ISSN 0233–7657. Biopolymers and Cell. 2013. Vol. 29. N 6. P. 506–510 doi: 10.7124/bc.000841

UDC 577.11 + 577.21

EPHA1 gene SNPs analysis in population of Ukraine

R. V. Gulkovskyi[1, 2], S. Y. Chernushyn[1], S. A. Kravchenko[1], G. M. Bychkova[1], L. A. Livshits[1]

[1]Institute of Molecular Biology and Genetics, NAS of Ukraine
150, Akademika Zabolotnoho Str., Kyiv, Ukraine, 03680

[2]Taras Shevchenko National University of Kyiv, Educational and Scientific Centre «Institute of Biology»
64, Volodymyrska Str., Kyiv, Ukraine, 03601

livshits@imbg.org.ua

Aim. *Analysis the EPHA1 gene G1475A and G1891A alleles distribution in the population of Ukraine, and to study the protein secondary structure as the first step in the investigation of EPHA1 gene involvement in intellectual disability pathogenesis.* ***Methods***. *Observation group consisted of 300 individuals, including 164 (54.6 %) male and 136 (45.3 %) female individuals. Polymorphic variants were detected using PCR followed by Kpn1 RFLP analysis for G1475A and ARMS PCR analysis for G1891A.* ***Results***. *The data concerning EPHA1 genotypes and allelic variants distribution were obtained. The low frequency of 1475A allele (0,012) and the absence of 1891A allele (not previously described) were revealed in the population of Ukrainian.* ***Conclusions***. *Assays for the detection of EPHA1 gene G1475A and G1891A SNPs based on ARMS and restriction analysis were developed and the preliminary data on distribution of G1475A and G1891A polymorphisms in the population of Ukraine were obtained.*

Keywords: EPHA1 gene, polymorphism, population.

Introduction. In the frame of CHERISH project (no. 223692) concerning the intellectual disability (ID) genetic basis, the next generation exome sequencing was conducted for a Ukrainian family (UKR 094) with two affected children (proband and his younger sibling) and their healthy non-consanguineous parents. The proband (UKR 263) is a 12 year old boy with ID (IQ = 43) and hyperactivity. His brother (UKR 264) is a 4 year old boy with febrile convulsions, ID and multiple congenital anomalies (Fig. 1). Both affected probands underwent physical and neurological examination, standard G-banding karyotype analysis and biochemical tests: amino-acids and acylcarnitines (TANDEM MS) as well as molecular genetic screening of common ID syndromes (FRAXA, FRAXE, FRAXF loci together with 15q11-q13 locus). Standard karyotype, biochemical and molecular genetic tests as well as array-CGH analysis (400 K resolution) were normal in both patients.

Exome analysis revealed several variants in either homozygous or compound heterozygous state in five genes. Among these genes, we decided to concentrate on the *EPHA1* gene, where two non-synonymous substitutions were detected in both patients: G1475A (rs11768549) inherited from the mother and G1891A (has not been previously described) inherited from the father. The both non-synonymous SNPs have been identified in the coding region (exons 7 and 11) of the *EPHA1* receptor gene. *EPHA1* has been reported to be highly expressed in the neural tissues during embryogenesis and it seemed a good candidate for autosomal recessive inherited ID [1, 2].

EPHA1 is the founding member of the erythropoietin-producing hepatocellular (Eph) receptors family, but little is known about its function. Eph kinases constitute the largest family of receptor tyrosine kinases, with 16 distinct members. The receptors are termed EPHA (EPHA1–8, only including those found in mammals) or EPHB (EPHB1–4, EPHB6) on the basis of se-

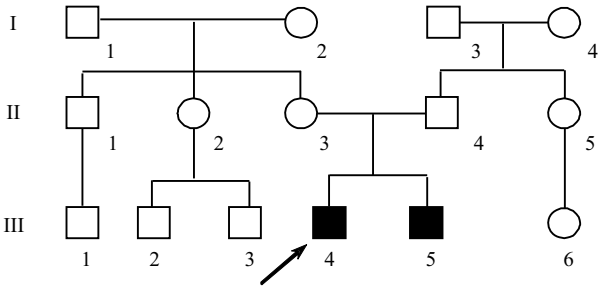

Fig. 1. Family UKR 094 pedigree: *III-4* – UKR 263; *III-5* – UKR 264; *II-3* – UKR 265; *II-4* – UKR 266

quence homologies and depending on the subgroup of ligands that they bind [3]. Ligand for the *EPHA* receptor is called ephrin-A. It has been recently demonstrated that EPHA/ephrin-A signalling has a proapoptotic effect on the embryonic cortex without affecting the proliferation or cell cycle progression [4]. Ephrin and Eph receptor genes are implicated in the control of cell and axon guidance in neural systems of different species [1, 2, 5, 6].

EPHA1 gene is located at chromosome 7q34 and contains 18 exons, two more than the related tyrosine kinase. The human *EPHA1* gene is transcribed as a single 3.5 kb mRNA, and encodes a polypeptide of 984 amino acids with a molecular mass of 130 kDa [7]. It is important to know that *EPHA1* is characterized by high conservative level among many species, indicating that there may be an evolutionarily conserved function [7]. It was shown that this gene is highly expressed in liver, lung, and kidney, but not in the adult brain. However, it is highly expressed in the spinal neural tube during neurulation [8]. The role of *EPHA1* during tissue morphogenesis is highlighted by the expression studies performed on mouse embryos, where the pattern of expression is dynamic throughout the immediate postimplantation, gastrulation, neurulation, and early somitogenesis periods [9].

In this study the analysis of single nucleotide polymorphisms G1475A and G1891A in *EPHA1* gene was performed in population of Ukraine and it is the first step in the analysis of *EPHA1* gene involvement in intellectual disability pathogenesis. To understand a biochemical effect of these substitutions we tried to predict possible changes in the EPHA1 protein structure.

Materials and methods. The DNA-samples were extracted from peripheral blood leucocytes of unrelated volunteer donors from different regions of Ukraine by standard methods [10]. Informed consents were obtained from all individuals participating in our study. The observation group consisted of 300 individuals, including 164 (54.6 %) male and 136 (45.3 %) female individuals. This group may be considered representative for the estimation of DNA polymorphism frequency in autosomal genes [11, 12].

Table 1
Specific oligonucleotides, designed and synthesized in accordance to corresponding exon sequences of EPHA1 gene, were used as primers

Substitution	Nucleotide sequence	Amplicon size, bp
G1475A	AAAGGGCCAGGACCCAGTGGGGTC AGC – forward	194
	CATGTGCTCTGATGCTGTCC – reverse	–
G1891A	CCTCTTTGAACCATTGCGTT – common	204
	GCTGATGGTGGACACTGTCATAG – wild type	–
	GCTGATGGTGGACACTGTCATAA – polymorphic	–

DNA sequencing was performed by the dideoxynucleotide chain-termination method [13] using [^{35}S]-dATP or [^{35}S]-dCTP (ICN), the Sequenase version 2.0 DNA sequencing kit (USB) and Thermo Sequenase cycle sequencing kit («Amersham», Great Britain).

An analysis of the G1891A polymorphism was performed by ARMS (amplification refractory mutation system) PCR. Specific oligonucleotide primers were complementary either to the wild type or to the sequence with substitution (Table 1). The PCR amplification was performed in a final volume of 15 μl containing 1× PCR buffer, 1.5 mM $MgCl_2$, 200 μM of each dNTP, 0.8 μM of each primer, 0.2 units of Taq-DNA polymerase and 200 ng of the DNA template. The cycling conditions for G1891A variant were as follows: initial denaturation at 94 °C for 5 min, 30 cycles consisting of denaturation at 94 °C for 30 s, annealing at 63 °C for 30 s, extension at 72 °C for 30 s, and a final elongation step at 72 °C for 5 min.

The presence of G1475A EPHA1 polymorphism was examined by PCR-RFLP (restriction fragment length polymorphism) analysis using the specific oligonucleotide primers described in the table. The PCR amplifica-

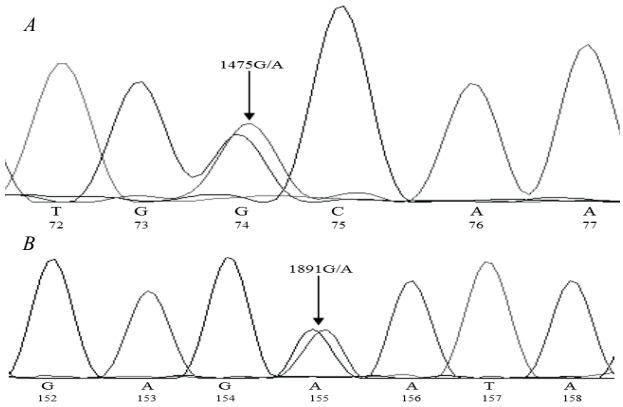

Fig. 2. Representative sequence chromatogram of the *EPHA1* gene exons 7 (*A*) and 11 (*B*) PCR products: *A* – UKR 265 (mother) – heterozygous carrier for the G1475A polymorphism; *B* – UKR 266 (father) – heterozygous carrier for the G1891A polymorphism

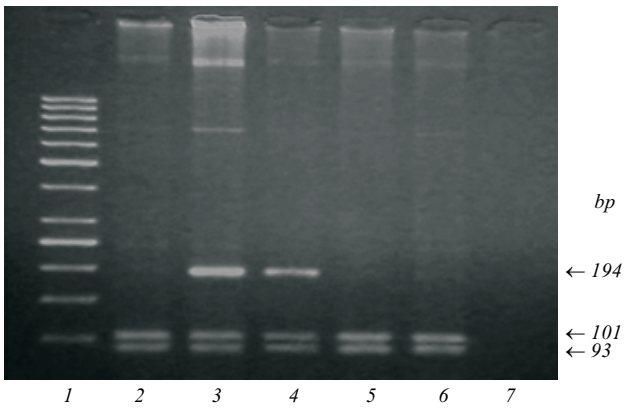

Fig. 3. RFLP analysis of G1475A *EPHA1* receptor gene variant (4 % agarose gel electrophoresis): *1* – molecular weight marker (Ladder 50 bp); *2*, *5*, *6* – individuals with homozygous genotype G1475G; *3*, *4* – individual with heterozygous genotype G1475A; *7* – negative control

tion was performed in a final volume of 25 μl containing 1 × PCR buffer, 1.5 mM $MgCl_2$, 200 μM of each dNTP, 1 μM of each primer, 0.2 units of Taq-DNA polymerase and 200 ng of the DNA template. The cycling conditions for G1475A variant were as follows: initial denaturation at 94 °C for 5 min, 30 cycles consisting of denaturation at 94 °C for 40 s, annealing at 55 °C for 40 s, extension at 72 °C for 40 s, and a final elongation step at 72 °C for 3 min. The amplified fragments were digested with *KpnI*. Digestion was performed in 15 μl reaction volume containing 1 X reaction buffer, 0.5 units of the restriction enzyme and 10 μl of purified PCR product, incubated at 37 °C overnight and analyzed by 6 % standard agarose gel electrophoresis.

Primers were designed using the web-based PRIMER 3.0 program (http://workbench.sdsc.edu). We used the «BLAST» program at http://www.ncbi.nlm.nih.gov/blast to check the primers specificity. Hypothetical RFLP results were tested using NEBcutter V2.0 (http://tools.neb.com/NEBcutter2). We also used the PSIPRED Protein Sequence Analysis Workbench (http://bioinf.cs.ucl.ac.uk/psipred) to predict the protein secondary structure.

Results and discussion. The Sanger sequencing of *EPHA1* gene loci with substitutions was performed. Sequencing confirmed the results of the exome analysis of UKR 094 family members and revealed that the father is a heterozygous carrier for the C1891T polymorphism and the mother is a heterozygous carrier for the C1475T polymorphism (Fig. 2). The sample that had previously undergone Sanger sequencing was used as a positive control.

RFLP analysis of the G1475A *EPHA1* variants. The designed primers successfully amplified the exon 7 fragment (194 bp) of *EPHA1* gene. The G to A transition in G1475A variant creates a restriction site for endonuclease *KpnI*. Thereby three different patterns could be observed for G1475A variant after the restriction digestion: a 194 bp band (for 1475A/A); a 194 bp, a 101 bp and a 93 bp bands (for 1475G/A); a 101 bp and a 93 bp bands (1475G/G) (Fig. 3).

ARMS for G1891A analysis is based on the observation that PCR amplification is inefficient or completely refractory if there is a mismatch between the 3' terminal nucleotide of a PCR primer and a corresponding template. This approach implies two PCR reactions. Amplification of the allele 1891G is accomplished using a complementary primer «wild type» (Table 1). Conversely, only the 1891A allele will be amplified if the 3' residue is complementary to the sequence with substitution (Table 1). Thus, as a result a normal (G1891G) individual generates PCR product only in the «normal» reaction; a heterozygote gives products in both reactions, and a homozygous A1891A individual gives amplification only in the «mutant» reaction. The results of analysis of individuals with different genotypes are presented in Fig. 4.

Based on the RFLP analysis of G1475A variant, individuals were classified into three groups: GG (1475G/G), GA (1475G/A) and AA (1475A/A). Of the 300 analyzed samples from the present study, we have found 7 heterozygous carriers for the G1475A polymorphism only. Genotypes and allele frequencies of the G1475A

Table 2
Genotype and allele frequencies distribution of EPHA1 gene variants in different populations

Populations	Chromosome sample count	Homozygote GG (wild type)	Heterozygote GA	Homozygote AA (polymorphic allele)	Polymorphic allele frequency
Ukraine	600	0.977	0.023	0.000	0.012
European	226	0.965	0.035	0.000	0.018
Asian	174	1.000	0.000	0.000	0.000
African	116	1.000	0.000	0.000	0.000

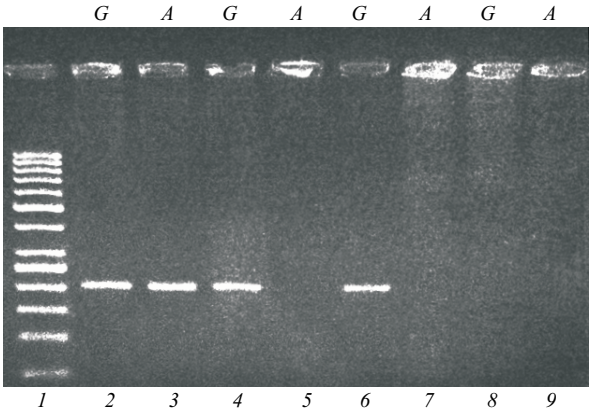

Fig. 4. ARMS analysis of G1891A *EPHA1* receptor gene variant (2 % agarose gel electrophoresis): *1* – molecular weight marker (50 bp ladder); *2*, *3* – individual with heterozygous genotype G1891A; *4–7* – individuals with homozygous genotype G1891G; *8*, *9* – negative control

polymorphism are presented in Table 2. The observed genotype distributions showed no deviations from Hardy-Weinberg expectations.

The frequency of allelic variants of *EPHA1* gene in our study is in agreement with the published data obtained within the International HapMap Project (http://hapmap.ncbi.nlm.nih.gov/) (Table 2).

We did not identify G1891A substitution in any of 300 investigated volunteer donors. It was found only in the members of aforesaid Ukrainian family with ID in children. Previous reports have demonstrated a significant association of T3022C (rs11767557) polymorphism within this gene with Alzheimer's disease [14, 15]. *EPHA1* has been reported to be highly expressed in the neural tissues during embryogenesis [3, 4].

The first analyzed SNP – G1475A – exchanges codon 492 from arginine (CGG) to a glutamine (CAG). The second one – G1891A – leads to an amino acid change at position 631 from glycine (GGA) to arginine (AGA). To understand a biochemical effect of these substitutions we tried to predict possible changes in the EPHA1 protein structure. A comparative analysis of the predicted secondary structure revealed that both amino acid substitutions lead to significant changes in the probability of the β-sheet forming. The change of Arg492 to Glu results in extending the 11th and 12th β-sheets in the fibronectin type III repeat of EPHA1 protein by 1 amino acid. The change of Gly631 to Arg results in extending of the 2nd β-sheet in the tyrosine kinase domain by 1 amino acid. Thereby these substitutions may directly cause changes in protein solubility or binding with ligands or protein-partners.

In the frame of CHERISH FP7 project, the exome NGS was performed for two affected probands from Ukrainian family with moderate ID, which revealed two polymorphisms in the *EPHA1* gene: G1475A and novel SNP G1891A. The assays for detection of these SNPs based on ARMS and restriction analysis were developed and the preliminary studies on the distribution of G1475A and G1891A polymorphisms in the population of Ukraine were conducted. Further case-control *EPHA1* exons sequencing will be carried out to evaluate the involvement of the *EPHA1* mutations in ID development.

Acknowledgements. We thank M. Noukas, M. Sauk, L. Milani, T. Pippucci, F. Balombo for their kind assistance in next generation exome sequencing of the Ukr 094 family members. We thank V. Kitam for his kind assistance in analysis of the proteins secondary structure.

Р. В. Гулковський, С. Ю. Чернушин, С. А. Кравченко, Г. М. Бичкова, Л. А. Лівшиць

Аналіз SNPs гена *EPHA1* у популяції України

Резюме

Мета. *Аналіз розподілу алельних варіантів поліморфізмів G1475A і G1891A гена EPHA1 у популяції України та вивчення вторинної структури білка як перший етап визначення ролі гена EPHA1 у патогенезі інтелектуальної недіздатності.* ***Методи***. *Алельні варі-*

анти SNP для G1475A і G1891A досліджували методами ПЛР-ПДРФ та алель-специфічної ПЛР відповідно. ***Результати****. Отримано дані стосовно розподілу генотипів і алельних варіантів гена EPHA1. Виявлено низьку частоту алеля 1475A (0,012) та відсутність алеля 1891A в популяції України.* ***Висновки****. Створено методики детекції замін G1475A і G1891A в гені EPHA1 та проведено попередні дослідження розподілу поліморфізмів G1475A і G1891A у популяції України.*

Ключові слова: ген EPHA1, поліморфізм, популяція.

Р. В. Гулковский, С. Ю. Чернушин, С. А. Кравченко,А. М. Бычкова, Л. А. Лившиц

Анализ SNPs гена *EPHA1* в популяции Украины

Резюме

Цель*. Анализ распределения аллельных вариантов полиморфизмов G1475A и G1891A гена EPHA1 в популяции Украины и изучение вторичной структуры белка как первый этап определения роли гена EPHA1 в патогенезе интеллектуальной недееспособности.* ***Методы****. Аллельные варианты SNP для G1475A и G1891A исследовали методами ПЦР-ПДРФ и аллель-специфической ПЦР соответственно.* ***Результаты****. Получены данные относительно распределения генотипов и аллельных вариантов гена EPHA1. Выявлена низкая частота аллеля 1475A (0,012) и отсутствие аллеля 1891A в популяции Украины.* ***Выводы****. Созданы методики обнаружения замен G1475A и G1891A в гене EPHA1 и проведены предварительные исследования распределения полиморфизмов G1475A и G1891A в популяции Украины.*

Ключевые слова: ген EPHA1, полиморфизм, популяция.

REFERENCES

1. *Klein R.* Eph/ephrin signaling in morphogenesis, neural development and plasticity // Curr. Opin. Cell Biol.–2004.–**16**, N 5.–P. 580–589.
2. *Poliakov A., Cotrina M., Wilkinson D. G.* Diverse roles of eph receptors and ephrins in the regulation of cell migration and tissue assembly // Dev. Cell.–2004.–**7**, N 4.–P. 465–480.
3. *Zisch A. H., Pasquale E. B.* The Eph family: a multitude of receptors that mediate cell recognition signals // Cell Tissue Res.–1997.–**290**, N 2.–P. 217–226.
4. *Depaepe V., Suarez-Gonzalez N., Dufour A., Passante L., Gorski J. A., Jones K. R., Ledent C., Vanderhaeghen P.* Ephrin signalling controls brain size by regulating apoptosis of neural progenitors // Nature.–2005.–**435**, N 7046.–P. 1244–1250.
5. *Cooke J. E., Moens C. B.* Boundary formation in the hindbrain: Eph only it were simple... // Trends Neurosci.–2002.–**25**, N 5.–P. 260–267.
6. *Flanagan J. G., Vanderhaeghen P.* The ephrins and Eph receptors in neural development // Annu. Rev. Neurosci.–1998.–**21**.–P. 309–345.
7. *Maru Y., Hirai H., Takaku F.* Overexpression confers an oncogenic potential upon the *eph* gene // Oncogene.–1990.–**5**, N 3.–P. 445–447.
8. *Abdul-Aziz N. M., Turmaine M., Greene N. D., Copp A. J.* Ephrin A-EphA receptor interactions in mouse spinal neurulation: implications for neural fold fusion // Int. J. Dev. Biol.–2009.–**53**, N 4.– P. 559–568.
9. *Duffy S. L., Steiner K. A., Tam P. P., Boyd A. W.* Expression analysis of the Epha1 receptor tyrosine kinase and its high-affinity ligands Efna1 and Efna3 during early mouse development // Gene Expr. Patterns.–2006.–**6**, N 7.–P. 719–723.
10. Maniatis *T., Fritsch E. F., Sambrook J.* Molecular cloning: a laboratory manual.–New York: Cold Spring Harbor Lab. press, 1982.–545 p.
11. *Cordell H. J., Clayton D. G.* Genetic association studies // Lancet.–2005.–**366**, N 9491.–P. 1121–1131.
12. *Balding D. J.* A tutorial on statistical methods for population association studies // Nat. Rev. Genet.–2006.–**7**, N 10.–P. 781–791.
13. *Sanger F., Nicklen S., Coulson A. R.* DNA sequencing with chain-terminating inhibitors // Proc. Natl Acad. Sci. USA.–1977.–**74**, N 12.–P. 5463–5467.
14. *Naj A.C., Jun G., Beecham G. W. et al.* Common variants at MS 4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's Disease // Nat. Genet.–2011.–**43**, N 5.–P. 436–441.
15. *Hollingworth P., Harold D., Sims R. et al.* Common variants at ABCA7, MS4A6A/MS4A4E, EPHA 1, CD33 and CD2AP are associated with Alzheimer's disease // Nat. Genet.–2011.–**43**, N 5.–P. 429–435.

Received 02.04.13