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.

The authors:
1. Vladimir L. Osakovsky, Institute of Health, North-Eastern Federal University named after M.K. Ammosov, Yakutsk, Russia, is_labgene@mail.ru
2. Address: Institute of Health of the North-Eastern Federal University named after M.K. Ammosov, Kulakovsky street, build. 46, 104 room, Yakutsk, Russia, 677000.
3. Tatyana M. Sivtseva
4. Institute of Health, North-Eastern Federal University named after M.K. Ammosov, Yakutsk, Russia, tm.sivtseva@s-vfu.ru, +7 914 2237432
5. Address: Institute of Health of the North-Eastern Federal University named after M.K. Ammosov, Kulakovsky street, build. 46, 104 room, Yakutsk, Russia, 677000.

V.G. Pshennikova, N.A. Barashkov, A.V. Solovyev, G.P. Romanov, F.M. Teryutin, N.N. Sazonov, O.L. Posukh, L.U. Dzhemileva, E.K. Khusnutdinova, M.I. Tomsky, S.A. Fedorova

INFORMATIVE PARAMETERS OF PREDICTIVE IN *SILICO* COMPUTER PROGRAMS IN ASSESSING CLINICAL SIGNIFICANCE OF MISSENSE VARIANTS OF GJB2 *(CX26)* GENE

ABSTRACT

Currently, in the HGMD database (The Human Gene Mutation Database) in the *GJB2* gene, coding protein connexin 26 (Cx26), 390 different nucleotide changes have been announced, most of which are associated with deafness, of which 73% are single nucleotide (missense/nonsense) variants. The pathogenetic role of most nonsense substitutions is obvious, as they lead to premature termination of translation and interruption of protein synthesis. It is more difficult to assess the mechanism of action of the missens replacement on protein, since they may have a damaging/partially damaging or neutral effect, depending on the location in the amino acid sequence of the polypeptide chain. To assess the possible effect of amino acid substitutions on the structure and/or function of the protein, in the absence of structural and functional studies, the *in silico* prognostic method is used, which is completely performed by simulation computer programs. In this study, based on the established clinical significance, 7 missense variants of the *GJB2* gene, detected as a result of the molecular genetics study of congenital deafness in Yakutia, 9 computer-*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* 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* gene were given by two programs - SIFT and PROVEAN (R = 0,73). The results obtained can help in carrying out bioinformatic analysis, in the case of detection of missense variants substitutions of the *GJB2* gene, which were not described before in the literature.

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

INTRODUCTION

Currently, methods of molecular and genetic analysis of the *GJB2* gene (Cx26) (13q12.11, MIM 121011) provide a high information content for the diagnosis of hereditary nonsyndromic hearing loss. The vast majority of patients with congenital hearing loss have recessive mutations in the *GJB2* gene in the homozygous or compound heterozygous state, which corresponds to a diagnosis of autosomal recessive deafness type 1A (DFNB1A, MIM 220290) [29]. According to the database of The Human Gene Mutation Database (HGMD), about 390 different nucleotide changes have been announced in the *GJB2* gene, of which 73% occupy single nucleotide missense/nonsense variants (http://www.hgmd.cf.ac.uk/ac/index.php). The pathogenic role of the majority of nonsense variants is obvious, as they lead to premature termination of translation and interruption of protein synthesis. Missense variants, depending on their location in the amino acid sequence of the polypeptide chain, can be neutral, damaging or partially damaging the structure of the protein, slightly affecting its function. In consequence, the pathogenetic role of many missense variants in the development of hearing impairment is difficult to assess.

To assess the possible effect of amino acid substitutions on the function and/or structure of the protein, in the absence of structural and functional studies, use the method *in silico* is used, which is fully implemented by simulation computer programs. However, existing *in silico* computer programs have a number of drawbacks, since the applied predictive algorithm can be radically different. This is due to the fact that each individual program works with different computational methods (BLAST, PSI-BLAST, PSIC) and tools (*Matrix Dirichlet*, Hidden Markov Model, Naive Bayesian classifier). In this regard, the accuracy of the predictive evaluation of *in silico* computer programs can vary widely [10, 19, 22, 23, 30].

In the previously conducted molecular genetic studies of DFNB1A in Yakutia, 12 allelic variants of the *GJB2* gene (8 pathogenic and 4 benign variants) were identified, among which 7 missense

variants led to amino acid substitution [1, 21]. The pathogenicity or non-pathogenicity of identified missense variants has been confirmed through clinical genealogical research methods, in addition to literature sources and data from databases. . Thus, the confirmed data on clinical significance of the identified *GJB2* gene missense variants , enable us to estimate the accuracy of *in silico* computer programs predictions.

The purpose of this paper is to compare the informative parameters of the most popular computer *in silico* predictive programs, to select the most accurate program for predicting the clinical significance of missense variants of the *GJB2* (Cx26) gene.

MATERIALS AND METHODS

Missense variants of the GJB2 (Cx26) gene

For *in silico* programs testing, seven missense variants of the *GJB2* (Cx26) gene were used: c.79G>A (p.Val27Ile), c.101T>C (p.Met34Thr), c.109G>A (p.Val37Ile), c.269T>C (p.Leu90Pro), c.341A>G (p.Glu114Gly), c.368C>A (p.Thr123Asn) *и* c.457G>A (p.Val153Ile), which were identified during previous molecular genetic studies of congenital deafness in Yakutia (Figure 1) [1, 21]. All of these missense variants of the *GJB2* gene were annotated in the following databases: OMIM (http://www.omim.org/); Human Gene Mutation Database (http://www.hgmd.cf.ac.uk/ac/index.php); ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/); the Exome Aggregation Consortium (http://exac.broadinstitute.org); 1000 Genomes Project (http://browser.1000genomes.org/index); dbSNP (http://www.ncbi.nlm.nih.gov/snphtml).

Clinical significance of listed *GJB2* sequence variants was interpreted following the American College of Medical Genetics and Genomics (ACMG) guidance [22]. This report recommends the use of specific standard terminology: «pathogenic», «probably pathogenic», «uncertain significance», «probably benign» and «benign» to describe variants identified in Mendelian disorders [22]. In the group of pathogenic variants associated with hearing impairment (deafness/DFNB1A), three variants were classified as c.269T>C (p.Leu90Pro) - as a «pathogenic», c.101T>C (p.Met34Thr) and c.109G>A (p.Val37Ile) - as «likely pathogenic». The remaining 4 variants were classified as benign variants: c.79G>A (p.Val27Ile) and c.457G>A (p.Val153Ile) - as «benign», c.341A>G (p.Glu114Gly) - «benign/likely benign», c.368C>A (p.Thr123Asn) - as «likely benign». To assess the clinical significance of the 7 missense variants of the *GJB2* gene, we relied not only on the supervised databases, but also on the results obtained among the 580 people studied from Yakutia (393 patients with congenital hearing impairment and 187 normal hearing individuals).

In silico computer programs

Table 1 shows the *in silico* prediction programs selected for this study: 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) and PROVEAN (Protein Variation Effect Analyzer). Each *in silico* classification tool utilizes different parameters for variant classification, the full details of which can be found online (websites are showed in table 1).

For the query in the search windows of *in silico* programs, the genetic sequence identifiers (nucleotide, amino acid and protein) were used in FASTA format (GenBank: AHB08964.1 - gap junction beta-2 protein [Homo sapiens]) from the NCBI database - Reference Sequence (http://www.ncbi.nlm.nih.gov/protein) and in the Ensembl ID format (ENSG00000165474 - for the gene, ENSP00000372299 - for the protein, ENST00000382848 - for the transcript) from the database of The Human Protein Atlas (http://www.proteinatlas.org).

Analysis of the informative content of in silico computer programs

When obtaining predictive results of *in silico* computer programs, their analytical parameters were calculated as follows [4, 7, 23]:

Sensitivity (Se) - *part of the truly positive results (correct identification of pathogenic variants), according to formula Se=TP/TP+FN*, where *TP* - are true positive cases and *FN* - are false negative cases;

Specificity (Sp) - part of the truly negative results (the correct identification of benign variants that do not have clinical significance), according to the formula *Sp=TN/TN+FP*, where *TN* - are true negative cases and *FP* - false positive cases;

Accuracy (A) - the ratio of complete correct predictions to the total number of predictions, according to formula *A=TP+TN/TP+TN+FP+FN*;

Positive predictive values (PPV) - the proportion of positive results that were true positives, according to formula *PPV=TP/TP+FP*;

Negative predictive values (NPV) - proportion of negative results that were true negatives, according to formula *NPV=TN/TN+FN*.

To determine the relationship between the analyzed parameters (the clinical value of missense variants with predictive evaluation of programs), we calculated the correlation coefficient (in Microsoft Excel) [4, 7, 13, 23].

RESULTS

As a result of the *in silico* programs predictions, all missense variants of the *GJB2* gene (3 pathogenic, 4 benign) were assessed ambiguously, that is, they had more than one prediction (simultaneous different interpretation). Nevertheless, one variant c.269T>C (p.Leu90Pro), was evaluated by all programs as a damaging variant. The predictive estimate of missense variants of the *GJB2* gene issued by computer *in silico* programs in comparison with their established clinical significance are presented in table 2.

In table 3, informative parameters of the compared prediction programs *in silico* are presented. The accuracy of the predictions of the clinical significance of missense replacements among the analyzed programs ranged from 42.9% (FATHMM), to 85.5% (achieved by the two programs SIFT and PROVEAN). These programs also showed high sensitivity and specificity parameters - 66.7% and 100%, respectively. The programs MutationAssessor, MutationTaster, CONDEL and FATHMM had 100% sensitivity, but showed a low specificity of 50%, and CONDEL total absence. The SIFT and PROVEAN programs had the highest predictive value of the positive result (100%) and the predictivity of the negative result (80%) . The MutPred had 100% PPV, but showed a low NPV value of 57.2%.

To determine the accuracy of the predictions of computer *in silico* programs and in order to maximize the overall quality of their analytical parameters, we calculated the correlation coefficient (R). Figure 2, shows the general correlation coefficients, where the highest correlation *in silico* predictions with the clinical value of missense variants of the *GJB2* gene was detected in the SIFT and PROVEAN programs (R = 0,73), which corresponds to their analytical parameters obtained

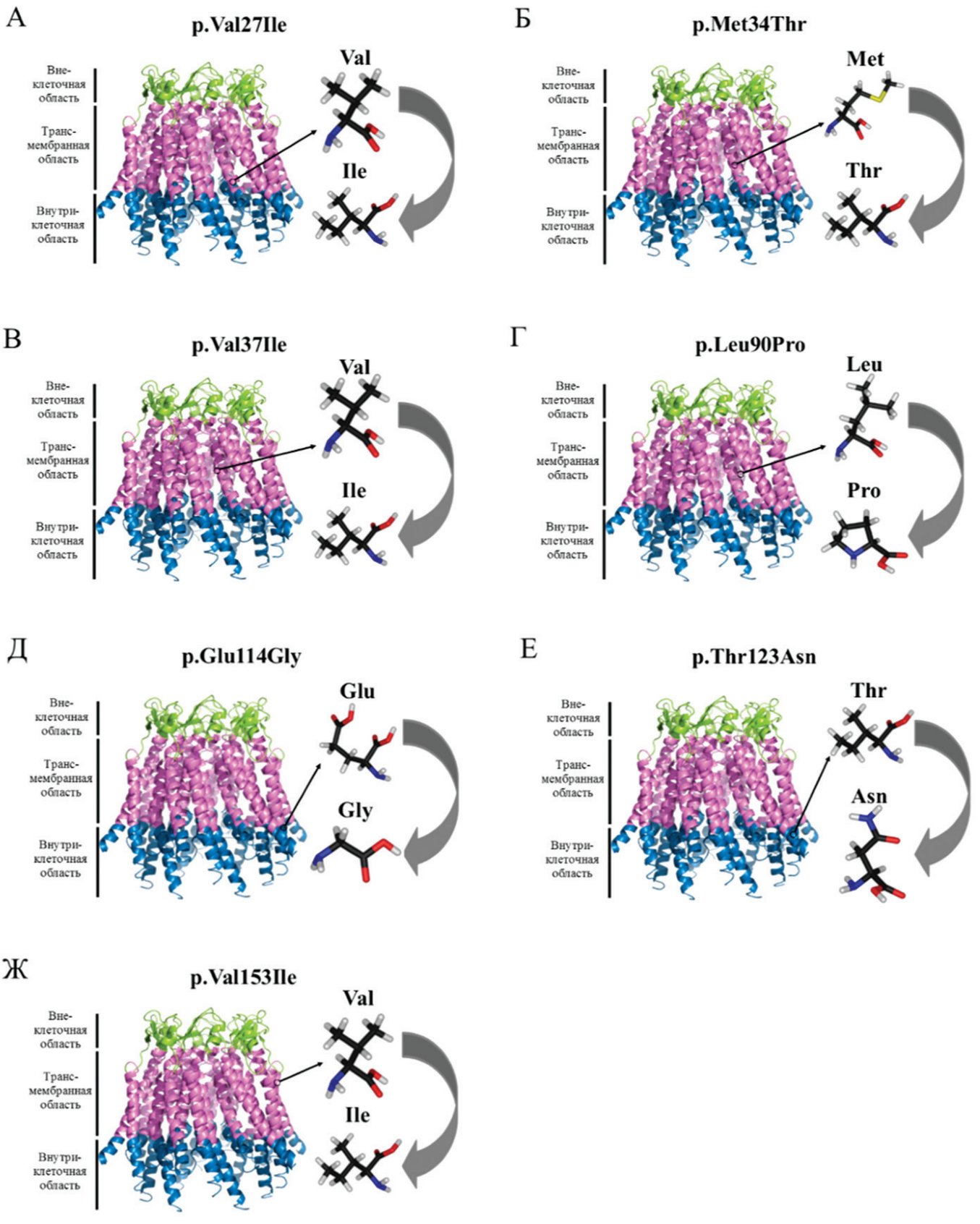

Figure 1. Localization of the revealed nonsynonymous (missense) amino acid substitutions in the protein-connexin sequence 26.
Note: 3D visualization of the structural model of the Cx26 protein was obtained using the PolyPhen-2 program (http://genetics.bwh.harvard.edu/pph2).

Computer *in silico* predictive programs

Name and web-site	The basis			Classification			References
	Algorithm	Method	Computing tools	Effect	Score	Prediction	
SIFT http://sift.jcvi.org	Evolutionary conservation	Compilation of a data set of functionally linked protein sequences using BLAST/PSI-BLAST	Matrix Dirichlet	The effect of the amino acid substitution on the structure/function of the protein	0.00 - 1	<0.05 = «Damaging»; >0.05 = «Tolerated»	[12, 16-19]
FATHMM http://fathmm.biocompute.org.uk/			Hidden Markov Model (HMM)		> -1.5 <	<-1.5 = «Damaging»; >-1.5 = «Tolerated»	[25, 26]
MutationAssessor http://mutationassessor.org/		It provides data from other databases, such as COSMIC, UniProt and the Pfam, as well as its own «functional point of influence» on the mutation	Cross-Entropy Method		-5.76 – 5.76	≤0.8 = «neutral»; 0.8≤1.9 = «low»; 1.9≤3.5 = «medium»; >3.5 = «high»	[27, 28]
PolyPhen2 http://genetics.bwh.harvard.edu/pph2/	Protein structure/function and evolutionary conservation	A statistical method of weighting and profiling sequences from subsets of identical sequences in several alignments using PSIC	Naive Bayesian classifier		Two models: HumDiv: 0.00 - 1 HumVar: 0.00 - 1	0.0 - 0.15 = «benign»; 0.15 - 1.0 = «possibly damaging»; 0.85 - 1.0 = «damaging»	[2]
Condel http://bg.upf.edu/fannsdb/		Combines SIFT, PolyPhen-2, MutationAssessor and FATHMM			0.00 - 1	0.0 = «Neutral»; 1.0 = «Deleterious»	[11]
MutationTaster http://www.mutationtaster.org/	Protein structure/function and evolutionary conservation	Integration of information from various biomedical databases (Ensembl, UniProt, ClinVar, ExAC, 1000 Genomes Project, phyloP, phastCons)	Naive Bayesian classifier	Cause of the disease	0,0 – 215 (Does not affect the forecast)	«disease causing»; «disease causing automatic»; «polymorphism»; «polymorphism automatic»	[14, 15]
MutPred http://mutpred.mutdb.org/	Protein structure/function and evolutionary conservation	Based on the established SIFT method	Support Vector Machines (SVM)	The effect of the amino acid substitution on the structure/function of the protein	g = 0.00 - 1 (g - Total score) (p - Rating 5 properties)	g>0.5, p<0.05 = «actionable hypotheses»; g>0.75, p<0.05 = «confident hypotheses»; g>0.75, p<0.01 = «very confident hypotheses»	[3]
Align GVGD http://agvgd.hci.utah.edu/agvgd_input.php	Protein structure/function and evolutionary conservation	Measurement of biochemical distances between amino acids (norm/substitution), according to MSA	Matrix of Grantham Grantham Variation (GV) Grantham Deviation (GD)	Pathogenetic effect on protein structure/ unction of the protein	GVGD = Class: C0, C15, C25, C35, C45, C55, C65	C65 - The most probable; C0 - Least probable	[8, 9]
PROVEAN http://provean.jcvi.org/index.php	Alignment and measurement of similarity between variant sequence and protein sequence homology	Compilation of a data set of functionally linked protein sequences using BLASTP, with further processing of large databases CD-HIT (ver.4.5.5)	Blocks Substitution Matrix (BLOSUM62)	Functional effect on protein	- 40 – 12.5 (threshold: -2.5)	≥- 2.5 = «deleterious»; ≤- 2.5 = «neutral»	[6, 24]

during the analysis of the informative parameters (Table 3). With an average correlation, MutationAssessor, MutationTaster and CONDEL (R = 0,54) were identified (Figure 2), which also corresponds to their analytical parameters (Table 3). A weak correlation was shown by MutPred (R = 0,47), PolyPhen2 (R = 0,41), very weak Align GVGD (R = 0,16), and the FATHMM program showed a zero value, which indicates a lack of correlation between the observed values (Figure 2).

DISCUSSION

Based on the known clinical significance of 7 missense variants of the *GJB2* gene, detected as a result of the molecular genetic study of congenital deafness in Yakutia, 9 most often used for bioinformatics analysis computer programs were tested. After the *in silico* predictions were obtained, a comparative analysis of the informative parameters (accuracy, sensitivity and specificity) of the programs was carried out, and the correlation coefficient (R) of the predicted programs with the clinical significance of the missense variants of the *GJB2* gene was calculated.

As a result, of the analysis of all the obtained parameters, of the analyzed *in silico* programs and the calculation of the correlation coefficient between the values (clinical significance with *in silico* predictions), the SIFT and PROVEAN programs were identified with the best indicators. In these programs, 85.8% of predictions are accurate which in frame of this study can be considered as the best indicator, since in the literature, where similar parameter comparisons are given, the highest accuracy of predictions reached 80% - 90% [4, 7, 14, 23, 25, 30]. The SIFT and PROVEAN programs also showed high sensitivity (66.7%) and specificity (100%) parameters, which in the context of this study is of high value, as the higher sensitivity parameters, the more effectively the pathogenic variants are predicted, and the higher the specificity, the more effectively the benign options. The middle accuracy was in the PolyPhen2 program (71.5%), and all other indicators - 66.7%. Bad results were shown by two programs FATHMM and MutPred, which produced a large number of incorrect predictions (Figure 2, Table 3).

In 2015, A. Yilmaz [31] published a study of the results of bioinformatics analysis of 211 missense mutations of the *GJB2* gene, announced in the

Table 2

Evaluation of missense variants of the *GJB2* (Cx26) gene by computer *in silico* predictive programs

Missense variant *GJB2* (Cx26)	Clinical significance	SIFT	MutationAssessor	FATHMM	Polyphen-2	MutationTaster	PROVEAN	Align GVGD	MutPred*	CONDEL
c.79G>A p.Val27Ile rs2274084	Benign	**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	Pathogenic	**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	**Deleterious** 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	Pathogenic	**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	Pathogenic	**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	Probability of deleterious mutation: **Confident Hypotheses** Gain of sheet (P = 0.039) general score:0.915	**Deleterious** Calculated Condel score: 0.676708483818
c.341A>G p.Glu114Gly rs2274083	Benign	**Tolerated** score: 0.16	**Medium** FI score:2.005 VC score:2.40 VS score:161	**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	Benign	**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
c.457G>A p.Val153Ile rs111033186	Benign	**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 GD 29.61	Probability of deleterious mutation: - general score:0.488	**Neutral** Calculated Condel score: 0.491937780564

Note: Gray is highlighted with «truly» positive and «truly» negative results.

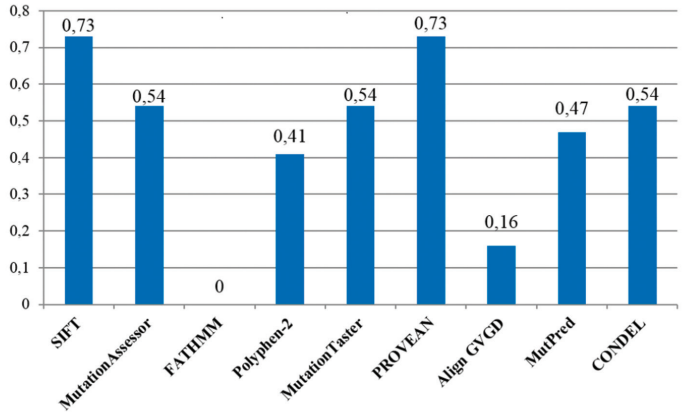

Figure 2. Histogram of the correlation coefficient (R).
Note: R - is the relationship between the known clinical value of missense variants of the GJB2 gene with in silico evaluation given by 9 computer predictive programs.

Table 3

Analytical informative parameters of computer in silico predictive programs

In silico predictive programs	Accuracy	Sensitivity	Specificity	PPV	NPV
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%

Note: PPV - prognosticity of positive results (predictive of pathogenetic variants); NPV - prognosticity of negative results (predictive of benign variants).

Ensembl and HGMD databases, using four predictive computer *in silico* programs: SIFT, PANTHER, PolyPhen-2 and FATHMM [31]. Results of the study demonstrated the applicability of bioinformatics algorithms in predictions of the hearing impairment causing mutations effect. However, in this study, a comparative analysis of informative parameters (accuracy, sensitivity, specificity) of the programs was not carried out. The literature data, which presents the results of a comparative analysis of the parameters of computer *in silico* programs, show that of the existing *in silico* programs, not all are equally suitable for predicting the pathogenicity of missense mutations in specific genes responsible for a specific hereditary disease. For example, in testing 1118 variants of four *BRCA1*, BRCA2, *MLH1* and *MLH2* genes associated with cancer, the most accurate was Align GVGD [4]. As a result, of the bioinformatics analysis of the variants of the *RYR1* and *CACNA1S* genes, associated with malignant hyperthermia, four out of 8 *in silico* programs: MutPred, SNPs & GO, PhD-SNP and CADD were able to accurately classify the variants of the *RYR1* gene, but ambiguously evaluated variants of another *CACNA1S* gene [20]. In another study, which included 14 191 deleterious Mendelian disease causing mutations, and 22 001 neutral mutations, which were annotated as not known to be associated with any phenotypes, all based on Uniprot annotation found, that FATHMM and KGGSeq had the highest accuracy [7]. MAPP was the most accurate tool and MAPP + PolyPhen-2.1 provided the best-combined model of 74 missense substitutions of *MMR* genes (*MLH1*, *MSH2*, *MSH6* and *PMS2*) that confer colon cancer susceptibility in Lynch syndrome [5]. Using a dataset consisting of 122 credibly pathogenic and benign variants in genes associated with the RASopathy disorders and limb-girdle muscular dystrophy (LGMD), 17 *in silico* predictive programs were analyzed, and result was that MutPred was the most accurate, whith a weighted accuracy of 82.6% in the full dataset. [30].

Thus, out of 9 analyzed programs, two programs - SIFT and PROVEAN (R = 0,73) - gave the most accurate *in silico* predictive estimates of the clinical significance of missense variants of the *GJB2* gene. The obtained results can help in the of bioinformatics analysis, in the case of detection of previously not described missense variants of the *GJB2* gene associated with hearing impairment.

REFERENCES

1. *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.

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 eightrecurrent 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. Rödelsperger, 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.N. Romanova, M.I. Voevoda, V.N. Maksimov

ASSOCIATION OF FOUR SINGLE NUCLEOTIDE POLYMORPHISMS WITH ARTERIAL HYPERTENSION AND MYOCARDIAL INFARCTION IN THE RS (YA): ETHNIC AND GENDER FEATURES

ABSTRACT

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

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

In Yakutia the cardiovascular diseases (CVD) in the structure of causes, both morbidity (19.1%) and mortality (47.4%), occupy a leading position [1]. Arterial hypertension (AH) is one of the main risk factors for the development of CVD and their complications, such as myocardial infarction (MI) and stroke. Currently in Russia, about 40% of the population (more than 42 million people) suffers from AH [2].

In recent years, a trend has been formed such as genetic cardiology, which integrates the concepts and technologies of molecular genetics for understanding the etiology and pathogenesis of CVD clinical polymorphism. The genetic approach allows creating a base for con-