

Hereditary diseases of nervous system in republic Sakha (Yakutia) / I.A. Nikolaeva, M.N. Korotov, E.E. Gurinova [et.al.] // Yakutskij medicinskij zhurnal [Yakut medical journal]. – 2009. -№2. – P.52-54.

3. Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts / B. Pakendorf, I.N. Novgorodov, V.L. Osakovskij [et.al.] // Hum Genet. – 2006. – 120. – P. 334-353.

4. Investigating the Prehistory of Tungusic Peoples of Siberia and the Amur-Ussuri Region with Complete mtDNA Genome Sequences

and Y-chromosomal Markers /A.Duggan, M. Whitten,V. Wiebe [et.al.] // PLOS one. 2013. – 8(12): e83570. doi:10.1371/ journal.pone.0083570.

5. Lifestyle mediates seasonal changes in metabolic health among the yakut (sakha) of northeastern Siberia /S.B. Levy, W.R. Leonard, L.A. Tarskia [et al.] // Am J Hum Biol. – 2016. – Nov;28(6):868-878. doi: 10.1002/ajhb.22879.

6. Population specific analysis of Yakuts exomes /A.S. Zlobin, S.Zn. Sarapov, V.P. Gur'ev [et al.] // Доклады академии наук – 2017. – Том 474, №4. – С. 505-509.

7. The complex Admixture History and Recent Southern Origins of Siberian Populations / Pugach I., Matveev R., Spitsyn V. [et.al.] // Molecular Biology and Evolution. – 2016. – 33(7): 1777-1795 doi:10.1093/molbev/msw055.

8. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans / M.Raghavan, P. Skoglund, K.E. Graf [et.al.] // Nature. – 2014. – 505. – P. 87–91 doi:10.1038/nature12736.

9. Y-chromosomal evidence for a strong reduction in male population size of Yakuts /B. Pakendorf [et. al.] // Hum.Genet – 2002. -110:198-200.

В.Г. Пшенникова, Н.А. Барашков, А.В. Соловьев, Г.П. Романов, Ф.М. Терютин, О.Л. Посух, Н.Н.Сазонов, Л.У. Джемилева, Э.К. Хуснутдинова, М.И. Томский, С.А. Федорова

АНАЛИЗ ИНФОРМАТИВНОСТИ КОМПЬЮТЕРНЫХ *IN SILICO* ПРЕДСКАЗАТЕЛЬНЫХ ПРОГРАММ ПРИ ОЦЕНКЕ КЛИНИЧЕСКОЙ ЗНАЧИМОСТИ МИССЕНС-ЗАМЕН ГЕНА *GJB2* (CX26)

УДК 575.224.22; 616.28-008

В данной работе на основе установленной клинической значимости 7 миссенс-замен гена *GJB2* (Cx26), обнаруженных в результате молекулярно-генетического исследования врожденной глухоты в Якутии, были протестированы 9 компьютерных *in silico* предсказательных программ. Для выявления программы с наиболее точными предсказательными оценками клинической значимости миссенс-замен гена *GJB2* (Cx26) проведен сравнительный анализ параметров информативности (точность, чувствительность и специфичность) с расчетом коэффициента корреляции между известным клиническим значением миссенс-замен с *in silico* оценкой, выдаваемой программами. В целом из 9 проанализированных программ наиболее точные *in silico* предсказательные оценки клинической значимости миссенс-замен гена *GJB2* (Cx26) дали две программы – SIFT и PROVEAN (R = 0,73). Полученные результаты могут помочь при проведении биоинформационного анализа, в случае обнаружения ранее не описанных в литературе миссенс-замен гена *GJB2* (Cx26).

Ключевые слова: анализ *in silico*, ген *GJB2*, коннексин 26 (Cx26), миссенс-замены, врожденная тугоухость и глухота, Якутия.

In this study, based on the established clinical significance, 7 missense variants of the *GJB2* (Cx26) gene, detected as a result of the molecular genetics study of congenital deafness in Yakutia, 9 computers *in silico* predictive programs were tested. In order to identify the program with the most accurate prediction of the clinical significance of missense variants substitutions of the *GJB2* (Cx26) gene, a comparative analysis of the informative parameters (accuracy, sensitivity and specificity) was carried out with the calculation of the correlation coefficient between the known clinical values of missense variants with *in silico* evaluation by the programs. In total, of the 9 analyzed programs, the most accurate *in silico* predictive estimates of the clinical significance of missense variants of the *GJB2* (Cx26) gene were given by two programs – SIFT and PROVEAN (R = 0.73). The obtained results can help in carrying out bioinformatic analysis, in case of detection of missense variants substitutions of the *GJB2* (Cx26) gene, which were not described before in the literature.

Keywords: *in silico* analysis, *GJB2* gene, connexin 26 (Cx26), missense variants, deafness, Yakutia.

Введение. В настоящее время в диагностике наследственной несиндромальной потери слуха высокую информативность обеспечивают методы

ЯНЦ КМП: ПШЕННИКОВА Вера Геннадиевна – н.с., psennikovavera@mail.ru, БАРАШКОВ Николай Алексеевич – к.б.н., руковод. лаб., barashkov2004@mail.ru, ТЕРЮТИН Федор Михайлович – к.м.н., с.н.с., ТОМСКИЙ Михаил Иннокентьевич – д.м.н., директор; СВФУ им. М.К. Аммосова: СОЛОВЬЕВ Айсен Васильевич – аспирант, РОМАНОВ Георгий Прокопьевич – аспирант, САЗОНОВ Николай Никитич – д.б.н., проф., ФЕДОРОВА Сардана Аркадьевна – д.б.н., зав. лаб.; ФИЦ ИЦИГ СО РАН: ПОСУХ Ольга Леонидовна – к.б.н., с.н.с.; ИБГ УНЦ РАН: ДЖЕМИЛЕВА Лиля Усеиновна – д.м.н., с.н.с., ХУСНУТДИНОВА Эльза Камилевна – д.б.н., проф., директор.

молекулярно-генетического анализа гена *GJB2*, кодирующего белок коннексин 26 (Cx26) (13q12.11, MIM 121011). У подавляющего большинства пациентов с врожденной потерей слуха идентифицируются рецессивные мутации в гене *GJB2* в гомозиготном или компаунд-гетерозиготном состоянии, что соответствует диагнозу аутосомно-рецессивной глухоты 1А типа (АРГ 1А – OMIM#220290), что делает его наиболее часто исследуемым геном у больных с врожденным нарушением слуха [29]. В настоящее время в базе данных The Human Gene Mutation Database (HGMD) в гене *GJB2* анонсировано около 390 различных нуклеотидных изменений, из которых 73% занимают однонуклеотидные (миссенс/нонсенс) замены ([\[ac.uk/ac/index.php\]\(http://www.hgmd.ac.uk/ac/index.php\)\). Патогенетическая роль большинства нонсенс-замен довольно очевидна, так как они приводят к преждевременной терминации трансляции и прерыванию синтеза белка. Миссенс-замены, в зависимости от их локализации в аминокислотной последовательности полипептидной цепи, могут быть нейтральными, или полностью или частично повреждающими структуру белка, незначительно влияя на его функцию. Вследствие этого патогенетическую роль многих миссенс-замен в развитии нарушений слуха сложно оценить.](http://www.hgmd.</p>
</div>
<div data-bbox=)

Для оценки возможного влияния аминокислотных замен на функцию и/или структуру белка в отсутствие структурно-функциональных исследований используют прогностический

метод *in silico*, который полностью осуществляется симуляционными компьютерными программами. Однако эти компьютерные программы имеют ряд недостатков, поскольку применяемый предсказательный алгоритм может кардинально отличаться. Это связано с тем, что каждая отдельно взятая программа работает с применением разных вычислительных методов (BLAST, PSI-BLAST, PSIC) и инструментов (матрица Дирихле, скрытая марковская модель, наивный байесовский классификатор). В связи с этим точность предсказательной оценки *in silico* компьютерных программ может широко варьировать [10, 19, 22, 23, 30].

В ранее проведенных молекулярно-генетических исследованиях ауто-сомно-рецессивной глухоты 1А типа в Якутии было выявлено 12 аллельных вариантов гена *GJB2* (8 патогенных и 4 доброкачественных варианта), среди которых 7 вариантов приводили к миссенс-заменам [1, 21]. Патогенность или непатогенность выявленных миссенс-замен помимо литературных источников и сведений из баз данных была подтверждена с помощью клинико-генеалогических методов исследования. Таким образом, полученные нами данные по клинической значимости выявленных миссенс-замен гена *GJB2* дают нам возможность ретроспективно оценить точность предсказаний *in silico* компьютерных программ.

Целью настоящей работы является сравнительный анализ параметров информативности наиболее популярных компьютерных *in silico* предсказательных программ для выбора наиболее точной программы для предсказания клинической значимости миссенс-замен гена *GJB2* (Cx26).

Материалы и методы исследования

Миссенс варианты гена *GJB2*. Для тестирования *in silico* программ были использованы 7 миссенс-замен гена *GJB2* (Cx26): с.79G>A (p.Val27Ile), с.101T>C (p.Met34Thr), с.109G>A (p.Val37Ile), с.269T>C (p.Leu90Pro), с.341A>G (p.Glu114Gly), с.368C>A (p.Thr123Asn) и с.457G>A (p.Val153Ile), обнаруженных при молекулярно-генетическом исследовании врожденной глухоты в Якутии (рис. 1) [1, 21]. Все данные миссенс *GJB2*-варианты присутствовали в базах данных аннотированных вариантов (фенотипические: OMIM – <http://www.omim.org/>, Human Gene Mutation Databases – <http://www.hgmd.cf.ac.uk/ac/index.php>, ClinVar – <http://www.ncbi.nlm.nih.gov/clinvar/>; популяционные: the Exome Aggregation

Consortium – <http://exac.broadinstitute.org>, 1000 Genomes Project – <http://browser.1000genomes.org/index>, dbSNP – <http://www.ncbi.nlm.nih.gov/snphtml>). Далее все представленные варианты были распределены по пяти классификациям, разработанным ACMG (Американская коллегия медицинской генетики и геномики), по рекомендациям которой используется специальный стандарт терминов для обозначения клинической значимости вариантов, выявленных в генах: «pathogenic» – патогенный, «likely pathogenic» – вероятно патогенный, «uncertain significance» – неопределенного значения, «likely benign» – вероятно доброкачественный, «benign» – доброкачественный [22]. Так, в группу патогенных вариантов, ассоциированных с нарушениями слуха (глухота/APГ 1А), вошли три варианта: с.269T>C (p.Leu90Pro) – как патогенный, «pathogenic», с.101T>C (p.Met34Thr) и с.109G>A (p.Val37Ile) – как вероятно патогенные, «likely pathogenic». Остальные 4 варианта были интерпретированы нами как доброкачественные варианты, не имеющие клинического значения: с.79G>A (p.Val27Ile) и с.457G>A (p.Val153Ile) – как доброкачественные, «benign», с.341A>G (p.Glu114Gly) – как доброкачественный/вероятно доброкачественный, «benign/likely benign», с.368C>A (p.Thr123Asn) – вероятно доброкачественный, «likely benign». Для оценки клинической значимости представленных 7 миссенс-замен гена *GJB2* мы опирались не только на курируемые базы данных, но и на результаты, полученные среди исследованных 580 чел. из Якутии (393 – с врожденным нарушением слуха и 187 – без нарушения слуха).

Компьютерные *in silico* программы. Для данного исследования выбраны компьютерные *in silico* предсказательные программы: SIFT (Sorting Intolerant From Tolerant), FATHMM (Functional Analysis through Hidden Markov Models), MutationAssessor, PolyPhen2 (Polymorphism Phenotyping v-2), Condel (Consensus Deleteriousness), MutationTaster, MutPred (Mutation Prediction), Align GVGD (Align Grantham Variation/Grantham Deviation) и PROVEAN (Protein Variation Effect Analyzer). Каждый *in silico* инструмент прогностических программ для классификации вариантов использует различные параметры, полная информация о которых доступна в онлайн интернет ресурсе (веб-сайты программ указаны в табл.1).

Для прогностического запроса в поисковых окнах *in silico* программ использованы идентификаторы генетических последовательностей (нуклеотидных, аминокислотных и белковых) в формате FASTA (GenBank: ANB08964.1 – gap junction beta-2 protein [Homo sapiens]) из базы данных NCBI – Reference Sequence (<http://www.ncbi.nlm.nih.gov/protein>), и в формате Ensembl ID (ENSG00000165474 – для гена, ENSP00000372299 – для белка, ENST00000382848 – для транскрипта) из базы The Human Protein Atlas (<http://www.proteinatlas.org>).

В исследовании использовались материалы из банка данных ЯНЦ КМП.

Анализ информативности компьютерных *in silico* программ

При получении прогностических результатов компьютерных *in silico* программ были вычислены их аналитические параметры по следующим операционным характеристикам [4, 7, 23]:

чувствительность (англ. Sensitivity, Se) – часть истинно положительных результатов (правильная идентификация патогенных вариантов), по формуле $Se = TP / (TP + FN)$, где TP – истинно положительные случаи и FN – ложноотрицательные случаи;

специфичность (англ. Specificity, Sp) – часть истинно отрицательных результатов (правильная идентификация доброкачественных вариантов, не имеющих клинического значения), по формуле $Sp = TN / (TN + FP)$, где TN – истинно отрицательные случаи и FP – ложноположительные случаи;

точность (англ. Accuracy, A) – отношение полных правильных предсказаний к общему числу предсказаний, по формуле $A = (TP + TN) / (TP + TN + FP + FN)$;

прогностичность положительно го результата (ППР) (англ. positive predictive values, PPV) – пропорция истинно положительных результатов среди всех положительных предсказаний, по формуле $PPV = TP / (TP + FP)$;

прогностичность отрицательно го результата (ПОР) (англ. negative predictive values, NPV) – пропорция истинно отрицательных результатов среди всех отрицательных предсказаний, по формуле $NPV = TN / (TN + FN)$.

Для определения наличия взаимосвязи между анализируемыми параметрами (клиническое значение миссенс-замены с предсказательной оценкой программ) нами был проведен расчет коэффициента корреляции с помощью функции «КОРРЕЛ» в среде Microsoft Excel, согласно рекомендациям [4, 7, 13, 23].

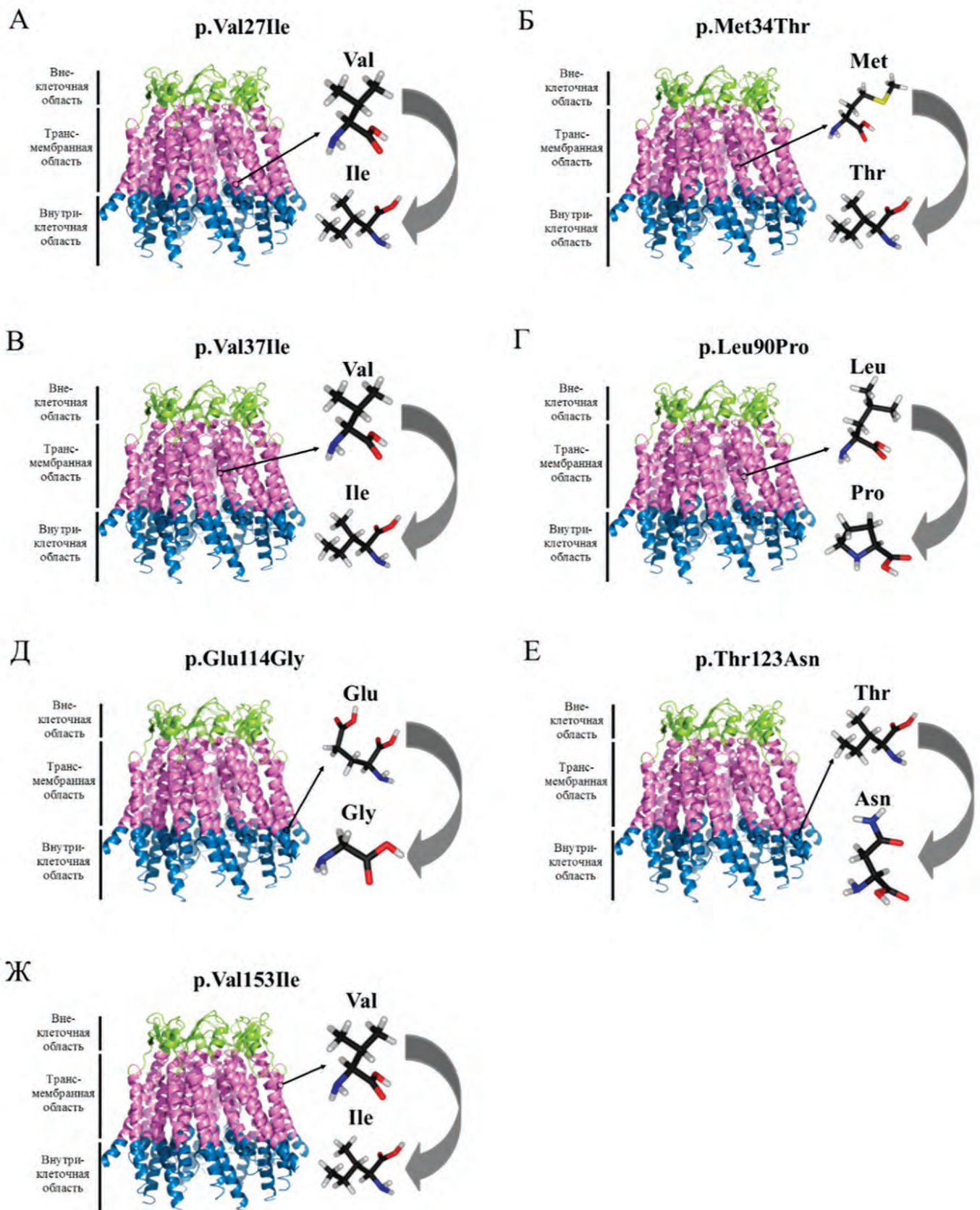


Рис.1. Локализация выявленных несинонимичных (миссенс) аминокислотных замен (А-Е) в последовательности белка-коннексина 26. Примечание. 3D-визуализация структурной модели белка Cx26 получена с помощью программы PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2>)

Компьютерные *in silico* предсказательные программы

Название и вебсайт	Основа			Классификация варианта			Ссылки
	Алгоритм	Метод	Вычислительный инструмент	Эффект	Диапазон оценок (score)	Прогноз	
SIFT http://sift.jcvi.org	Эволюционная консервативность	Компиляция набора данных функционально связанных белковых последовательностей с помощью BLAST/PSI-BLAST	Матрица Дирихле		0.00 – 1	<0.05 = «Damaging» (повреждающий); >0.05 = «Tolerated» (толерантный)	[12, 16-19]
FATHMM http://fathmm.biocompute.org.uk/							
MutationAssessor http://mutationassessor.org/	Эволюционная консервативность и структура/функция белка	Предоставляет данные из других баз данных, таких как COSMIC, UniProt и Pfam, а также свой собственный «функциональный балл воздействия» на мутацию	Метод перекрестной энтропии (Cross-Entropy Method)	Воздействие аминокислотной замены на структуру/функцию белка	-5.76 – 5.76	<0.8 = «neutral» (нейтральный); 0.8 ≤ 1.9 = «low» (возможно нейтральный); 1.9 ≤ 3.5 = «medium» (возможно паготенный); >3.5 = «high (паготенный)	[27, 28]
PolyPhen2 http://genetics.bwh.harvard.edu/pph2/							
Condel http://bg.upf.edu/famnsdb/	Эволюционная консервативность и структура/функция белка	Статистический метод взвешивания и профилирования последовательностей из подмножеств одинаковых последовательностей в нескольких выравниваниях с помощью PSIC	Наивный байсовский классификатор		Две модели: HumDiv: 0.00 – 1 HumVar: 0.00 – 1	0.0 – 0.15 = «benign» (доброкачественный); 0.15 – 1.0 = «possibly damaging» (возможно повреждающий); 0.85 – 1.0 = «damaging» (повреждающий)	[2]
MutationTaster http://www.mutationtaster.org/							
MutPred http://mutpred.mutdb.org/	Эволюционная консервативность и структура/функция белка	Объединяет SIFT, PolyPhen2, MutationAssessor и FATHMM	Наивный байсовский классификатор	Причина заболевания	0.0 – 1	0.0 = «Neutral» (нейтральный); 1.0 = «Deleterious» (повреждающий)	[11]
Align GVGD http://agvgd.hci.utah.edu/agvgd_input.php							
PROVEAN http://provean.jcvi.org/index.php	Выравнивание и измерение сходства между последовательностью варианта и последовательностью гомологичного белка	Интеграция информации из различных биомедицинских баз данных (Ensembl, UniProt, ClinVar, ExAC, 1000 Genomes Project, phyloP, phastCons)	Метод опорных векторов (Support Vector Machines – SVM)		0.0 – 215 (не влияет на прогноз)	«disease causing» (является); «disease causing automatic» (является автоматически); «polymorphism» (полиморфный); «polymorphism automatic» (полиморфный, автоматически)	[14, 15]
MutPred http://mutpred.mutdb.org/							
PROVEAN http://provean.jcvi.org/index.php	Выравнивание и измерение сходства между последовательностью варианта и последовательностью гомологичного белка	Основывается на установленном методе SIFT	Матрица Грантэма GV (Grantham Variation) GD (Grantham Deviation)	Воздействие аминокислотной замены на структуру/функцию белка	g = 0.00 – 1 (g – общий балл) (p – оценка 5 свойств)	g > 0.5, p < 0.05 = «actionable hypotheses» (вероятная гипотеза) g > 0.75, p < 0.05 = «confident hypotheses» (истинная гипотеза); g > 0.75, p < 0.01 = «very confident hypotheses» (точная гипотеза)	[3]
MutPred http://mutpred.mutdb.org/							
PROVEAN http://provean.jcvi.org/index.php	Выравнивание и измерение сходства между последовательностью варианта и последовательностью гомологичного белка	Измерение биохимических расстояний между аминокислотами (норма/замена), в соответствии с MSA	Матрица Грантэма GV (Grantham Variation) GD (Grantham Deviation)	Патогенетическое влияние на структуру/функцию белка	GVGD = Класс: C0, C15, C25, C35, C45, C55, C65	C65 – наиболее вероятное; C0 – наименее вероятное	[8, 9]
MutPred http://mutpred.mutdb.org/							
PROVEAN http://provean.jcvi.org/index.php	Выравнивание и измерение сходства между последовательностью варианта и последовательностью гомологичного белка	Компиляция набора данных функционально связанных белковых последовательностей с помощью BLAST, с дальнейшей обработкой больших баз данных CD-HIT (ver.4.5.5)	Эмпирическая матрица замены (BLOSUM62)	Функциональное воздействие на белок	-40 – 12.5 (порог: -2.5)	≥ -2.5 = «deleterious» (повреждающий); ≤ -2.5 = «neutral» (нейтральный)	[6, 24]
MutPred http://mutpred.mutdb.org/							

Результаты. В результате полученных предсказаний *in silico* программ все миссенс-замены гена *GJB2* (3 патогенных, 4 доброкачественных) были оценены неоднозначно, то есть имели более чем одно предсказание (одновременно несколько классификаций). Тем не менее один вариант – с.269T>C (p.Leu90Pro), всеми программами был оценен как вариант с повреждающим влиянием (табл. 2).

Точность предсказаний клинической значимости миссенс-замен среди анализируемых программ варьировала от 42,9%, которую показала FATHMM, до 85,5%, достигнутой двумя программами SIFT и PROVEAN (табл. 3). Данные программы показали также высокие параметры чувствительности и специфичности – 66,7% и 100% соответственно. Программы MutationAssessor, MutationTaster, CONDEL и FATHMM имели 100%-ную чувствительность, но при этом показали низкую специфичность – 50%, а CONDEL – ее полное отсутствие. Наиболее высокий показатель ППР (100%) и ПОР (80%) имели программы SIFT и PROVEAN. Программа MutPred имела 100% ППР, но показала низкое значение ПОР – 57,2%.

Для определения точности предсказаний компьютерных *in silico* программ и для того, чтобы максимизировать общее качество их аналитических параметров, нами был рассчитан коэффициент корреляции R (рис. 2). Наиболее высокая взаимосвязь *in silico* предсказаний с клиническим значением миссенс-замен гена *GJB2* выявлена у программ SIFT и PROVEAN (R = 0,73), что соответствует их аналитическим параметрам, полученным в ходе проведенного анализа параметров информативности (табл.3). Со средней корреляцией выявлены MutationAssessor, MutationTaster и CONDEL (R = 0,54) (рис. 2), что также соответствует их аналитическим параметрам (табл. 3). Слабую корреляцию показали MutPred (R = 0,47), PolyPhen2 (R = 0,41), очень слабую Align GVGD (R = 0,16), а программа FATHMM показала нулевое значение, что свидетельствует об отсутствии связи между наблюдаемыми значениями.

Обсуждение

На основе клинической значимости 7 вариантов (миссенс-замен) гена *GJB2*, обнаруженных в результате молекулярно-генетического исследования врожденной глухоты в Якутии, были протестированы 9 компьютерных программ, наиболее часто используемых для биоин-

Таблица 2

Оценка миссенс-замен гена *GJB2* (Cx26) с помощью компьютерных *in silico* предсказательных программ

Миссенс-замена <i>GJB2</i> (Cx26)	Клиническая значимость	SIFT	MutationAssessor	FATHMM	Polyphen-2	MutationTaster	PROVEAN	Align GVGD	MutPred*	CONDEL
c.79G>A p.Val27Ile rs2274084	Доброкачественный	Tolerated score: 0.21	Medium FI score: 2.28 VC score: 2.16 VS score: 2.40	Damaging score: -5.59	Probably damaging HumDiv score: 1.000 HumVar score: 0.998	Polymorphism score: 29	Neutral score: -0.660	Unclassified Class C25 GV 0.00 GD 29.61	Probability of deleterious mutation: - general score: 0.321	Deleterious Calculated Condel score: 0.612278613903
c.101T>C p.Met34Thr rs35887622	Патогенный	Damaging score: 0.01	Medium FI score: 2.315 VC score: 2.43 VS score: 2.20	Damaging score: -5.41	Benign HumDiv score: 0.038 HumVar score: 0.083	Disease causing score: 81	Deleterious score: -3.801	Class C65 GV 0.00 GD 81.04	Probability of deleterious mutation: - general score: 0.969	Deleterious Calculated Condel score: 0.58786807751
c.109G>A p.Val37Ile rs72474224	Патогенный	Tolerated score: 0.34	Medium FI score: 2.095 VC score: 2.58 VS score: 1.61	Damaging score: -5.46	Probably damaging HumDiv score: 1.000 HumVar score: 0.996	Disease causing score: 29	Neutral score: -0.857	Unclassified Class C25 GV 0.00 GD 29.61	Probability of deleterious mutation: - general score: 0.902	Deleterious Calculated Condel score: 0.61487213316
c.269T>C p.Leu90Pro rs80338945	Патогенный	Damaging score: 0	Medium FI score: 3.33 VC score: 4.26 VS score: 2.40	Damaging score: -5.64	Probably damaging HumDiv score: 1.000 HumVar score: 0.996	Disease causing score: 98	Deleterious score: -6.482	Deleterious C65 GV 0.00 GD 97.78	Confident Hypotheses Gain of sheet (P = 0.039) general score: 0.915	Deleterious Calculated Condel score: 0.676708483818
341A>G p.Glu114Gly rs2274083	Доброкачественный	Tolerated score: 0.16	Medium FI score: 2.005 VC score: 2.40 VS score: 1.61	Damaging score: -4.58	Benign HumDiv score: 0.001 HumVar score: 0.001	Polymorphism score: 98	Neutral score: -0.123	Deleterious Class C65 GV 0.00 GD 97.85	Probability of deleterious mutation: - general score: 0.232	Deleterious Calculated Condel score: 0.556433693212
c.368C>A p.Thr123Asn rs111033188	Доброкачественный	Tolerated score: 0.59	Neutral FI score: -0.305 VC score: -0.61 VS score: -0	Damaging score: -4.42	Benign HumDiv score: 0.000 HumVar score: 0.000	Disease causing score: 53	Neutral score: 0.797	Deleterious Class C55 GV 0.00 GD 64.77	Probability of deleterious mutation: - general score: 0.201	Neutral Calculated Condel score: 0.513276654484
.457G>A p.Val153Ile rs111033186	Доброкачественный	Tolerated score: 1	Neutral FI score: -0.305 VC score: -0.43 VS score: -0.18	Damaging score: -3.69	Benign HumDiv score: 0.003 HumVar score: 0.007	Disease causing score: 29	Neutral score: 0.138	Unclassified Class C25 GV 0.00	Probability of deleterious mutation: - general score: 0.488	Neutral Calculated Condel score: 0.491937780564

Примечание. Серым цветом выделены «истинно» положительные и «истинно» отрицательные результаты.

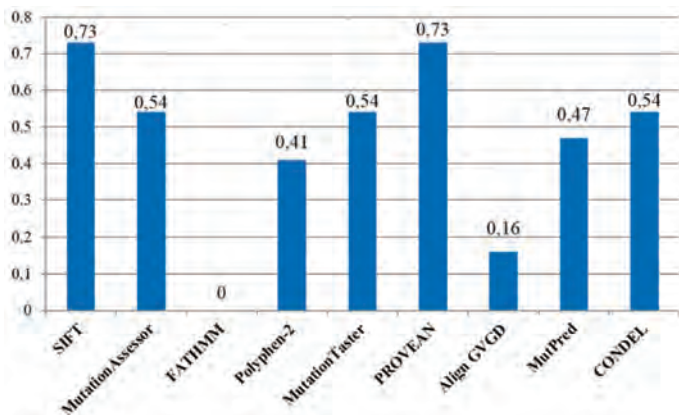


Рис.2. Гистограмма коэффициента корреляции (R).

Примечание. R – взаимосвязь между известным клиническим значением миссенс-замен гена *GJB2* с *in silico* оценкой, выдаваемой 9 компьютерными предсказательными программами.

формационного анализа. После полученных *in silico* предсказательных оценок был проведен сравнительный анализ параметров информативности (точность, чувствительность и специфичность) программ, а также рассчитан коэффициент корреляции полученных предсказаний программ с клинической значимостью миссенс-замен гена *GJB2*.

В результате анализа всех полученных параметров анализируемых *in silico* программ и расчета коэффициента корреляции между значениями (клиническое значение с *in silico* предсказаниями) с наилучшими показателями выявлены программы SIFT и PROVEAN. У данных программ 85,8% точных предсказаний, что в данном исследовании можно считать лучшим показателем, так как в литературных данных, где приводятся аналогичные сравнения параметров, наиболее высокая точность предсказаний достигала 80 – 90% [4, 7, 14, 23, 25, 30]. Программы SIFT и PROVEAN показали также высокие параметры чувствительности (66,7%) и специфичности (100%), что в контексте данного исследования имеет высокую ценность, поскольку чем выше параметры чувствительности, тем эффективнее предсказываются патогенетические варианты, и чем выше специфичность, тем эффективнее предсказываются доброкачественные варианты. Средние показатели точности были у программы PolyPhen2 (71,5%), все остальные ее показатели были не выше 66,7%. Неудовлетворительные результаты показали две программы FATHMM и MutPred, выдавшие большое количество неправильных предсказаний (рис. 2, табл. 3).

В 2015 г. было опубликовано ис-

следование А. Yilmaz [31], в котором отражены результаты биоинформационного анализа 211 миссенс-мутаций гена *GJB2*, анонсированных в базах данных Ensembl и HGMD, полученных с помощью четырех предсказательных компьютерных *in silico* программ: SIFT, PANTHER, PolyPhen-2 и FATHMM [31]. Авторы пришли к выводу, что полученные результаты предсказаний *in silico* программ продемонстрировали возможность их применения для биоинформационного анализа при прогнозировании эффектов миссенс-мутаций, ассоциированных с нарушениями слуха. Однако в данном исследовании сравнительный анализ их параметров информативности (точность, чувствительность, специфичность) программ не был проведен. Литературные данные, в которых представлены результаты сравнительного анализа параметров компьютерных *in silico* программ, свидетельствуют, что из существующего на сегодняшний день множества программ не все одинаково подходят для предсказания патогенности миссенс-мутаций в конкретных генах, ответственных за конкретное наследственное заболевание. Так, например, при тестировании 1118 вариантов четырех генов *BRCA1*, *BRCA2*, *MLH1*, и *MLH2*, ассоциированных с онкологическими заболеваниями, наиболее точной оказалась программа Align GVGD [4]. В результате биоинформационного анализа вариантов генов *RYR1* и *CACNA1S*, ассоциированных со злокачественной гипертермией, из 8 *in silico* программ четыре: MutPred, SNPs & GO, PhD-SNP и CADD смогли точно классифицировать варианты гена *RYR1*, но неоднозначно оценили варианты другого гена – *CACNA1S* [20]. В результате анализа эффективности

11 *in silico* программ, на основе 14191 патогенных мутаций и 22001 нейтральных вариантов из базы данных UniProt с наиболее точными предсказаниями были выявлены две программы – FATHMM и KGGSeq [7]. Программы MAPP и PolyPhen2 наиболее точно предсказали эффекты 74 миссенс-замен генов *MMR* (*MLH1*, *MSH2*, *MSH6* и *PMS2*), которые связаны с синдромом Линча (наследственная форма рака толстой кишки) [5]. Используя набор данных, состоящий из 122 достоверно патогенных и доброкачественных вариантов генов, связанных с группой заболеваний Ras-патий (*RAS-pathies*) и конечностно-поясных мышечных дистрофий (*Limb Girdle Muscular Dystrophies – LGMD*), были проанализированы 17 *in silico* предсказательных программ, и наиболее точной оказалась программа MutPred [30].

Таким образом, из 9 проанализированных программ наиболее точные *in silico* предсказательные оценки клинической значимости миссенс-замен гена *GJB2* дали две программы – SIFT и PROVEAN ($R = 0,73$). Полученные результаты могут помочь при проведении биоинформационного анализа в случае обнаружения ранее не описанных в литературе миссенс-замен гена *GJB2*, ассоциированных с нарушениями слуха.

Литература

1. Анализ спектра и частоты *GJB2*-мутаций у пациентов с врожденными нарушениями слуха в Республике Саха (Якутия) / В.Г. Пшенникова, Н.А. Барашков, Ф.М. Терютин [и др.] // Мед. генетика. – 2015. – Т. 14. – №6 (156). – С.10-23.
2. *GJB2* mutation spectrum in patients with congenital hearing loss in Yakutia / V.G. Pshennikova, N.A. Barashkov, F.M. Teryutin [et al.] // Russ. J. Med. Genetics. – 2015. – Vol. 6 (156). – P. 10-23.

Таблица 3

Аналитические параметры информативности компьютерных *in silico* предсказательных программ, %

Предсказательные <i>in silico</i> программы	Точность	Чувствительность	Специфичность	ППР	ПОР
SIFT	85,8	66,7	100	100	80
MutationAssessor	71,5	100	50	60	100
FATHMM	42,9	100	0	42,9	0
Polyphen-2	71,5	66,7	66,7	66,7	66,7
MutationTaster	71,5	100	50	60	50
PROVEAN	85,8	66,7	100	100	80
Align GVGD	57,2	66,7	50	50	66,7
MutPred	71,5	33,4	100	100	57,2
CONDEL	71,5	100	50	60	100

Примечание. ППР – прогностичность положительных результатов (прогноз патогенетических вариантов); ПОР – прогностичность отрицательных результатов (прогноз доброкачественных вариантов, не имеющих клинического значения).

2. A method and server for predicting damaging missense mutations / I.A. Adzhubei, S. Schmidt, L. Peshkin [et al.] // *Nat. Methods*. – 2010. – Vol. 7(4). – P. 248-249.
3. Automated inference of molecular mechanisms of disease from amino acid substitutions / B. Li, V.G. Krishnan, M.E. Mort. // *Bioinformatics*. – 2009. – Vol. 25(21). – P. 2744-50. doi: 10.1093/bioinformatics/btp528
4. Assessment of *in silico* protein sequence analysis in the clinical classification of variants in cancer risk genes / I.D. Kerr, H.C. Cox, K. Moyes [et al.] // *J Community Genet*. – 2017. – Vol. 8(2). – P. 87-95. doi: 10.1007/s12687-016-0289-x
5. Calibration of multiple *in silico* tools for predicting pathogenicity of mismatch repair gene missense substitutions / B.A. Thompson, M.S. Greenblatt, M.P. Vallee [et al.] // *Hum Mutat*. – 2013. – Vol. 34(1). – P. 255-65. doi: 10.1002/humu.22214
6. Choi Y. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels / Y. Choi, A.P. Chan // *Bioinformatics*. – 2015. – Vol. 31(16). – P. 2745-2747.
7. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies / C. Dong, P. Wei, X. Jian [et al.] // *Hum Mol Genet*. – 2015. – Vol. 24(8). – P. 2125-21237. doi: 10.1093/hmg/ddu733
8. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral / S.V. Tavtigian, A.M. Deffenbaugh, L. Yin [et al.] // *J Med Genet*. – 2006. – Vol. 43(4). – P. 295-305.
9. Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods / E. Mathe, M. Olivier, S. Kato [et al.] // *Nucleic Acids Res*. – 2006. – Vol. 34(5). – P. 1317-1325.
10. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs / X. Liu, C. Wu, C. Li [et al.] // *Hum Mutat*. – 2016. – Vol. 37(3). – P. 235-241. doi: 10.1002/humu.22932
11. González-Pérez A. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel / A. González-Pérez, N. López-Bigas // *Am J Hum Genet*. – 2011. – Vol. 88(4). – P. 440-449. doi: 10.1016/j.ajhg.2011.03.004
12. Kumar P. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm / P. Kumar, S. Henikoff, P.C. Ng // *Nat Protoc*. – 2009. – Vol. 4(7). – P. 1073-1081.
13. Matthews B.W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme / B.W. Matthews // *J Biochim Biophys Acta*. – 1975. – Vol. 405. – P. 442-451.
14. MutationTaster evaluates disease-causing potential of sequence alterations / J.M. Schwarz, C. Rödelberger, M. Schuelke [et al.] // *Nat Methods*. – 2010. – Vol. 7(8). – P. 575-576.
15. MutationTaster2: mutation prediction for the deep-sequencing age / J.M. Schwarz, D.N. Cooper, M. Schuelke, D. Seelow // *Nat Methods*. – 2014. – Vol. 11(4). – P. 361-362. doi: 10.1038/nmeth.2890
16. Ng P.C. Predicting Deleterious Amino Acid Substitutions / P.C. Ng, S. Henikoff // *Genome Res*. – 2001. – Vol. 11(5). – P. 863-874.
17. Ng P.C. Accounting for Human Polymorphisms Predicted to Affect Protein Function / P.C. Ng, S. Henikoff // *Genome Res*. – 2002. – Vol. 12(3). – P. 436-446.
18. Ng P.C. SIFT: predicting amino acid changes that affect protein function / P.C. Ng, S. Henikoff // *Nucleic Acids Res*. – 2003. – Vol. 31(13). – P. 3812-3814.
19. Ng P.C. Predicting the Effects of Amino Acid Substitutions on Protein Function / P.C. Ng, S. Henikoff // *Annu Rev Genomics Hum Genet*. – 2006. – Vol. 7. – P. 61-80.
20. Schiemann A.H. Comparison of pathogenicity prediction tools on missense variants in RYR1 and CACNA1S associated with malignant hyperthermia / A.H. Schiemann, K.M. Stowell // *Br J Anaesth*. – 2016. – Vol. 117(1). – P. 124-128. doi: 10.1093/bja/aew065.
21. Spectrum and Frequency of the *GJB2* Gene Pathogenic Variants in a Large Cohort of Patients with Hearing Impairment Living in a Subarctic Region of Russia (the Sakha Republic) / N.A. Barashkov, V.G. Pshennikova, O.L. Posukh [et al.] // *PLoS One*. – 2016. – Vol. 11(5):e0156300. doi: 10.1371/journal.pone.0156300.
22. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology / S. Richards, N. Aziz, S. Bale [et al]; ACMG Laboratory Quality Assurance Committee // *Genet Med*. – 2015. – Vol. 17(5). – P. 405-24. doi: 10.1038/gim.2015.30.
23. Thusberg J. Performance of mutation pathogenicity prediction methods on missense variants / J. Thusberg, A. Olatubosun, M. Vihinen // *Hum Mutat*. – 2011. – Vol. 32(4). – P. 358-368. doi: 10.1002/humu.21445.
24. Predicting the functional effect of amino acid substitutions and indels / Y. Choi, G.E. Sims, S. Murphy [et al.] // *PLoS One*. – 2012. – Vol. 7(10):e46688. doi: 10.1371/journal.pone.0046688.
25. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models / H.A. Shihab, J. Gough, D.N. Cooper [et al.] // *Hum Mutat*. – 2013. – Vol. 34(1). – P. 57-65. doi: 10.1002/humu.22225.
26. Ranking Non-Synonymous Single Nucleotide Polymorphisms based on Disease Concepts / H.A. Shihab, J. Gough, M. Mort [et al.] // *Human Genomics*. – 2014. – Vol. 8. – P. 11. doi: 10.1186/1479-7364-8-11.
27. Reva B. Determinants of protein function revealed by combinatorial entropy optimization / B. Reva, Y. Antipin, C. Sander // *Genome Biol*. – 2007. – Vol. 8(11). – P. 232.
28. Reva B. Predicting the functional impact of protein mutations: application to cancer genomics / B. Reva, Y. Antipin, C. Sander // *Nucleic Acids Res*. – 2011. – Vol. 39(17). – P. 118. doi: 10.1093/nar/gkr407.
29. Van Camp G. Hereditary Hearing Loss Homepage: URL: <http://hereditaryhearingloss.org>, 2016 / G. Van Camp, R.J.H. Smith.
30. Variability in pathogenicity prediction programs: impact on clinical diagnostics / L.C. Walters-Sen, S. Hashimoto, D.L. Thrush [et al.] // *Mol Genet Genomic Med*. – 2015. – Vol. 3(2). – P. 99-110. doi: 10.1002/mgg3.116.
31. Yilmaz A. Bioinformatic Analysis of *GJB2* Gene Missense Mutations / A. Yilmaz // *Cell Biochem Biophys*. – 2015. – Vol. 71(3). – P. 1623-1642. doi: 10.1007/s12013-014-0385-7.

A.Н. Романова, М.И. Воевода, В.Н. Максимов

ИССЛЕДОВАНИЕ АССОЦИАЦИИ ЧЕТЫРЕХ ОДНОНУКЛЕОТИДНЫХ ПОЛИМОРФИЗМОВ С АРТЕРИАЛЬНОЙ ГИПЕРТЕНЗИЕЙ И ИНФАРКТОМ МИОКАРДА В РЕСПУБЛИКЕ САХА (ЯКУТИЯ): ЭТНИЧЕСКИЕ И ГЕНДЕРНЫЕ ОСОБЕННОСТИ

УДК 616.12-008.1(571.56)

РОМАНОВА Анна Николаевна – д.м.н., гл.н.с.-руков. отд. ЯНЦ КМП, ranik@mail.ru; **ВОЕВОДА Михаил Иванович** – акад. РАН, д.м.н., проф. НИИ терапии и профилактической медицины, mvoevoda@ya.ru; **МАКСИМОВ Владимир Николаевич** – д.м.н., доцент, зав. лаб. НИИ терапии и профилактической медицины, medik11@mail.ru.

Проведено исследование ассоциации 4 однонуклеотидных полиморфизмов с артериальной гипертензией и инфарктом миокарда у жителей Республики Саха (Якутия) в зависимости от этнической и гендерной принадлежности.

Ключевые слова: однонуклеотидные полиморфизмы, артериальная гипертензия, инфаркт миокарда, гендерные и этнические особенности, Республика Саха (Якутия).

The research of 4 single nucleotide polymorphisms association with arterial hypertension and myocardial infarction in population of the Republic of Sakha (Yakutia) depending on ethnicity and gender is carried out.

Keywords: single nucleotide polymorphisms, arterial hypertension, myocardial infarction, gender and ethnic features, Republic of Sakha (Yakutia).