Comparative Analysis of Phenylalanine Hydroxylase Mutations Spectrum in Novosibirsk and Kemerovo regions of Western Siberia, Russia

1, *Olga A. Baturina
1, 2 Igor V. Morozov

1 *Corresponding author
Institute of Chemical Biology and Fundamental Medicine SB RAS, Russian Federation
E-mail: Olga.Baturina@niboch.nsc.ru
8 Lavrentiev Avenue, Novosibirsk 630090
E-mail: Olga.Baturina@niboch.nsc.ru
2 Novosibirsk State University, 2 Pirogova Str., Novosibirsk, 630090, Russian Federation
E-mail: mor@niboch.nsc.ru

Abstract

Keywords: phenylketonuria, PKU, phenylalanine hydroxylase, PAH, genotype, phenotype.

Introduction
Phenylketonuria (PKU; MIM 261600) is a severe genetic disorder caused in most cases by the lack of phenylalanine hydroxylase (PAH; EC 1.14.16.1) activity leading to a failure in phenylalanine (Phe) to tyrosine (Tyr) conversion [1]. Accumulation of phenylalanine and toxic alternative pathways byproducts like phenylpyruvate, vinylacetate, phenyllactate, phenylacetylglutamine results innumerable symptoms including mental retardation. The lack of phenylalanine hydroxylase activity is commonly caused by the mutations in the corresponding 90 Kb long gene (PAH) located in chromosome 12 long arm segment q22-q24. Gene includes 13 exons encoding 451 a.a. protein [2]. The trait has autosomal recessive inheritance. There are more than 800 types of PKU-associated mutations in PAH locus known up to date and the amount is steadily increasing (http://www.pahdb.mcgill.ca). PKU is one of the most common genetic disorders with average frequency estimated as 1 per 10000 newborns. It is even more common in Western Siberia – 1 per 7000 newborns [3].

The present study was aimed to assess the spectrum of PKU-associated PAH gene mutations in PKU patients of Western Siberia and to compare mutations spectra of two different regions:

UDC 575.224.22
Novosibirsk region and Kemerovo region. The comparative analysis could provide the insights for understanding of the genetic structure and genetic history of populations. The precise identification of PKU-associated mutations types is important for providing personalized treatment and family planning for PKU-patients and their families involved in the study.

**Materials and Methods**

**Patients**

The cohort studied was composed of 115 unrelated PKU patients aged 5 years or less from Western Siberia, Russia. The PKU diagnosis was primary established by neonatal biochemical tests during years 2005-2013. Blood Phe concentration was assessed at day 3 or 4 after birth, and, if found elevated, the test was repeated later to confirm the diagnosis. Only patients with blood Phe levels of 120 microM or more were included in the studied cohort. The mutations types and inheritance was further confirmed by genotyping of patient’s parents and sibs. Of total 115 patients studied 67.6% were residents of Novosibirsk region and 32.4% were residents of Kemerovo region.

**Methods**

The Phe concentration was assessed via fluorescent analyser Delphia-Victor (Perkin Elmer, Finland) according to manufacturer instructions. DNA was isolated from blood nuclear cells and purified by peptides precipitation in the presence of NaCl according to [4]. Exons and adjacent introns regions were PCR-amplified as 13 separate amplicons. PCR reaction mix of 40 mkl contained the following: 65 mM Tris:HCl (pH 8.9), 16 mM (NH4)2 SO4, 1.5 mM MgCl2, 0.01% Tween-20, 10 mM 2-mercaptoethanol, 0.1 mM dNTP, 0.2 mM oligonucleotide primers, 50-100 ng of genomic DNA, 2 u.a. Taq DNA pol (ICBFM SD RAS, Russia). PCR products amount and size were confirmed by agarose gel electrophoresis. Reaction mix for Sanger reaction of total volume 30 mkl contained 0.25-0.5 pmole of PCR product, 10 pmole of oligonucleotide primer, was done in 1 mkl of BigDye v.3.1 reagent and 6 mkl 5x sequencing buffer from BigDye Cycle Terminators Sequencing Kit (Applied Biosystems, UK). Cycling conditions for Sanger reaction were: 96°C for 1 min followed by 38 cycles 3 steps each: 98°C for 10 sec; 50°C for 5 sec; 60°C for 4 min. Unincorporated dyes and low M.W. components were removed via CentriSep spin columns (Princeton Separations, USA) according to manufacturer instructions. Purified Sanger reaction products were analyzed on ABI3130xl Genetic Analyser (Applied Biosystems, USA) in SB RAS Genomics Core Facility (ICBFM SB RAS, Novosibirsk, Russia).

The homozygosity level of alleles was assessed using equation $j=\Sigma f_i^2$, where $f_i$ is allele frequency for allele i, as described in [5].

**Results**

We have identified 32 PKU-associated mutations types in the cohort studied, 86.7% of which being missense mutations, 9.5% - splicing mutations and 1.4% - deletions (see Table 1 for mutations types and frequencies). Vast majority of the mutations (55.8%) were in hemizygous state. Mutations were distributed over almost all of 13 PAH gene exons save exons 8 and 9. We also have found 7 neutral polymorphic sites: IVS1+62C>T, IVS1+134A>G, IVS2+19T>C, IVS3-22C>T, IVS4+47C>T, IVS5-54G>A, p.Q232Q.

Despite the relatively high number of identified mutations types only six of them (p.R408W, p.P281L, p.R261Q, p.R158Q, p.Y414C и IVS10-11G>A) account for more than 80% PKU-associated alleles. Relatively high diversity of mutation types could be accounted for intense migrations during contemporary population formation as well as it’s highly mixed ethnic nature.

Homozygosity index for the cohort studied was 0.38 which is comparable with ethnically heterogenous Northern and Eastern Europe populations. For comparison we calculated homozygosity indexes from the literature data on mutations frequencies for some European countries: 0.58 for Latvia [6], 0.55 for Lithuania [7], 0.31 for Czech Republic [8], 0.38 for Poland [9], 0.38 for Iceland [10], 0.20 for Denmark [11]. On the contrary, calculated homozygosity indexes for more genetically homogenous Asian appeared to be much lower: 0.12 for Japan [12], 0.051 for Korea [13], 0.043 for China [14].

In our study we compared PKU-associated mutations spectra for two vast and highly populated regions of Western Siberia. Novosibirsk region situated in the middle of Eurasia and almost in the middle of Russian Federation on the South-East part of the West Siberian Lowland, one of the greatest plains in the World. The region is 178 thousands square kilometers vast which is...
1% of Russian Federation territory. On 2010 it was populated by 2.66 million inhabitants representing 1.87% of Russian Federation population. Ethnically region population consists of mostly Russians (93.1%) with some Germans (1.2%), Ukrainians (0.9%), Tatars (0.6%) and some other nationalities (3.9% in total). Kemerovo region occupies the branches of Altay and Sayan Mountains at south-east of Western Siberia. It extends for 95.5 thousands square kilometers (4% of western Siberia and 0.56% of Russian Federation) and gives home for 2.76 million residents (1.87% of Russian Federation population) being the most dense populated part of Siberia. Russians also represent the majority of the region population (93.7%). Other ethnic groups present are also Tatars (1.5%), Germans (0.9%), Ukrainians (0.8%) and some other nationalities (4.1% in total).

Seventy nine unrelated PKU patients aged under 2 years participated in the study in Novosibirsk region. We were able to identify both PKU-associated PAH gene mutations in 75 of participants (94.9% of the cohort). Most widespread genotype p.R408W/p.R408W was identified in 30 patients (37.9% of the cohort); the next common genotype appeared to be p.R261Q/p.R408W identified in 9 patients. Missense mutation p.R408W appeared to be prevailing allele with allele frequency 63.6%. Only two other mutations (p.R261Q and p.R158Q) scored more than 5% allele frequency each. Several mutations (p.P281L, p.Y414C, IVS12+1G>A, IVS10-11G>A, IVS4+5G>T, p.L48S, p.R261X) had allele frequency from 1% to 3%. All other identified mutations were present as single cases. The notable feature of the region mutations spectrum was the discovery of several mutations (IVS2+5G>A, p.E390G, p.A403V, IVS1+5G>T, p.S349P) described as very rare or single cases in the populations of initial mutation discovery.

The cohort of Kemerovo region residents consisted of 37 patients and both PKU-associated PAH gene mutations were identified in 34 of them (91.9%). Mutation p.R408W was again the most prevailing (56.2% allele frequency) being found in homozygous state in 13 patients and in hemizygous with other mutations in another 15 patients. Only two mutations (p.Y414C and IVS10-11G>A) appeared to have allele frequency above 4%, two other rare mutations (p.R243Q and p.R155H) were discovered in one patient each but in homozygous state. All other identified alleles were present in just one occasion.

Discussion

Based on PAH gene mutations identification by the DNA sequencing of the corresponding loci we compared the PKU-associated mutation spectra for Novosibirsk and Kemerovo regions - two major districts of Western Siberia. As the possible factors playing role in this spectra formation one might regard the migration processes during the formation of contemporary regions populations. The initial development of Siberia by pioneers moving towards the Pacific coast was performed mainly by two social groups: developers (hunters, traders, manufacturers), capable for claiming of vast untouched territories for living and commercial use, and refugees expelled by authorities or life hardships into far previously unpopulated territories [15, 16]. Recently (since economic crisis of 1990-th) Novosibirsk had become the biggest migrations assimilation and redistribution center, where ample migrations streams of Russian-speaking people from Kazakhstan and former soviet Middle Asia republics were heading to. The natural conditions in the region are more favorable for housing and agriculture than in the other vast Siberian and Far East territories [17]. Contemporary Kemerovo region is one of the most industrial regions in Russia Federation [18], which population was also in significant part formed by active migration, particularly from the European part of the Soviet Union in the beginning and the middle of 20th century. The above mentioned migration flows could account for high polymorphism of PKU-associated mutations in Novosibirsk and Kemerovo regions.

The PKU-associated mutations spectra in Novosibirsk and Kemerovo regions share several common features. In both regions p.R408W absolutely dominates like in many European populations (76.0% in Latvia [6], 66.6% in Ukraine [19], 55.0% in Poland [9], 42.1% in Czech [8]. The next common mutation p.R261Q (13 hemizygous cases in Novosibirsk region and 3 – in Kemerovo region) is known to be wide spread in Switzerland and North Italy [10], Portugal [20] and Turkey [21]. The presence of p.R261Q could indicate the presence of Turkic alleles possibly introduced during Kipchak military tribes invasions [22]. The p.R158Q mutation frequency is below 4% in both regions. In many European populations this mutation is more common with frequencies from 5% to 10% [10]. The mutation p.P281L (3 hemizygous cases in each Novosibirsk and Kemerovo regions) is known to be common in South Europe [23 – 25] and to prevail in Iran,

Many rare mutations were identified in Novosibirsk region patients, for instance rare splice mutations: IVS4+5G>T, previously described in Poland (9), IVS2+5G>A - initially discovered in Germany [33], IVS2-13T>G first reported in Italy [23] and IVS1+5G>T discovered in Denmark [33]. Three single cases of deletions S16>XfsX1, IVS2+1delG, D222>STOP were also found in Novosibirsk region patients. Low frequency (less than 2%) mutations were also present in Novosibirsk region cohort in quite a variety: p.L48S, p.R243Q, p.R261X, p.R243X, p.E280K, p.E390G, p.A403V, p.P407L, p.R408Q (see Table 1).

The mutations spectrum for Kemerovo region was not so diverse. We identified single cases of splice mutation IVS11+1G>C first mentioned in 1995 in Indian patient [34] and deletion E221_D222>Efs single cases described in Germany and Denmark [35]. We also found rare mutation p.Y386C, previously described in single occasions in USA, Ireland and Italy [5, 36 and 37]. Rare mutation p.R243Q was identified in homozygous state in a patient from Kemerovo region, being probably the result of the marriage of close relatives. Mutations of moderate frequencies (2-3%) in Kemerovo region were presented by p.R68S, p.R155H, p.Y168H, p.R243Q, p.R243X, p.A342T, p.Y386C, p.Y414C (see Table 1). Common for European population’s mutation p.R252W was found in a single occasion in Kemerovo region, but not in Novosibirsk region.

Data on alleles (Table 1) and genotypes (Table 2) frequencies suggest strong influence of genes flows from Eastern (IVS4+5G>T, IVS2+5G>A, IVS2+5G>C, S16>XfsX1, D222>STOP, p.A403V, p.P407L), South (IVS2-13T>G) and Western (p.S349P, p.E280K) Europe for Novosibirsk region genes pool formation with significant income from Turkey (p.R261Q) as well. In Kemerovo region not only mutations of European origin were identified but some of South-Eastern Asia origin as well: p.R243Q and p.R155H. The mutation p.R243Q, discovered in homozygous state, is known to be common in Japan, Korea and China (18% in Chinese population) [38].

When comparing our data on PKU-associated PAH gene mutations diversity with the similar data for other regions [39] of Russian Federation one could notice general increase of rare mutations diversity by a price of the most common mutations (p.R408W, p.P281L) share when moving from West to the East. This tendency is particularly notable in Novosibirsk region, probably being a result of intense migrations flows during the area population formation especially in 20th century. For instance, migration income in Novosibirsk region during seven years from 2000 till 2006 reached more than 50 thousands with Kazakhstan, Uzbekistan, Kirgizia, Latvia, Moldova, Germany and Israel being the main sources of migration. During the same seven years total migrants to Kemerovo region reached more than 30.6 thousands. More than a half (52%) migrated to Kemerovo region from Kazakhstan, 35% migrated from Middle Asia (Kirgizia and Uzbekistan) and Ukraine, the rest 13% came from other ten CIS (Commonwealth of Independent States) and Baltic countries. It is also worth to note significant difference in the frequencies of the second by prevalence mutation p.R261Q between the two regions: 8.23% in Novosibirsk region versus 4.1% in Kemerovo region.

Acknowledgments
Authors thank Tatiana V. Lukjanova, Olga V. Podosinova for sharing PKU patient’s blood samples.

References:
34. Guldberg P. 1995; Nov 30/95 to Consortium.

Table 1: PKU-associated PAH gene alleles frequencies

<table>
<thead>
<tr>
<th>Location</th>
<th>Mutation</th>
<th>Allele numbers and frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>protein</td>
<td>cDNA</td>
</tr>
<tr>
<td>exon 1</td>
<td>S16&gt;XfsX1</td>
<td>c.47_48delCT</td>
</tr>
<tr>
<td>intron 1</td>
<td>IVD1+5G&gt;T</td>
<td>c.60+5G&gt;T</td>
</tr>
<tr>
<td>exon 2</td>
<td>p.L48S</td>
<td>c.143T&gt;C</td>
</tr>
<tr>
<td>intron 2</td>
<td>IVD2+1delG</td>
<td>c.196delG</td>
</tr>
<tr>
<td>intron 2</td>
<td>IVD2+5G&gt;A</td>
<td>c.168+5G&gt;A</td>
</tr>
<tr>
<td>intron 2</td>
<td>IVD2+5G&gt;C</td>
<td>c.168+5G&gt;C</td>
</tr>
<tr>
<td>intron 2</td>
<td>IVD2-13T&gt;G</td>
<td>c.169-13T&gt;G</td>
</tr>
<tr>
<td>exon 3</td>
<td>p.R68S</td>
<td>c.204A&gt;T</td>
</tr>
</tbody>
</table>
**Table 2: PKU-associated PAH gene genotypes frequencies**

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Novosibirsk region</th>
<th>Kemerovo region</th>
</tr>
</thead>
<tbody>
<tr>
<td>p.R408W/p.R408W</td>
<td>30 (38.5%)</td>
<td>13 (38.2%)</td>
</tr>
<tr>
<td>p.R408W/p.R261Q</td>
<td>9 (11.5%)</td>
<td>1 (2.9%)</td>
</tr>
<tr>
<td>p.R408W/p.R158Q</td>
<td>6 (7.7%)</td>
<td>1 (2.9%)</td>
</tr>
<tr>
<td>p.R408W/IVS10-11G&gt;A</td>
<td>3 (3.8%)</td>
<td>1 (2.9%)</td>
</tr>
<tr>
<td>p.R408W/IVS4+5G&gt;T</td>
<td>3 (3.8%)</td>
<td>1 (2.9%)</td>
</tr>
<tr>
<td>p.R408W/p.P281L</td>
<td>2 (2.6%)</td>
<td>2 (5.9%)</td>
</tr>
<tr>
<td>p.R408W/p.R261X</td>
<td>2 (2.6%)</td>
<td>1 (2.9%)</td>
</tr>
<tr>
<td>p.R408W/p.Y414C</td>
<td>1 (1.3%)</td>
<td>1 (2.9%)</td>
</tr>
<tr>
<td>p.R408W/p.E390G</td>
<td>1 (1.3%)</td>
<td>1 (2.9%)</td>
</tr>
<tr>
<td>p.R408W/IVS1+5G&gt;T</td>
<td>1 (1.3%)</td>
<td>1 (2.9%)</td>
</tr>
<tr>
<td>p.R408W/p.L48S</td>
<td>1 (1.3%)</td>
<td>1 (2.9%)</td>
</tr>
<tr>
<td>p.R158Q/p.R261Q</td>
<td>1 (1.3%)</td>
<td>1 (2.9%)</td>
</tr>
<tr>
<td>p.L48S/p.A403V</td>
<td>1 (1.3%)</td>
<td>1 (2.9%)</td>
</tr>
<tr>
<td>p.R408W/IVS12+1G&gt;A</td>
<td>3 (3.8%)</td>
<td>1 (2.9%)</td>
</tr>
<tr>
<td>p.R408W/X</td>
<td>2 (2.6%)</td>
<td>1 (2.9%)</td>
</tr>
<tr>
<td>p.R408W/S16&gt;XfsX1</td>
<td>1 (1.3%)</td>
<td>1 (2.9%)</td>
</tr>
<tr>
<td>p.R408W/IVS2-13T&gt;G</td>
<td>1 (1.3%)</td>
<td>1 (2.9%)</td>
</tr>
<tr>
<td>p.R408W/IVS2+5G&gt;A</td>
<td>1 (1.3%)</td>
<td>1 (2.9%)</td>
</tr>
<tr>
<td>p.R408W/p.R68S</td>
<td>1 (1.3%)</td>
<td>1 (2.9%)</td>
</tr>
<tr>
<td>p.R408W/p.Y168H</td>
<td>1 (1.3%)</td>
<td>1 (2.9%)</td>
</tr>
<tr>
<td>p.R408W/E221_D222&gt;Efs</td>
<td>1 (1.3%)</td>
<td>1 (2.9%)</td>
</tr>
<tr>
<td>X</td>
<td>4 (2.47%)</td>
<td>3 (4.1%)</td>
</tr>
</tbody>
</table>

X- mutations in any exon or adjacent intron regions not found.
p.R408W/p.E280K 1 (1.3%)
p.R408W/p.S349P 1 (1.3%)
p.R408W/p.Y386C 1 (2.9%)
p.R408W/p.P407L 1 (1.3%)
p.R408W/p.R408Q 1 (1.3%)
IVS2+1delG/p.P281L 1 (1.3%)
p.R155H/p.R155H 1 (2.9%)
p.R408W/p.R243X 1 (2.9%)
p.R243Q/p.R243Q 2 (5.9%)
p.R243Q/X 1 (1.3%)
p.R252W/p.Y414C 1 (2.9%)
p.R261Q/p.P281L 1 (1.3%)
p.R261Q/IVS10-11G>A 1 (1.3%)
p.R261Q/IVS12+1G>A 1 (2.9%)
p.R261Q/X 1 (1.3%)
p.P281L/p.A342T 1 (2.9%)
IVS10-11G>A/IVS11+1G>C 1 (2.9%)
p.Y386C/X 1 (2.9%)

X- mutations in any exon or adjacent intron regions not found