

## *In silico* поиск генов, контролирующих ишемическую болезнь сердца

И.В. Зоркольева<sup>1</sup>✉, Н.М. Белоногова<sup>1</sup>, Г.Р. Свищёва<sup>1,2</sup>, А.В. Кириченко<sup>1</sup>, Т.И. Аксенович<sup>1,3</sup>

<sup>1</sup> Федеральное исследовательское учреждение Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>2</sup> Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, Россия

<sup>3</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

✉ e-mail: zor@bionet.nsc.ru

К настоящему времени с помощью крупномасштабных полногеномных исследований обнаружено более 100 локусов, ассоциированных с ишемической болезнью сердца (ИБС). Для некоторых из нескольких сотен генов, лежащих в этих локусах, была показана их роль в патогенезе болезни. Тем не менее основные генетические механизмы и конкретные гены, контролирующие заболевание, все еще полностью не известны. Данное исследование посвящено *in silico* поиску новых генов, контролирующих ИБС. Проведен региональный анализ ассоциаций, при котором все полиморфные варианты внутри гена анализируются одновременно. Материалом для анализа служили результаты полногеномного анализа ассоциаций, депонированные в открытых базах данных MICAD (120575 человек, 85112 маркеров) и UK Biobank (337199 человек, 10894597 маркеров). Использовался программный пакет sumFREGAT, в котором реализован широкий спектр новых методов для тестирования генных ассоциаций с помощью суммарных статистик. Всего было обнаружено 88 генов. Из них 44 являются уже известными для ИБС генами. Кроме того, в известных локусах нами идентифицировано 28 дополнительных генов, которые можно рассматривать в качестве новых генов-кандидатов. 16 генов (*AGPAT4*, *ARHGEF12*, *BDP1*, *DHX58*, *EHBP1*, *FBF1*, *HSPB9*, *NPBWR2*, *PDLIM5*, *PLCB3*, *PLEKHM2*, *POU2F3*, *PRKD2*, *TMEM136*, *TTC29* и *UTP20*), обнаруженных нами вне известных локусов, являются новыми. Информация о функциональной роли этих генов позволяет рассматривать многие из них в качестве кандидатов для ИБС. Среди набора идентифицированных генов 41 ген не имел значимых сигналов полногеномного анализа ассоциаций и был идентифицирован только благодаря одновременному рассмотрению всех вариантов внутри гена в рамках регионального анализа ассоциаций. Полученные результаты демонстрируют, что региональный анализ ассоциаций представляет собой мощный инструмент для поиска новых генов. Он применим для анализа различных признаков и болезней с использованием накопленных в мире огромных объемов данных, полученных с помощью полногеномного анализа ассоциаций. Такие исследования общедоступны, поскольку не требуют дополнительных материальных затрат.

Ключевые слова: ишемическая болезнь сердца; полногеномный анализ ассоциаций; региональный анализ ассоциаций; суммарные статистики; *in silico* картирование.

**Для цитирования:** Зоркольева И.В., Белоногова Н.М., Свищёва Г.Р., Кириченко А.В., Аксенович Т.И. *In silico* поиск генов, контролирующих ишемическую болезнь сердца. Вавиловский журнал генетики и селекции. 2019;23(8): 1037-1046. DOI 10.18699/VJ19.585

## *In silico* mapping of coronary artery disease genes

I.V. Zorkoltseva<sup>1</sup>✉, N.M. Belonogova<sup>1</sup>, G.R. Svishcheva<sup>1,2</sup>, A.V. Kirichenko<sup>1</sup>, T.I. Aksenovich<sup>1,3</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Vavilov Institute of General Genetics, RAS, Moscow, Russia

<sup>3</sup> Novosibirsk State University, Novosibirsk, Russia

✉ e-mail: zor@bionet.nsc.ru

To date, more than 100 loci associated with coronary artery disease (CAD) have been detected in large-scale genome-wide studies. For some of the several hundreds of genes located in these loci, roles in the pathogenesis of the disease have been shown. However, the genetic mechanisms and specific genes controlling this disease are still not fully understood. This study is aimed at *in silico* search for new CAD genes. We performed a gene-based association analysis, where all polymorphic variants within a gene are analyzed simultaneously. The analysis was based on the results of the genome-wide association studies (GWAS) available from the open databases MICAD (120,575 people, 85,112 markers) and UK Biobank (337,199 people, 10,894,597 markers). We used the sumFREGAT package implementing a wide range of new methods for gene-based association analysis using summary statistics. We found 88 genes demonstrating significant gene-based associations. Forty-four of the identified genes were already known as CAD genes. Furthermore, we identified 28 additional genes in the known CAD loci. They can be considered as new candidate genes. Finally, we identified sixteen new genes (*AGPAT4*, *ARHGEF12*, *BDP1*, *DHX58*, *EHBP1*, *FBF1*, *HSPB9*, *NPBWR2*, *PDLIM5*, *PLCB3*, *PLEKHM2*, *POU2F3*, *PRKD2*, *TMEM136*, *TTC29* and *UTP20*) outside the known loci. Information about the functional role of these genes allows us to consider many of them as candidates for CAD. The 41 identified genes did not have significant GWAS signals and they were identified only due to simultaneous consideration of all variants within the gene in the

framework of gene-based analysis. These results demonstrate that gene-based association analysis is a powerful tool for gene mapping. The method can utilize huge amounts of GWAS results accumulated in the world to map different traits and diseases. This type of studies is widely available, as it does not require additional material costs.

Key words: coronary artery disease; gene-based association analysis; genome-wide association analysis; summary statistics; *in silico* mapping.

**For citation:** Zorkoltseva I.V., Belonogova N.M., Svishcheva G.R., Kirichenko A.V., Axenovich T.I. *In silico* mapping of coronary artery disease genes. *Vavilovskii Zhurnal Genetiki i Selekcii* = *Vavilov Journal of Genetics and Breeding*. 2019; 23(8):1037-1046. DOI 10.18699/VJ19.585 (in Russian)

## Введение

Ишемическая болезнь сердца (ИБС) – одна из основных причин потери трудоспособности и преждевременной смерти во всем мире. ИБС является следствием атеросклероза коронарных артерий, которые питают миокард и снабжают его кислородом. Согласно «Международной статистической классификации болезней, травм и причин смерти» (<http://apps.who.int/classifications/icd10/browse/2010/en#/I51.6>), основные формы ИБС включают стенокардию, инфаркт миокарда (ИМ), кардиосклероз, нарушение сердечного ритма и проводимости, сердечную недостаточность и внезапную коронарную смерть. Клинические проявления ИБС зависят от ее формы: ИМ и внезапная коронарная смерть развиваются в течение нескольких минут, а постинфарктный кардиосклероз – хроническое состояние, которое стабильно при адекватно подобранном лечении. ИБС часто сопровождается гипертонией, диабетом второго типа, гиперхолестеринемией, так как эти заболевания имеют общие факторы риска и патогенетические механизмы (Ehret et al., 2011; Johnson et al., 2011). Различные метаболические и физиологические расстройства, имеющиеся при перечисленных заболеваниях, также содействуют повышению риска ИБС. Развитие ИБС часто происходит на фоне нездорового образа жизни, связанного с курением, неправильным питанием, ожирением, недостатком физической активности и др. (Yusuf et al., 2004).

Известно, что наследуемость ИБС составляет 40–50 % (Peden, Farrall, 2011), но активный поиск молекулярных механизмов, реализующих наследственную предрасположенность к ИБС, длительное время оставался малоуспешным. Многочисленные молекулярно-генетические исследования были направлены на изучение полиморфизмов в функционально значимых генах, принимающих участие в процессах, связанных с патогенезом ИБС: обмене липидов, тромбообразовании, эндотелиальной дисфункции, регуляции ренин-ангиотензиновой системы, воспалительных реакций. Однако лишь для небольшого числа генов, главным образом связанных с метаболизмом липопротеинов низкой плотности (ЛПНП), удалось установить связь с риском развития ИБС (Brown, Goldstein, 1986; Abifadel et al., 2003; Cohen et al., 2006).

Новая эра в изучении ИБС началась, когда благодаря развитию технологий генотипирования стал доступен полногеномный анализ ассоциаций (ПГАА), ставший основным инструментом для выяснения генетической природы сложных признаков. Результаты первых полногеномных исследований ИБС были опубликованы в 2007 г. Три группы исследователей независимо друг от друга обнаружили локус 9p21.3, ассоциированный с ИБС (Hel-

gadottir et al., 2007; McPherson et al., 2007; Samani et al., 2007). Связь между 9p21.3 и ИБС является одной из самых сильных генетических ассоциаций, наблюдаемых в ПГАА (Schunkert et al., 2011), при этом данный локус расположен в так называемой геномной пустыне – регионе, в котором нет генов, кодирующих известные белки. Точный механизм, с помощью которого 9p21.3 влияет на ИБС, до сих пор остается неясным.

Спустя два года в нескольких относительно небольших индивидуальных исследованиях удалось выявить двенадцать дополнительных локусов (Erdmann et al., 2009; Gudbjartsson et al., 2009; Samani et al., 2009; Tregouet et al., 2009). Существенный прогресс был достигнут, когда стали проводить метаанализ, объединяя результаты анализов отдельных исследований. Это стало возможным благодаря созданию международных консорциумов по изучению ИБС и ИМ, например Myocardial Infarction Genetics Consortium (MIGen, 6042 чел.), Coronary Artery Disease (CAD) Genetics Consortium (68065 чел.), IBC 50K CAD Consortium (180183 чел.), Coronary ARtery Disease Genome wide Replication and Meta-analysis consortium (CARDIoGRAM, 86995 чел.).

К настоящему времени в общей сложности известно около 300 однонуклеотидных полиморфных вариантов (single nucleotide polymorphism, SNP), ассоциированных с ИБС, локализованных в более чем 100 локусах (Schunkert et al., 2011; The CARDIoGRAMplusC4D Consortium, 2013; Nikpay et al., 2015; Nioi et al., 2016; Stitzel et al., 2016; Howson et al., 2017; Klarin et al., 2017; Nelson et al., 2017; Webb et al., 2017; van der Harst, Verweij, 2018), которые совместно объясняют около 40 % генетической компоненты ИБС (Nikpay et al., 2015; Hartiala et al., 2017; Nelson et al., 2017; Vilne, Schunkert, 2018).

Трудно объяснить генетическую структуру заболевания только на основе результатов ПГАА, так как подавляющее большинство ассоциированных с болезнью SNP лежит в некодирующих регионах генома. Они не связаны с изменениями структуры и функции белка, а вероятнее всего, оказывают влияние на уровень экспрессии генов. Кроме того, идентифицированные с помощью ПГАА SNP служат своего рода флагом для всех SNP, находящихся в неравновесии по сцеплению с ними и которые также могут быть задействованы в контроле данного заболевания. Вследствие этих причин конкретные гены, непосредственно влияющие на риск развития ИБС, все еще остаются неизвестными.

Несколько лет назад был предложен подход, в основе которого лежит региональный анализ ассоциаций (РАА), или анализ ассоциаций на геномном уровне (gene-based association analysis), при котором все полиморфные варианты

внутри гена анализируются одновременно (Li, Leal, 2008; Eichler et al., 2010). Это позволяет установить статистически значимую ассоциацию между геном и признаком даже в том случае, когда ни один из вариантов внутри гена не дает значимого сигнала. При использовании PAA в анализе можно вовлечь не только распространенные, но и редкие генетические варианты, тем самым увеличивая статистическую мощность анализа. Кроме того, PAA сразу позволяет интерпретировать значимый сигнал ассоциации как указание на роль структуры соответствующего белка в контроле изучаемого признака или болезни.

Применение PAA для анализа ИБС до недавнего времени было затруднено тем, что он требовал огромных выборок, содержащих информацию об индивидуальных фенотипах и генотипах всех членов выборки. Такие данные никогда не депонируются в открытые базы данных, в отличие от результатов ПГАА и метаанализа, представленного набором суммарных статистик для каждого исследованного генетического варианта. В последних работах было показано, что PAA можно проводить *in silico*, используя суммарные статистики, не требуя индивидуальных фенотипов и генотипов (de Leeuw et al., 2015; Bakshi et al., 2016; Wang et al., 2017; Svishecheva et al., 2019).

Целью нашей работы является поиск новых генов, участвующих в контроле ИБС, путем *in silico* картирования с использованием регионального анализа ассоциаций.

## Материалы и методы

**Суммарные статистики.** Использовались два набора суммарных статистик, полученных для ИБС с помощью ПГАА и депонированных в общедоступных базах данных. Первый набор, MICAD (Myocardial Infarction Genetics and CARDIOGRAM Exome), представляет собой результаты метаанализа 20 отдельных когорт. Детальная информация о каждой из них представлена в оригинальной статье (Stitzel et al., 2016). Метаанализ включает 42335 человек с ИБС и/или ИМ и 78240 здоровых людей европейского происхождения. Генотипирование осуществлялось с использованием панелей Illumina HumanExome BeadChip array и Illumina OmniExome, дополнительная информация о которых размещена по адресу [http://genome.sph.umich.edu/wiki/Exome\\_Chip\\_Design](http://genome.sph.umich.edu/wiki/Exome_Chip_Design). Суммарные статистики ( $p$ -value, размер эффекта ( $\beta$ )) для 85 112 SNP были загружены с сайта [www.cardiogramplusc4d.org](http://www.cardiogramplusc4d.org).

Второй набор данных, UKbb, представляет результаты ПГАА хронической ишемической болезни сердца из проекта UK Biobank Resource, включающего 8755 больных и 328444 здоровых человека. Генотипирование осуществлялось на основе метабоципа, специально созданного для ИБС. Подробное описание критериев включения и характеристики фенотипа дано в оригинальных публикациях (Klarin et al., 2017; van der Harst, Verweij, 2018). Суммарные статистики для 10894597 SNP были загружены с сайта [www.ukbiobank.ac.uk](http://www.ukbiobank.ac.uk).

Таким образом, анализируемые выборки различались между собой как по структуре (метаанализ/индивидуальное исследование, число больных и здоровых), так и по набору генотипированных маркеров.

**Региональный анализ ассоциаций** проводили с использованием программного пакета sumFREGAT, соз-

данного коллективом авторов на языке программирования R. Свободный доступ к нему открыт по адресу: <https://CRAN.R-project.org/package=sumFREGAT>. Пакет предназначен для тестирования генных ассоциаций с использованием суммарных статистик. В нем реализован широкий спектр методов для регионального анализа ассоциаций, построенных на обобщенной регрессионной модели с аддитивными генетическими эффектами (Svishecheva, 2019). Эти методы включают: метод множественной линейной регрессии (MLR); методы, использующие функциональный анализ данных (FLM) и анализ главных компонент (PCA); метод коллапсинга или Бурден-тест (BT); метод SKAT, основанный на анализе компонент дисперсии и использующий принцип ядерного сглаживания, а также метод SKAT-O, объединяющий BT и SKAT. Перечисленные методы отличаются стратегией комбинирования суммарных статистик и являются эффективным инструментом для PAA. Кроме них в пакете реализованы еще три метода, использующие два традиционных подхода для объединения суммарных статистик. Один метод основан на сумме хи-квадратов ( $\sum \chi^2$ ) и в двух методах применен минимум  $p$ -value (min  $p$ -value и SimpleM).

Подробно методы и их имплементация описаны в нашей работе (Svishecheva et al., 2019), где также проведена апробация каждого метода на выборках различной структуры с использованием реальных данных, в том числе MICAD и UKbb.

Для проведения анализа на генном уровне с помощью пакета sumFREGAT достаточно задать в качестве входных данных суммарные статистики ПГАА, вычисленные для генетических вариантов исследуемого гена, а именно  $p$ -value и размер эффекта ( $\beta$ ), и матрицы корреляций между генотипами этих вариантов. Для подсчета матриц корреляций мы использовали референсную панель 1000 Genomes Phase 3, включающую 503 индивида европейского происхождения. Границы генов устанавливали в соответствии с GRCh37/hg19. Маркер приписывали к определенному гену, если он находится в его экзоне или интроне.

Уровень значимости определяли с помощью поправки Бонферрони. Сигнал ассоциации считали значимым, если значение  $p$ -value для гена не превышало порогового уровня  $4.58 \cdot 10^{-6}$  ( $0.05/10912$  генов) для выборки MICAD и  $2.78 \cdot 10^{-6}$  ( $0.05/17975$  генов) для выборки UKbb. Мы считали, что ген достоверно ассоциирован с болезнью, если хотя бы в одном из методов значение  $p$ -value достигало порогового уровня.

**Биоинформатический анализ.** Анализ представленности функциональных групп генов выполняли с помощью FUMA GWAS (Functional Mapping and Annotation of Genome-Wide Association Studies) (<http://fuma.ctglab.nl/>), используя стандартные настройки, предлагаемые сервисом.

**Известные локусы ИБС.** На начало данного исследования было известно 97 локусов, ассоциированных с ИБС. Их список приведен в работе (Klarin et al., 2017). Каждый из локусов в этом списке представлен одним или несколькими лидирующими SNP, удовлетворяющими следующим условиям: сигнал ПГАА достигает полногеномного уровня значимости  $5.00 \cdot 10^{-8}$  в оригинальном исследовании и реплицирован на независимой выборке

или в метаанализе. Для каждого локуса в списке приводится также набор генов, которые содержат эти SNP и/или являются ближайшими к ним. В дальнейшем такие локусы и гены мы называем известными.

## Результаты

Анализируемые выборки UKbb и MICAD были генотипированы на основе двух различных платформ, поэтому число генов и число SNP в них сильно варьировало (табл. 1). Гены, содержащие менее двух SNP, были исключены из анализа.

Для выборки UKbb мы получили статистически значимый сигнал ассоциации ( $p$ -value  $< 2.78 \cdot 10^{-6}$ ) хотя бы в одном из тестов для 49 генов. Результаты PAA для этих генов отражены в табл. 2, где для каждого гена приведено наименьшее среди всех тестов значение  $p$ -value. Аналогичная таблица приведена для выборки MICAD (табл. 3). Она содержит список из 61 гена, сигнал ассоциации которых достиг порогового уровня  $4.58 \cdot 10^{-6}$  хотя бы в одном из тестов.

Всего при анализе выборок UKbb и MICAD было идентифицировано 88 генов, ассоциированных с ИБС, 22 из них были определены на обеих выборках (см. рисунок).

Мы разделили все идентифицированные гены на две категории в зависимости от наличия в них значимого SNP-сигнала. Гены со значимым SNP-сигналом могут быть идентифицированы уже на уровне ПГАА, в то время как гены, не включающие такие сигналы, можно идентифицировать только с использованием PAA.

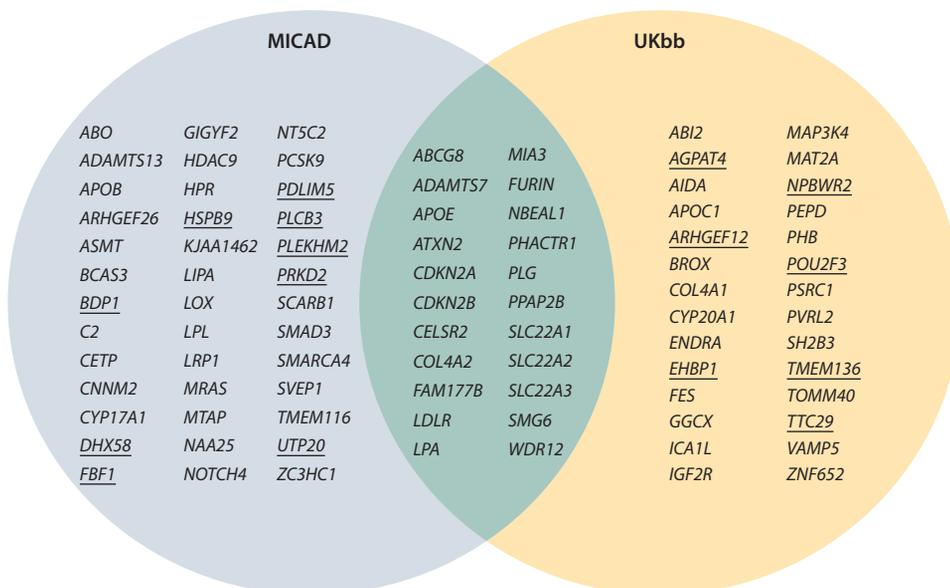
Наибольший интерес представляет вторая категория, к которой относится 41 ген (45.5 % от всех идентифицированных) (табл. 4). Три из них (*COL4A2*, *FURIN* и *SMG6*) были обнаружены с помощью PAA на обеих выборках.

Кроме того, мы сопоставили локализацию идентифицированных генов с известными локусами. Мы определили известный локус как регион  $\pm 500$  тыс. п. н. от позиции полногеномно значимого SNP, маркирующего этот локус. Регионы, находящиеся от него на расстоянии более 500 тыс. п. н., считали новыми. Оказалось, что 72 из 88 генов (81.8 %) лежат в уже известных локусах, ассоциированных с ИБС. 44 из этих генов входят в список известных генов-кандидатов ИБС, а остальные 28 генов могут рассматриваться как новые гены-кандидаты ИБС, лежащие в известных локусах. 16 генов (*AGPAT4*, *ARHGEF12*, *BDP1*, *DHX58*, *EHBP1*, *FBF1*, *HSPB9*, *NPBWR2*, *PDLIM5*, *PLCB3*, *PLEKHM2*, *POU2F3*, *PRKD2*, *TMEM136*, *TTC29* и *UTP20*) были идентифицированы нами за пределами известных локусов, что позволяет считать их новыми потенциальными генами для ИБС (см. табл. 4). Эти гены расположены в 13 новых локусах на хромосомах 1, 2, 4, 5, 6, 11, 12, 17, 19 и 20.

Анализ представленности функциональных групп генов был проведен для всего набора идентифицированных генов. Значительная часть этих групп включала метаболические пути, связанные с обменом липидов, например белок-липидный комплекс, связывание липопротеинов,

**Table 1.** Statistics for UKbb and MICAD data

Sample	Platform	Number of SNPs in gene				Number of genes	Total number of SNPs
		min	mean	median	max		
MICAD	Exome-chip	2	5.3	4	231	10912	57815
UKbb	MetaboChip for CAD	2	237.4	85	21287	17975	4267351



Genes associated with CAD in two samples. Sixteen new genes are underlined.

**Table 2.** Results of gene-based analyses of UKbb data

Chr	Gene	Position		Number of SNPs	p-value	Method	Lead SNP	
		Start	End				p-value	SNP
1	<i>PPAP2B</i>	56960419	57045257	325	2.96E-07	PCA	6.94E-08	rs17114036
	<i>CELSR2</i>	109792641	109818378	99	1.58E-14	min p-val	<b>4.10E-16</b>	rs660240
	<i>PSRC1</i>	109822176	109825790	14	2.07E-06	MLR. FLM	2.66E-04	rs35358959
	<i>MIA3</i>	222791444	222841351	170	5.85E-09	sum $\chi^2$	<b>2.28E-09</b>	rs2133189
	<i>AIDA</i>	222841355	222886602	107	6.14E-08	SKAT-O	<b>3.89E-09</b>	rs36079339
	<i>BROX</i>	222885906	222908529	49	1.19E-07	SKAT-O	<b>3.97E-08</b>	rs12043288
	<i>FAM177B</i>	222910558	222926932	44	2.01E-07	sum $\chi^2$	<b>2.90E-08</b>	rs34885880
2	<i>ABCG8</i>	44066103	44105947	248	3.91E-07	PCA	1.97E-07	rs116426890
	<i>EHBP1</i>	62900986	63273622	739	9.67E-07	SKAT	4.23E-05	rs13410889
	<i>MAT2A</i>	85766101	85772403	14	9.96E-08	PCA	<b>1.39E-08</b>	rs2028900
	<i>GGCX</i>	85771978	85788657	60	2.01E-07	SKAT-O	8.43E-08	rs7605975
	<i>VAMP5</i>	85811531	85820511	29	1.67E-06	BT	1.64E-07	rs55971080
	<i>ICA1L</i>	203637873	203736708	292	2.07E-08	sum $\chi^2$	<b>4.60E-09</b>	rs72932711
	<i>WDR12</i>	203745323	203776949	110	1.49E-08	sum $\chi^2$	<b>7.47E-09</b>	rs6722332
	<i>NBEAL1</i>	203879597	204091101	485	3.12E-07	min p-val	<b>2.53E-09</b>	rs72934573
	<i>CYP20A1</i>	204103164	204170563	188	8.33E-07	PCA	5.65E-08	rs115600411
4	<i>TTC29</i>	147628179	147867034	758	3.43E-07	SimpleM	<b>1.89E-09</b>	rs534388792
	<i>EDNRA</i>	148402069	148466106	171	1.68E-09	FLM	<b>8.29E-09</b>	rs77249653
6	<i>PHACTR1</i>	12716888	13290476	1980	1.99E-11	min p-val	<b>1.52E-13</b>	rs9349379
	<i>IGF2R</i>	160390131	160527583	473	3.94E-12	PCA	<b>7.58E-14</b>	rs3777402
	<i>SLC22A1</i>	160542821	160580147	139	5.54E-12	min p-val	<b>1.33E-13</b>	rs2282143
	<i>SLC22A2</i>	160637794	160679963	146	1.97E-07	min p-val	<b>4.29E-09</b>	rs184091076
	<i>SLC22A3</i>	160769405	160873613	361	4.47E-50	MLR	<b>2.65E-27</b>	rs4709431
	<i>LPA</i>	160952506	161097478	410	2.85E-97	MLR	<b>1.03E-48</b>	rs55730499
	<i>PLG</i>	161123225	161175086	203	1.69E-52	MLR	<b>1.10E-32</b>	rs4252185
	<i>MAP3K4</i>	161412816	161538417	461	7.22E-11	min p-val	<b>1.03E-12</b>	rs141766382
	<i>AGPAT4</i>	161551057	161695107	555	7.08E-07	min p-val	<b>7.81E-09</b>	rs148349043
9	<i>CDKN2A</i>	21967751	21994490	62	1.82E-17	MLR	<b>4.21E-14</b>	rs3731239
	<i>CDKN2B</i>	22002902	22009312	15	1.20E-33	PCA	<b>1.19E-32</b>	rs3217992
11	<i>POU2F3</i>	120107349	120190653	305	1.28E-06	PCA	2.01E-04	rs138488120
	<i>TMEM136</i>	120195838	120204397	27	2.27E-06	MLR. FLM	2.86E-06	rs1893261
	<i>ARHGEF12</i>	120207264	120360645	398	1.08E-06	sum $\chi^2$	3.07E-07	rs12417256
12	<i>SH2B3</i>	111843720	111889427	85	8.79E-07	PCA	2.35E-07	rs3184504
	<i>ATXN2</i>	111890018	112037480	258	1.27E-06	PCA	1.78E-07	rs4766578
13	<i>COL4A1</i>	110801310	110959496	863	4.36E-11	PCA	3.56E-06	rs9521642
	<i>COL4A2</i>	110959631	111165374	1252	1.79E-06	FLM	5.46E-06	rs9515203
15	<i>ADAMTS7</i>	79051545	79103805	202	1.66E-09	PCA	<b>1.47E-08</b>	rs2904223
	<i>FURIN</i>	91411885	91426687	37	4.91E-07	sum $\chi^2$	6.27E-08	rs8039305
	<i>FES</i>	91427665	91439006	38	2.58E-08	sum $\chi^2$	1.28E-07	rs7177338
17	<i>SMG6</i>	1963133	2207069	846	1.85E-06	BT	1.11E-06	rs67457628
	<i>ZNF652</i>	47366568	47439835	209	3.05E-06	sum $\chi^2$	1.13E-06	rs62076439
	<i>PHB</i>	47481410	47492267	36	2.63E-07	sum $\chi^2$	<b>2.30E-08</b>	rs7502499
19	<i>LDLR</i>	11200037	11244506	220	2.14E-08	PCA	<b>2.62E-08</b>	rs6511720
	<i>PEPD</i>	33877855	34012799	754	1.64E-07	min p-val	<b>1.91E-09</b>	rs145436496
	<i>PVRL2</i>	45349393	45392485	221	4.61E-09	FLM	9.24E-08	rs6857
	<i>TOMM40</i>	45394477	45406946	61	6.74E-07	PCA	7.47E-08	rs2075650
	<i>APOE</i>	45409039	45412650	9	4.00E-15	PCA	<b>2.64E-09</b>	rs429358
	<i>APOC1</i>	45417577	45422606	11	3.99E-09	SKAT	<b>4.01E-08</b>	rs12721051
20	<i>NPBWR2</i>	62737183	62738184	6	1.92E-07	SKAT-O	1.06E-07	rs13036542

Notes. Here and in Table 3: gene positions are indicated with reference to GRCh37/hg19; new genes are indicated in bold. p-values < 5.0 · 10<sup>-8</sup> are indicated in bold.

**Table 3.** Results of gene-based analyses of MICAD data

Chr	Gene	Position		Number of SNPs	p-value	Method	Lead SNP	
		Start	End				p-value	SNP
1	<b>PLEKHM2</b>	16010632	16061264	4	2.47E-06	SKAT	3.72E-06	rs12091750
	PCSK9	55505149	55530526	5	3.21E-11	MLR	<b>8.03E-10</b>	rs11591147
	PPAP2B	56960419	57045257	5	2.79E-10	SKAT-O	<b>4.55E-10</b>	rs9970807
	CELSR2	109792641	109818378	18	1.43E-22	SKAT	<b>1.13E-23</b>	rs12740374
	MIA3	222791444	222841351	11	9.03E-14	min p-val	<b>1.22E-14</b>	rs17465637
	FAM177B	222910558	222926932	3	1.75E-10	SKAT	<b>1.12E-09</b>	rs2378607
2	APOB	21224301	21266945	33	2.60E-06	MLR	1.09E-04	rs533617
	ABCG8	44066103	44105947	10	1.30E-09	SKAT	<b>4.88E-09</b>	rs4299376
	WDR12	203745323	203776949	2	1.24E-15	PCA	<b>1.25E-15</b>	rs35212307
	NBEAL1	203879597	204091101	14	3.11E-14	min p-val	<b>2.26E-15</b>	rs2351524
	GIGYF2	233562015	233725287	5	8.17E-09	PCA	<b>1.46E-08</b>	rs1801251
3	MRAS	138066490	138124377	2	1.83E-08	min p-val	<b>1.83E-08</b>	rs2306374
	ARHGGEF26	153838792	153975616	3	4.92E-09	sum $\chi^2$	<b>8.28E-09</b>	rs12493885
4	<b>PDLIM5</b>	95373008	95589378	6	2.30E-06	sum $\chi^2$	5.97E-07	rs2452600
5	<b>BDP1</b>	70751442	70863649	18	8.26E-07	BT	3.84E-03	rs12187098
	LOX	121398890	121414206	2	1.40E-07	SKAT	1.72E-07	rs1800449
6	PHACTR1	12716888	13290476	7	5.17E-35	min p-val	<b>8.61E-36</b>	rs9349379
	C2	31865562	31913449	18	1.02E-06	SimpleM	7.87E-08	rs3130683
	NOTCH4	32162620	32191844	33	2.41E-06	MLR	8.98E-06	rs394657
	SLC22A1	160542821	160580147	15	2.47E-11	min p-val	<b>1.78E-12</b>	rs2282143
	SLC22A2	160637794	160679963	3	2.18E-10	min p-val	<b>1.09E-10</b>	rs2279463
	SLC22A3	160769405	160873613	5	7.41E-26	MLR	<b>6.98E-21</b>	rs1810126
	LPA	160952506	161097478	17	1.16E-28	MLR	<b>9.74E-24</b>	rs3798220
	PLG	161123225	161175086	11	8.76E-10	PCA	<b>2.62E-10</b>	rs4252120
7	HDAC9	18126572	19039135	6	1.92E-06	min p-val	3.20E-07	rs2023938
	ZC3HC1	129658126	129691233	4	1.30E-12	min p-val	<b>4.11E-13</b>	rs11556924
8	LPL	19796582	19824770	9	6.78E-09	MLR	6.99E-07	rs264
9	MTAP	21802635	21941040	4	2.34E-07	min p-val	7.76E-08	rs7023954
	CDKN2A	21967751	21994490	2	8.11E-07	PCA	8.46E-07	rs3731249
	CDKN2B	22002902	22009312	2	1.67E-37	PCA	<b>1.47E-33</b>	rs1063192
	SVEP1	113127528	113342160	35	2.91E-06	min p-val	1.05E-07	rs111245230
	ABO	136130563	136150630	22	2.42E-18	MLR	<b>9.31E-17</b>	rs507666
	ADAMTS13	136279459	136324525	14	9.78E-07	PCA	1.63E-07	rs4962153
10	KIAA1462	30301729	30348488	8	8.68E-09	SKAT-O	<b>3.94E-09</b>	rs3739998
	LIPA	90973326	91011660	6	1.22E-15	min p-val	<b>3.04E-16</b>	rs2246942
	CYP17A1	104590288	104597290	3	1.41E-06	SKAT-O	1.85E-06	rs1004467
	AS3MT	104629183	104661656	4	2.97E-11	SKAT-O	<b>2.16E-08</b>	rs11191447
	CNNM2	104678075	104838344	2	2.86E-09	sum $\chi^2$	<b>4.46E-08</b>	rs12413409
	NT5C2	104847774	104953063	2	5.56E-08	min p-val	<b>2.78E-08</b>	rs11191580
11	<b>PLCB3</b>	64018995	64036924	6	3.75E-07	MLR	3.98E-05	rs12146487
	LRP1	57522282	57607142	17	6.10E-08	PCA	<b>1.78E-08</b>	rs11172113
12	UTP20	101673905	101780397	6	2.42E-07	BT	3.31E-05	rs2305858
	ATXN2	111890018	112037480	7	1.67E-13	PCA	<b>2.50E-13</b>	rs653178
	TMEM116	112369086	112451023	2	2.91E-07	PCA	2.91E-07	rs3752630
	NAA25	112464493	112546826	9	4.50E-08	min p-val	<b>8.20E-09</b>	rs17696736
	SCARB1	125262174	125348519	5	1.48E-09	min p-val	<b>3.69E-10</b>	rs11057830
	COL4A2	110959631	111165374	13	6.00E-07	SKAT	9.09E-05	rs3809346
15	SMAD3	67358036	67487533	3	2.15E-06	min p-val	8.77E-07	rs17293632
	ADAMTS7	79051545	79103805	7	8.87E-07	min p-val	4.43E-07	rs1994016
	FURIN	91411885	91426687	4	1.64E-06	PCA	1.64E-06	rs17514846
	CETP	56995835	57017757	7	8.80E-07	SKAT	1.12E-06	rs1532624
16	HPR	72097125	72111145	2	7.55E-08	PCA	7.57E-08	rs2000999
	SMG6	1963133	2207069	11	7.76E-07	SKAT-O	3.17E-06	rs903160
17	<b>DHX58</b>	40253422	40264751	23	2.97E-06	min p-val	5.93E-07	rs2074158
	HSPB9	40274756	40275371	2	2.41E-06	SKAT-O	3.01E-06	rs1122326
	BCAS3	58755172	59470199	9	2.06E-07	PCA	7.25E-08	rs8080784
	FBF1	73905925	73937119	6	1.48E-06	SKAT	9.13E-07	rs1135889
	SMARCA4	11071598	11172958	4	1.45E-10	SKAT	<b>1.02E-10</b>	rs1122608
19	LDLR	11200037	11244506	3	6.84E-19	min p-val	<b>2.28E-19</b>	rs6511720
	APOE	45409039	45412650	3	5.17E-19	PCA	<b>9.97E-16</b>	rs7412
	PRKD2	47177573	47220384	11	2.60E-06	SKAT	2.05E-05	rs425105

**Table 4.** Location of CAD-associated genes identified by gene-based analyses

Location	Genes with significant GWAS signal	Genes without significant GWAS signal
Known genes in known loci	<b>ABCG8</b> , <b>ABO</b> , <b>ADAMTS7</b> , <b>APOC1</b> , <b>APOE</b> , <b>ARHGEF26</b> , <b>CDKN2B</b> , <b>CNNM2</b> , <b>EDNRA</b> , <b>GIGYF2</b> , <b>KIAA1462</b> , <b>LDLR</b> , <b>LIPA</b> , <b>LPA</b> , <b>LPL</b> , <b>LRP1</b> , <b>MIA3</b> , <b>MRAS</b> , <b>NT5C2</b> , <b>PCSK9</b> , <b>PHACTR1</b> , <b>PLG</b> , <b>PPAP2B</b> , <b>SCARB1</b> , <b>SLC22A3</b> , <b>WDR12</b> , <b>ZC3HC1</b>	<b>APOB</b> , <b>BCAS3</b> , <b>C2</b> , <b>CETP</b> , <b>COL4A1</b> , <b>COL4A2</b> , <b>CYP17A1</b> , <b>FES</b> , <b>FURIN</b> , <b>GGCX</b> , <b>HDAC9</b> , <b>LOX</b> , <b>SH2B3</b> , <b>SMAD3</b> , <b>SMG6</b> , <b>SVEP1</b> , <b>VAMP5</b>
New genes in known loci	<b>AIDA</b> , <b>AS3MT</b> , <b>ATXN2</b> , <b>BROX</b> , <b>CDKN2A</b> , <b>CELSR2</b> , <b>FAM177B</b> , <b>ICA1L</b> , <b>IGF2R</b> , <b>MAP3K4</b> , <b>MAT2A</b> , <b>NAA25</b> , <b>NBEAL1</b> , <b>PEPD</b> , <b>PHB</b> , <b>SLC22A1</b> , <b>SLC22A2</b> , <b>SMARCA4</b>	<b>ADAMTS13</b> , <b>CYP20A1</b> , <b>HPR</b> , <b>MTAP</b> , <b>NOTCH4</b> , <b>PSRC1</b> , <b>PVRL2</b> , <b>TMEM116</b> , <b>TOMM40</b> , <b>ZNF652</b>
New genes in new loci	<b>AGPAT4</b> , <b>TTC29</b>	<b>ARHGEF12</b> , <b>BDP1</b> , <b>DHX58</b> , <b>EHBP1</b> , <b>FBF1</b> , <b>HSPB9</b> , <b>NPBWR2</b> , <b>PDLIM5</b> , <b>PLCB3</b> , <b>PLEKHM2</b> , <b>POU2F3</b> , <b>PRKD2</b> , <b>TMEM136</b> , <b>UTP20</b>

Notes. Genes identified in both samples are indicated in bold.

**Table 5.** Ontologies of new genes

Gene ontology	Number of genes		p-value	Genes
	in gene ontology	from the given set		
Neutral lipid metabolic process	83	7	9.39E-08	<b>PCSK9</b> , <b>CETP</b> , <b>APOE</b> , <b>APOC1</b> , <b>APOB</b> , <b>AGPAT4</b> , <b>LPL</b>
Catabolic processes within a body	947	17	9.42E-07	<b>LIPA</b> , <b>NT5C2</b> , <b>PLCB3</b> , <b>SCARB1</b> , <b>COL4A1</b> , <b>COL4A2</b> , <b>ADAMTS7</b> , <b>FURIN</b> , <b>HPR</b> , <b>PEPD</b> , <b>APOE</b> , <b>PRKD2</b> , <b>APOB</b> , <b>SLC22A3</b> , <b>LPL</b> , <b>MTAP</b> , <b>CDKN2A</b>
Alcohol metabolism	347	10	3.68E-06	<b>PCSK9</b> , <b>LIPA</b> , <b>CYP17A1</b> , <b>PLCB3</b> , <b>SCARB1</b> , <b>CETP</b> , <b>LDLR</b> , <b>APOE</b> , <b>APOC1</b> , <b>APOB</b>
Vasculature development	468	11	8.17E-06	<b>PPAP2B</b> , <b>LRP1</b> , <b>COL4A1</b> , <b>COL4A2</b> , <b>BCAS3</b> , <b>APOE</b> , <b>PRKD2</b> , <b>APOB</b> , <b>EDNRA</b> , <b>LOX</b> , <b>NOTCH4</b>
Circulatory system development	786	14	1.07E-05	<b>PPAP2B</b> , <b>LRP1</b> , <b>COL4A1</b> , <b>COL4A2</b> , <b>SMAD3</b> , <b>BCAS3</b> , <b>APOE</b> , <b>PRKD2</b> , <b>APOB</b> , <b>PDLIM5</b> , <b>EDNRA</b> , <b>LOX</b> , <b>NOTCH4</b> , <b>HDAC9</b>
Regulation of endothelial cell migration	114	6	1.28E-05	<b>PPAP2B</b> , <b>SCARB1</b> , <b>BCAS3</b> , <b>APOE</b> , <b>PRKD2</b> , <b>HDAC9</b>
Vessel morphogenesis	364	9	3.88E-05	<b>LRP1</b> , <b>COL4A1</b> , <b>COL4A2</b> , <b>BCAS3</b> , <b>APOE</b> , <b>PRKD2</b> , <b>APOB</b> , <b>EDNRA</b> , <b>NOTCH4</b>
Organic hydroxy compound metabolism	481	10	6.12E-05	<b>PCSK9</b> , <b>LIPA</b> , <b>CYP17A1</b> , <b>PLCB3</b> , <b>SCARB1</b> , <b>CETP</b> , <b>LDLR</b> , <b>APOE</b> , <b>APOC1</b> , <b>APOB</b>

Notes. New genes are indicated in bold.

транспорт фосфолипидов и др. Этот результат хорошо согласуется с липопротеидной теорией развития ИБС. Онтологии, включающие новые гены, приведены в табл. 5. Как видно, во все функциональные группы, обнаруженные для новых генов, входят также известные гены. Большая часть этих групп относится к формированию кровеносных сосудов.

## Обсуждение

В данном исследовании было проведено *in silico* картирование ИБС с использованием методов РАА. При анализе двух выборок европейского происхождения МІСАД и UKbb нами идентифицировано 88 генов, ассоциированных с ИБС, 16 из которых являются новыми.

Научная новизна нашего подхода обеспечена как спецификой материала, так и новизной статистических методов анализа. Анализ был выполнен на материале, представленном суммарными статистиками, такими как *p*-value и размер эффекта ( $\beta$ ), являющимися результатом ПГАА. Использование суммарных статистик для РАА обладает рядом существенных преимуществ перед традиционным

подходом, основанным на материале, представленном индивидуальными фенотипами и генотипами. Во-первых, при использовании суммарных статистик не возникает этических проблем, связанных с доступом к персональной информации, что позволяет размещать их в базах данных с открытым доступом (Pasaniuc, Price, 2017). К настоящему времени существуют уже десятки таких баз данных (Zheng et al., 2017). Во-вторых, такой подход позволяет вовлечь в анализ материал, на порядки превышающий по объему каждое индивидуальное исследование, так как суммарные статистики часто являются результатом метаанализа, объединяющего ресурсы отдельных исследований. Кроме того, анализ суммарных статистик освобождает исследователя от проблем, связанных с подразделенностью выборки, наличием в ней скрытого родства, отклонением распределения анализируемого признака от нормального, наличием выбросов и др., поскольку эти проблемы обычно устранены при проведении ПГАА (Yang et al., 2018).

Новизна методического подхода заключалась в следующем. Во-первых, мы использовали РАА, который позволил нам обнаружить 40 генов, не имеющих значимых сигналов

ПГАА в анализируемых выборках. При этом для трех генов, *COL4A2*, *FURIN* и *SMG6*, связь с ИБС была ранее установлена на других выборках, и они уже вошли в список генов-кандидатов в известных локусах (Klägin et al., 2017). То, что мы их обнаружили с помощью ПГАА, служит положительным контролем нашего анализа. Во-вторых, мы использовали широкий спектр методов объединения суммарных статистик, реализованных в созданном ранее авторами этой статьи пакете sumFREGAT (Svishcheva et al., 2019). Кроме методов минимума *p*-value и суммы хи-квадратов, этот пакет содержит ряд мощных методов, не реализованных ни в одном из существующих пакетов для ПГАА. Применение всего набора методов позволяет существенно увеличить мощность анализа и получить максимально полную информацию о генетической природе заболевания.

Анализ ассоциаций на геномном уровне позволил приоритизировать гены в известных локусах. Традиционно в качестве генов-кандидатов рассматривают гены, включающие самый сильный сигнал ПГАА и/или ближайшие к нему гены. Однако из-за неравновесия по сцеплению к развитию заболевания могут быть причастны и более удаленные от лидирующего SNP гены. Мы рассмотрели регион 1 млн п. н. вокруг позиции SNP, маркирующего известный локус, так как показано, что свыше 99 % генетических вариантов, находящихся в неравновесии по сцеплению с этим SNP, локализованы в окне размером 1 сМ (Bulik-Sullivan et al., 2015). Помимо 44 генов из числа известных, в эти регионы попали 28 новых генов, которые можно считать приоритетными для дальнейшего исследования известных локусов. Некоторые из них уже появлялись в поле зрения ученых в связи с участием продуктов этих генов в различных метаболических процессах. Например, в исследовании (Samani et al., 2008) авторы предположили, что локус вблизи генов *PSRC1* и *CELSR2* повышает риск развития коронарного атеросклероза посредством влияния на уровень липопротеинов низкой плотности. Важную роль в метаболизме лекарственных средств играют гены *SLC22A1* и *SLC22A2*, которые кодируют белки-транспортёры, отвечающие за попадание в клетки органических катионов. Мутации в гене *SLC22A1*, снижающие его экспрессию, приводят к целому ряду различных негативных последствий на энергетический обмен (Zhou et al., 2009). Известно также, что ключевым ферментом для метаболизма лекарств и синтеза холестерина, стероидов и других липидов является цитохром P450, кодируемый семейством генов *CYP*, к которому принадлежит *CYP20A1* (Zanger, Schwab, 2013).

Идентифицированные нами 16 генов попали в 13 новых локусов, расположенных на 10 хромосомах. Гены *AGPAT4*, *PLCB3*, *PRKD2*, *PDLIM5* были кластеризованы в группы, ассоциированные с несколькими биологическими процессами, которые раньше уже были ассоциированы с ИБС и ИМ. Известно, что белок, кодируемый геном *AGPAT4*, играет важную роль в поддержании нормального уровня докозагексаеновой кислоты (основного структурного компонента липидов в мембране клеток головного мозга) (Eto et al., 2014). Гены *PLCB3* и *PRKD2* участвуют в процессах пролиферации и миграции эндотелиальных клеток, необходимых для ангиогенеза (Bhattacharya et al., 2009). Ген

*PDLIM5* участвует в процессе экспансии кардиомиоцитов и связан с развитием дилатационной кардиомиопатии (Bang et al., 2014).

Двенадцать новых генов, *BDP1*, *ARHGEF12*, *DHX58*, *EHBPI*, *FBF1*, *HSPB9*, *NPBWR2*, *PLEKHM2*, *POU2F3*, *TMEM136*, *TTC29* и *UTP20*, не вошли в состав статистически значимых функциональных кластеров. Тем не менее известно, что ген *UTP20* контролирует кальцификацию коронарной артерии (O'Donnell et al., 2007), а *EHBPI* связан с уровнями липидов (Willer et al., 2013). Мутации в гене *PLEKHM2* приводят к аномальной локализации лизосом и нарушению механизмов аутофагии и являются причиной рецессивной дилатационной кардиомиопатии и некомпактной кардиомиопатии левого желудочка (Muhammad et al., 2015). Полученные нами данные о причастности этих генов к ИБС могут способствовать лучшему пониманию этиологии и патогенеза этой болезни.

## Заключение

Основной задачей современного здравоохранения является переход к персонализированной медицине, в основе которой лежит учет индивидуальных особенностей пациентов, в частности наличие генетических маркеров болезни. Очевидно, чем полнее список таких маркеров, тем эффективнее профилактика и лечение болезни. Полученные нами результаты открывают новые возможности в этом направлении. Они демонстрируют, что ПГАА служит мощным инструментом для поиска новых генов, позволяющим извлекать новую информацию из накопленных в мире огромных объемов данных, полученных с помощью ПГАА, без дополнительных материальных затрат.

## Список литературы / References

- Abifadel M., Varret M., Rabes J.P., Allard D., Ouguerram K., Devillers M., Cruaud C., Benjannet S., Wickham L., Erlich D., Derre A., Villegier L., Farnier M., Beucler I., Bruckert E., Chambaz J., Chanut B., Lecerf J.M., Luc G., Moulin P., Weissenbach J., Prat A., Krempf M., Junien C., Seidah N.G., Boileau C. Mutations in *PCSK9* cause autosomal dominant hypercholesterolemia. *Nat. Genet.* 2003; 34:154-156. DOI 10.1038/ng1161.
- Bakshi A., Zhu Z., Vinkhuyzen A.A., Hill W.D., McRae A.F., Visscher P.M., Yang J. Fast set-based association analysis using summary data from GWAS identifies novel gene loci for human complex traits. *Sci. Rep.* 2016;6:32894. DOI 10.1038/srep32894.
- Bang C., Batkai S., Dangwal S., Gupta S.K., Foinquinos A., Holzmann A., Just A., Remke J., Zimmer K., Zeug A., Ponimaskin E., Schmiedl A., Yin X., Mayr M., Halder R., Fischer A., Engelhardt S., Wei Y., Schober A., Fiedler J., Thum T. Cardiac fibroblast-derived microRNA passenger strand-enriched exosomes mediate cardiomyocyte hypertrophy. *J. Clin. Invest.* 2014;124:2136-2146. DOI 10.1172/JCI70577.
- Bhattacharya R., Kwon J., Li X., Wang E., Patra S., Bida J.P., Bajzer Z., Claesson-Welsh L., Mukhopadhyay D. Distinct role of PLCβ3 in VEGF-mediated directional migration and vascular sprouting. *J. Cell Sci.* 2009;122:1025-1034. DOI 10.1242/jcs.041913.
- Brown M.S., Goldstein J.L. A receptor-mediated pathway for cholesterol homeostasis. *Science.* 1986;232:34-47.
- Bulik-Sullivan B.K., Loh P.R., Finucane H.K., Ripke S., Yang J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson N., Daly M.J., Price A.L., Neale B.M. LD Score regression distinguishes confounding from polygenicity in ge-

- nome-wide association studies. *Nat. Genet.* 2015;47:291-295. DOI 10.1038/ng.3211.
- Cohen J.C., Boerwinkle E., Mosley T.H., Jr., Hobbs H.H. Sequence variations in *PCSK9*, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* 2006;354:1264-1272. DOI 10.1056/NEJMoa054013.
- de Leeuw C.A., Mooij J.M., Heskes T., Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* 2015;11:e1004219. DOI 10.1371/journal.pcbi.1004219.
- Ehret G.B., Munroe P.B., Rice K.M., Bochud M., Johnson A.D., Chasman D.I., Smith A.V., ... van Duijn C.M., Newton-Cheh C., Levy D., Caulfield M.J., Johnson T. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature.* 2011;478:103-109. DOI 10.1038/nature10405.
- Eichler E.E., Flint J., Gibson G., Kong A., Leal S.M., Moore J.H., Nadeau J.H. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat. Rev. Genet.* 2010;11:446-450. DOI 10.1038/nrg2809.
- Erdmann J., Grosshennig A., Braund P.S., König I.R., Hengstenberg C., Hall A.S., Linsel-Nitschke P., ... Deloukas P., Thompson J.R., Ziegler A., Samani N.J., Schunkert H. New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat. Genet.* 2009;41:280-282. DOI 10.1038/ng.307.
- Eto M., Shindou H., Shimizu T. A novel lysophosphatidic acid acyltransferase enzyme (LPAAT4) with a possible role for incorporating docosahexaenoic acid into brain glycerophospholipids. *Biochem. Biophys. Res. Commun.* 2014;443:718-724. DOI 10.1016/j.bbrc.2013.12.043.
- Gudbjartsson D.F., Holm H., Gretarsdottir S., Thorleifsson G., Walters G.B., Thorgeirsson G., Gulcher J., Mathiesen E.B., Njolstad I., Nyrnes A., Wilsgaard T., Hald E.M., Hveem K., Stoltenberg C., Kucera G., Stubblefield T., Carter S., Roden D., Ng M.C., Baum L., So W.Y., Wong K.S., Chan J.C., Gieger C., Wichmann H.E., Gschwendtner A., Dichgans M., Kuhlbaumer G., Berger K., Ringelstein E.B., Bevan S., Markus H.S., Kostulas K., Hillert J., Sveinbjornsdottir S., Valdimarsson E.M., Lochen M.L., Ma R.C., Darbar D., Kong A., Arnar D.O., Thorsteinsdottir U., Stefansson K. A sequence variant in *ZFHX3* on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat. Genet.* 2009;41:876-878. DOI 10.1038/ng.417.
- Hartiala J., Schwartzman W.S., Gabbay J., Ghazalpour A., Bennett B.J., Allayee H. The genetic architecture of coronary artery disease: current knowledge and future opportunities. *Curr. Atheroscler. Rep.* 2017;19:6. DOI 10.1007/s11883-017-0641-6.
- Helgadóttir A., Thorleifsson G., Manolescu A., Gretarsdottir S., Blondal T., Jonasdóttir A., Jonasdóttir A., Sigurdsson A., Baker A., Palsson A., Masson G., Gudbjartsson D.F., Magnusson K.P., Andersen K., Levey A.I., Backman V.M., Matthiasdóttir S., Jonsdóttir T., Palsson S., Einarsdóttir H., Gunnarsdóttir S., Gylfason A., Vaccarino V., Hooper W.C., Reilly M.P., Granger C.B., Austin H., Rader D.J., Shah S.H., Quyyumi A.A., Gulcher J.R., Thorgeirsson G., Thorsteinsdóttir U., Kong A., Stefansson K. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science.* 2007;316:1491-1493. DOI 10.1126/science.1142842.
- Howson J.M.M., Zhao W., Barnes D.R., Ho W.K., Young R., Paul D.S., Waite L.L., ... Nordestgaard B.G., Assimes T.L., Danesh J., Butterworth A.S., Saleheen D. Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. *Nat. Genet.* 2017;49:1113-1119. DOI 10.1038/ng.3874.
- Johnson A.D., Newton-Cheh C., Chasman D.I., Ehret G.B., Johnson T., Rose L., Rice K., Verwoert G.C., Launer L.J., Gudnason V., Larson M.G., Chakravarti A., Psaty B.M., Caulfield M., van Duijn C.M., Ridker P.M., Munroe P.B., Levy D., on behalf of the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium, Global BPgen Consortium, Women's Genome Health Study. Association of hypertension drug target genes with blood pressure and hypertension in 86,588 individuals. *Hypertension.* 2011;57:903-910. DOI 10.1161/HYPERTENSIONAHA.110.158667.
- Klarin D., Zhu Q.M., Emdin C.A., Chaffin M., Horner S., McMillan B.J., Leed A., Weale M.E., Spencer C.C.A., Aguet F., Segre A.V., Ardlie K.G., Khera A.V., Kaushik V.K., Natarajan P., Consortium C.A.D., Kathiresan S. Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. *Nat. Genet.* 2017;49:1392-1397. DOI 10.1038/ng.3914.
- Li B., Leal S.M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am. J. Hum. Genet.* 2008;83:311-321. DOI 10.1016/j.ajhg.2008.06.024.
- McPherson R., Pertsemliadis A., Kavaslar N., Stewart A., Roberts R., Cox D.R., Hinds D.A., Pennacchio L.A., Tybjaerg-Hansen A., Folsom A.R., Boerwinkle E., Hobbs H.H., Cohen J.C. A common allele on chromosome 9 associated with coronary heart disease. *Science.* 2007;316:1488-1491. DOI 10.1126/science.1142447.
- Muhammad E., Levitas A., Singh S.R., Braiman A., Ofir R., Etzion S., Sheffield V.C., Etzion Y., Carrier L., Parvari R. *PLEKHM2* mutation leads to abnormal localization of lysosomes, impaired autophagy flux and associates with recessive dilated cardiomyopathy and left ventricular noncompaction. *Hum. Mol. Genet.* 2015;24:7227-7240. DOI 10.1093/hmg/ddv423.
- Nelson C.P., Goel A., Butterworth A.S., Kanoni S., Webb T.R., Marrouli E., Zeng L., ... Farrall M., Danesh J., Samani N.J., Watkins H., Deloukas P. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat. Genet.* 2017;49:1385-1391. DOI 10.1038/ng.3913.
- Nikpay M., Goel A., Won H.H., Hall L.M., Willenborg C., Kanoni S., Saleheen D., ... McPherson R., Deloukas P., Schunkert H., Samani N.J., Farrall M. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* 2015;47:1121-1130. DOI 10.1038/ng.3396.
- Nioi P., Sigurdsson A., Thorleifsson G., Helgason H., Agustsdóttir A.B., Norddahl G.L., Helgadóttir A., ... Holm H., Gudbjartsson D., Sulem P., Thorsteinsdóttir U., Stefansson K. Variant *ASGR1* associated with a reduced risk of coronary artery disease. *N. Engl. J. Med.* 2016;374:2131-2141. DOI 10.1056/NEJMoa1508419.
- O'Donnell C.J., Cupples L.A., D'Agostino R.B., Fox C.S., Hoffmann U., Hwang S.J., Ingellson E., Liu C., Murabito J.M., Polak J.F., Wolf P.A., Demissie S. Genome-wide association study for subclinical atherosclerosis in major arterial territories in the NHLBI's Framingham Heart Study. *BMC Med. Genet.* 2007;8(Suppl.1):S4. DOI 10.1186/1471-2350-8-S1-S4.
- Pasaniuc B., Price A.L. Dissecting the genetics of complex traits using summary association statistics. *Nat. Rev. Genet.* 2017;18:117-127. DOI 10.1038/nrg.2016.142.
- Peden J.F., Farrall M. Thirty-five common variants for coronary artery disease: the fruits of much collaborative labour. *Hum. Mol. Genet.* 2011;20:R198-205. DOI 10.1093/hmg/ddr384.
- Samani N.J., Braund P.S., Erdmann J., Gotz A., Tomaszewski M., Linsel-Nitschke P., Hajat C., Mangino M., Hengstenberg C., Stark K., Ziegler A., Caulfield M., Burton P.R., Schunkert H., Tobin M.D. The novel genetic variant predisposing to coronary artery disease in the region of the *PSRC1* and *CELSR2* genes on chromosome 1 associates with serum cholesterol. *J. Mol. Med. (Berl).* 2008;86:1233-1241. DOI 10.1007/s00109-008-0387-2.
- Samani N.J., Deloukas P., Erdmann J., Hengstenberg C., Kuulasmaa K., McGinnis R., Schunkert H., Soranzo N., Thompson J., Tiret L., Ziegler A. Large scale association analysis of novel genetic loci for coronary artery disease. *Arterioscler. Thromb. Vasc. Biol.* 2009;29:774-780. DOI 10.1161/ATVBAHA.108.181388.
- Samani N.J., Erdmann J., Hall A.S., Hengstenberg C., Mangino M., Mayer B., Dixon R.J., Meitinger T., Braund P., Wichmann H.E., Barrett J.H., König I.R., Stevens S.E., Szymczak S., Tregouet D.A., Iles M.M., Pahlke F., Pollard H., Lieb W., Cambien F., Fischer M., Ouwehand W., Blankenberg S., Balmforth A.J., Baessler A., Ball S.G., Strom T.M., Braenne I., Gieger C., Deloukas P., Tobin M.D., Ziegler A., Thompson J.R., Schunkert H., WTCCC and the Cardiogenics Consortium. Genomewide association analysis of

- coronary artery disease. *N. Engl. J. Med.* 2007;357:443-453. DOI 10.1056/NEJMoa072366.
- Schunkert H., König I.R., Kathiresan S., Reilly M.P., Assimes T.L., Holm H., Preuss M., ... O'Donnell C.J., McPherson R., Erdmann J., Consortium C.A., Samani N.J. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* 2011;43:333-338. DOI 10.1038/ng.784.
- Stitzel N.O., the Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators. Variants in *ANGPTL4* and the risk of coronary artery disease. *N. Engl. J. Med.* 2016;375:2306. DOI 10.1056/NEJMc1607380.
- Svishcheva G.R. A generalized model for combining dependent SNP-level summary statistics and its extensions to statistics of other levels. *Sci. Rep.* 2019;9:5461. DOI 10.1038/s41598-019-41827-5.
- Svishcheva G.R., Belonogova N.M., Zorkoltseva I.V., Kirichenko A.V., Axenovich T.I. Gene-based association tests using GWAS summary statistics. *Bioinformatics.* 2019;35(19):3701-3708. DOI 10.1093/bioinformatics/btz172.
- The CARDIoGRAMplusC4D Consortium (Deloukas P., Kanoni S., Willenborg C., Farrall M., Assimes T.L., Thompson J.R., Ingelsson E., ... Palmer C.N., Roberts R., Watkins H., Schunkert H., Samani N.J.) Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat. Genet.* 2013;45:25-33. DOI 10.1038/ng.2480.
- Tregouet D.A., König I.R., Erdmann J., Munteanu A., Braund P.S., Hall A.S., Grosshennig A., Linsel-Nitschke P., Perret C., DeSuremain M., Meitinger T., Wright B.J., Preuss M., Balmforth A.J., Ball S.G., Meisinger C., Germain C., Evans A., Arveiler D., Luc G., Ruidavets J.B., Morrison C., van der Harst P., Schreiber S., Neureuther K., Schafer A., Bugert P., El Mokhtari N.E., Schrezenmeier J., Stark K., Rubin D., Wichmann H.E., Hengstenberg C., Ouwehand W., Ziegler A., Tiret L., Thompson J.R., Cambien F., Schunkert H., Samani N.J. Genome-wide haplotype association study identifies the *SLC22A3-LPAL2-LPA* gene cluster as a risk locus for coronary artery disease. *Nat. Genet.* 2009;41:283-285. DOI 10.1038/ng.314.
- van der Harst P., Verweij N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ. Res.* 2018;122:433-443. DOI 10.1161/CIRCRESAHA.117.312086.
- Vilne B., Schunkert H. Integrating genes affecting coronary artery disease in functional networks by multi-omics approach. *Front. Cardiovasc. Med.* 2018;5:89. DOI 10.3389/fcvm.2018.00089.
- Wang M., Huang J., Liu Y., Ma L., Potash J.B., Han S. COMBAT: A Combined Association Test for genes using summary statistics. *Genetics.* 2017;207:883-891. DOI 10.1534/genetics.117.300257.
- Webb T.R., Erdmann J., Stirrups K.E., Stitzel N.O., Masca N.G., Jansen H., Kanoni S., ... Samani N.J., Schunkert H., Deloukas P., Kathiresan S., for the Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators. Systematic evaluation of pleiotropy identifies 6 further loci associated with coronary artery disease. *J. Am. Coll. Cardiol.* 2017;69:823-836. DOI 10.1016/j.jacc.2016.11.056.
- Willer C.J., Schmidt E.M., Sengupta S., Peloso G.M., Gustafsson S., Kanoni S., Ganna A., ... Deloukas P., Kathiresan S., Mohlke K.L., Ingelsson E., Abecasis G.R. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* 2013;45:1274-1283. DOI 10.1038/ng.2797.
- Yang J., Chen S., Abecasis G., IAMDGC. Improved score statistics for meta-analysis in single-variant and gene-level association studies. *Genet. Epidemiol.* 2018;42:333-343. DOI 10.1002/gepi.22123.
- Yusuf S., Hawken S., Ounpuu S., Dans T., Avezum A., Lanas F., McQueen M., Budaj A., Pais P., Varigos J., Lisheng L., on behalf of the INTERHEART Study Investigators. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet.* 2004;364:937-952. DOI 10.1016/S0140-6736(04)17018-9.
- Zanger U.M., Schwab M. Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol. Ther.* 2013;138:103-141. DOI 10.1016/j.pharmthera.2012.12.007.
- Zheng J., Erzurumluoglu A.M., Elsworth B.L., Kemp J.P., Howe L., Haycock P.C., Hemani G., Tansey K., Laurin C., Early Genetics and Lifecourse Epidemiology (EAGLE) Eczema Consortium, Pourcain B.S., Warrington N.M., Finucane H.K., Price A.L., Bulik-Sullivan B.K., Anttila V., Paternoster L., Gaunt T.R., Evans D.M., Neale B.M. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics.* 2017;33:272-279. DOI 10.1093/bioinformatics/btw613.
- Zhou K., Donnelly L.A., Kimber C.H., Donnan P.T., Doney A.S., Leese G., Hattersley A.T., McCarthy M.I., Morris A.D., Palmer C.N., Pearson E.R. Reduced-function *SLC22A1* polymorphisms encoding organic cation transporter 1 and glycemic response to metformin: a GoDARTS study. *Diabetes.* 2009;58:1434-1439. DOI 10.2337/db08-0896.

**Acknowledgements.** This work was supported by the Russian Ministry of Education and Science (project 0324-2019-0040) and the Russian Foundation for Basic Research (project 18-04-00076).

**Conflict of interest.** The authors declare no conflict of interest.

Received April 2, 2019. Revised July 15, 2019. Accepted July 19, 2019.