogenomics J.* 2017;17(5): 461-70. doi:10.1038/tpj.2016.39.
27. Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4(1):s13742-015-0047-8. doi:10.1186/s13742-015-0047-8.
28. Balanovska EV, Zhabagin MK, Agdzhoyan AT, et al. Population biobanks: organizational models and prospects of application in gene geography and personalized medicine. *Genetics.* 2016;52(12):1371-87. (In Russ.) Балановская Е.В., Жабегин М.К., Агджоян А.Т. и др. Популяционные биобанки: принципы организации и перспективы применения в геногеографии и персонализированной медицине. *Генетика.* 2016;52(12):1371-87. doi:10.7868/S001667581612002X.
29. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2006;2(12):e190. doi:10.1371/journal.pgen.0020190.
30. Koshel' SM. Geoinformation technologies in genogeography. *Modern geographical cartography.* 2012;158-66. (In Russ.) Кошель С.М. Геоинформационные технологии в геногеографии. *Современная географическая картография.* 2012;158-66.