

Genetic variation in phenylketonuria: analysis of the PAHvdb database

Alin Remus Iuhas^{1,2}, Claudia Jurca^{1,3}, Marius Bembea^{1,3}

¹ Faculty of Medicine and Pharmacy, University of Oradea, Romania

² Pediatrics II, Emergency Clinical County Hospital, Oradea, Romania

³ Bihor Regional Center for Medical Genetics, Emergency Clinical County Hospital, Oradea, Romania

ABSTRACT

Introduction. Phenylketonuria (PKU) is the most frequent inborn metabolism error. The principal determinant factor for the metabolic phenotype in PKU is the residual enzymatic activity, which is determined by the variations in the phenylalanine hydroxylase (PAH) gene. To date, there are known over 1200 PAH gene variants contained in the PAH *International Database of Variations in Phenylalanine Hydroxylase Gene* (PAHvdb).

Aim. The aim of this study is to elaborate an updated PAH variant report with their type, localization, frequency and severity.

Material and method. The PAH variant analysis was made using PAHvdb. PAHvdb presently contains 1285 PAH variants and is connected to BIOPKU genotype-phenotype database which has anonymized information from nearly 18000 PKU patients, with data regarding the genotype and the correspondent phenotype.

Results. From the 1285 studied variants, the most frequent variants are substitutions – 1051 (81.8%) and deletions – 150 (11.7 %). The majority (723 - 56.3%) is represented by missense mutations, followed in frequency by frameshift variants - 177 (13.8%), splice ESS (Exonic Splicing Silencer) variants - 127 (9.9%) and nonsense mutations - 86 (6.7%). The most affected region of the gene is exon 6 (169 variants), followed by exon 7 (151 variants), exon 3 (130 variants) and exon 11 (121 variants). The majority of variants are located in the catalytic (57.66%) domain. The most frequent 3 alleles are c.1222C>T with a frequency of 19,2%, c.1066-11G>A with a frequency of 6,8% and c.782G>A with a frequency of 5.5%. From the total 1285 variants, 488 (38.5%) cause a severe phenotype, 57 (4.9%) cause a moderate phenotype and 74 (6%) cause a mild phenotype; in the case of 666 variants (50,6%) which have a low allelic frequency the metabolic phenotype couldn't be established.

Discussions. The majority of the variants are substitutions, missense in the catalytic domain of the gene, on the 6th exon. Approximately 50% of the alleles are found in single patients so they can't be used for phenotypic prediction. The majority of variants are causing a severe phenotype.

Conclusion. The existence of a database with a large number of PKU patients and PAH variants brings an important contribution to the understanding of the genotype-phenotype relation and to the capacity of better phenotypical prediction based on genotype.

Keywords: phenylketonuria, PAH, PAH genetic variants, PAHvdb

INTRODUCTION

Phenylketonuria (PKU), caused by a mutation in the phenylalanine hydroxylase gene (PAH), is the most frequent inborn metabolism error [1], with a global incidence of 1:24.000 newborns [2]. PAH vari-

ants, with autosomal recessive transmission, lead to a decrease in enzymatic activity and a phenylalanine (Phe) accumulation to neurotoxic levels [3]. In more seldom cases, the hyperphenylalaninemia is produced by tetrahydrobiopterin (BH4) deficit, a cofactor to Phe metabolism [4].

Corresponding author:

Alin Remus Iuhas

E-mail: alin.iuhas@yahoo.com

Article History:

Received: 1 March 2022

Accepted: 10 March 2022

The determinant factor for the phenotypic expression in PKU is the residual enzymatic activity. According to the pre-treatment levels of Phe, PKU classifies in: (i) classic PKU (cPKU) in which the enzymatic activity is completely (or almost completely) abolished, leading to a plasmatic Phe level of over 1200 $\mu\text{mol/L}$ (20 mg/dl); (ii) mild PKU (mPKU) with a residual PAH activity and Phe plasmatic levels between 600 and 1200 $\mu\text{mol/L}$ (10-20 mg/dl); and mild hyperphenylalaninemia (HPA) with plasmatic Phe levels of 120-360 $\mu\text{mol/L}$ (2-6 mg/dl) [2]. Some authors use an intermediary form of PKU – a moderate form, with Phe plasmatic levels of 900-1200 $\mu\text{mol/L}$ (15-20 mg/dl) [5]. The large phenotypic spectrum is due mainly to the many genetic variants. Presently, there are known over 1200 PAH mutations, contained in the *PAH International Database of Variations in Phenylalanine Hydroxylase Gene* (PAHvdb) [6].

PAH gene, responsible for the PAH enzyme, was identified on the human chromosome 12q22–24.1. It has a length of 90 kb (171 kb if the untranslated regions -UTR flanking regions are included), with 13 exons and 3 domains: regulatory, catalytic and oligomerization domain [1].

AIM AND IMPORTANCE OF THE STUDY

The objective of this article is to elaborate an updated report of the known PAH variants with their type, localization, frequency and their phenotypic implications, aiming to contribute to a better understanding of phenotype-genotype relation, critical in the disease severity prognostic.

MATERIAL AND METHOD

This article is an original study realized with data extracted from PAHvdb.

PAHvdb (PAH International Database of Variations in Phenylalanine Hydroxylase Gene) [6] contains 1285 PAH variants. It is connected to the BIOPKU genotype – phenotype database which offers supplementary information. BIOPKU database [7] gathers anonymized information from nearly 18.000 PKU patients, providing data about their genotype, the corresponding phenotype, BH4 response (when reported) or pre-treatment Phe levels (when reported).

The collected information was used to establish some indicators like:

- arbitrary assigned value (AV),
- allelic phenotype values (APV),
- allelic frequency

The AV, used for the first time in 1998 by Gulberg et al. is a numeric value assign to each genetic variant depending on the metabolic phenotype se-

verity. Initially, this classification was made using 686 patients with 133 different PAH variants [8]. Depending on the AV, the PAH variants are divided in: (i) variants that determine cPKU (AV=1), (ii) variants that determine mPKU (AV=4) and variants that determine HPA (AV=8). In the original article, the authors used an AV=2 to describe the moderate forms of PKU, not used in this study.

APV also defines the phenotypic severity of PKU by making a genotype – phenotype correlation [9]. In the BIOPKU database APV is calculated using over 10500 PKU or HPA patients, with over 800 different PAH variants. With values between 0 and 10 (0 being the value assigned to the most severe form of cPKU and 10 to the mildest form of HPA), APV classifies PKU in cPKU (APV= 0 – 2.7), mPKU (APV= 2.8 – 6.6) and HPA (APV= 6.7 – 10); an -11 value signifies an unknown APV.

The data accumulated from the PAHvdb and BIOPKU database provides information to a series of algorithms used to determine the protein stability – the principal determinant mechanism of the metabolic phenotype and the key in understanding the genotype – phenotype correlation. These algorithms are FoldX, SIFT, Polyphen 2 and SNPs 3D.

FoldX is an algorithm that uses an empiric force field that can establish the energetic effect of a mutation and also the energetic interaction of protein complexes [7,10].

SIFT (Sorting Intolerant From Tolerant) works on the supposition that some important position in the sequence of a protein was preserved during the evolution process and mutations at these positions may alter the protein function. SIFT sorts this mutation, on the important positions (intolerant) from the other mutations, with less impact (tolerant) [11].

PolyPhen-2 (Polymorphism Phenotyping v2) is an algorithm that uses physical consideration to compare a normal gene to a mutated one, and to assess the possible impact of the mutation on the gene structure and functionality [7,12].

SNPs 3D identifies, by protein analysis, which protein substitution may destabilize the protein [13].

The allele frequency was calculated based upon 16.270 alleles contained in the BIOPKU database in August 2015 [7].

For the statistical processing of the data IBM SPSS Statistics Version 26 was used. A $p < 0,05$ was considered statistically relevant.

RESULTS

From the 1285 analyzed variants, the majoritarian type was substitution, followed by deletion – 150 (11.7 %); the rest of the types together represent 6,5% from the total (Figure 1).

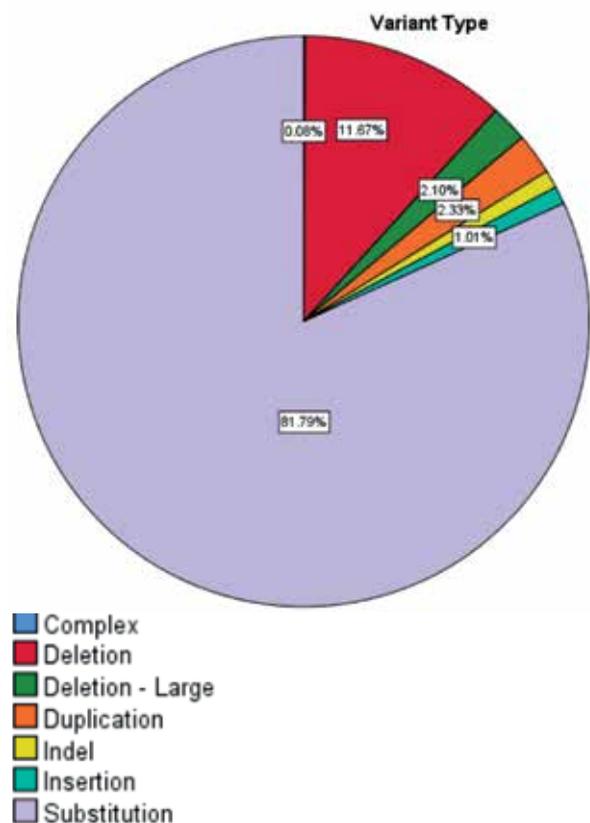


FIGURE 1. The distribution of PAH variants based on the type of the mutation

The majority of the variants (723 - 56.3%) are missense, 177 (13.8%) are frameshift mutations, 127 (9.9%) are Splice ESS - Exonic Splicing Silencer, 86 (6.7%) are nonsense mutations. The rest of the mutations are rarer and they are found in Table 1.

TABLE 1. The distribution of PAH variants based on the effect of the mutation

Effect	Frequency	Percentage
Complex	3	0.2
Extension	2	0.2
Frame shift	177	13.8
In-frame	18	1.4
Missense	723	56.3
Non-ESS	79	6.1
Nonsense	86	6.7
Splice ESS	127	9.9
Synonymous	55	4.3
Unknown	15	1.2
Total	1285	100.0

The most affected region of the gene is the 6th exon (169 variants, representing 13,2% from the total), followed by the 7th exon with 151 variants (11,8% from the total), 3rd exon (130 variants reported at this level, representing 10,1%) and 11th exon with 121 variants (9,4%) (Figure 2)

Most of the variants are located in the catalytic domain (57.66%), 18.28% are located in the regulatory domain and 5.29% on the oligomerizator domain. The rest of the variants (18.75%) are found on the introns or on the UTR regions.

The allele frequency is reported in 606 variants. The three most frequent allele are c.1222C>T (p. Arg408Trp) with 19,2% frequency, c.1066-11G>A (p. Gln355_Tyr356insGlyLeuGln/IVS10-11G>A) with 6,8% frequency and c.782G>A (p.Arg261Gln) with 5,5% frequency. Table 2 contains the most frequent 10 alleles from each metabolic phenotype.

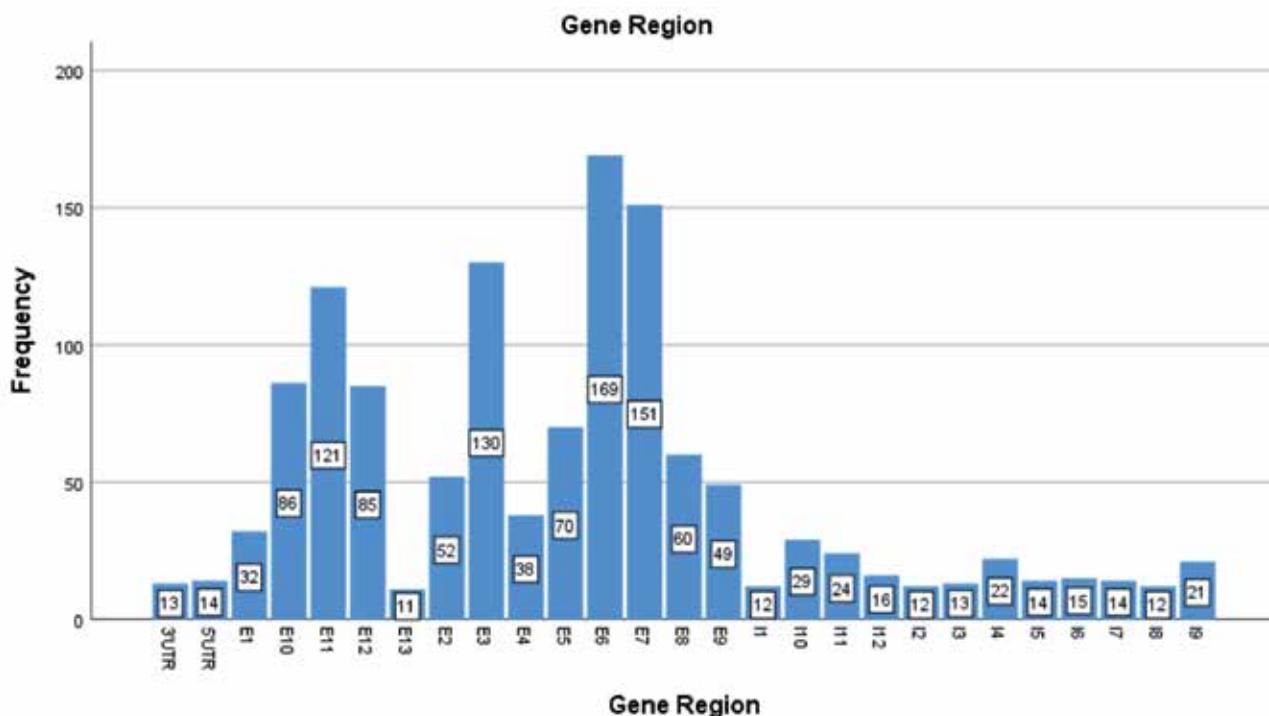


FIGURE 2. The distribution of PAH variants based on the affected region (UTR = UNTRANSLATED REGION, E = EXON, I = INTRON)

Protein Variant	cDNA Aberration	Enzyme Activity (%)	Assigned Value (AV)	Allelic Phenotype Value (APV)	Allele Frequency (%)	FoldX Interpretation	SIFT Interpretation	Polyphen 2 Interpretation	SNPs 3D Interpretation
cPKU									
p.Arg408Trp	c.1222C>T	2	1	0	19.2	destabilizing	deleterious	probably damaging	destabilizing
p.Gln355_Tyr356insGlyLeuGln / IVS10-11G>A	c.1066-11G>A	5	1	0	6.8	Not defined	Not defined	Not defined	Not defined
p.Arg261Gln	c.782G>A	44	4	1.6	5.5	destabilizing	deleterious	probably damaging	destabilizing
p.(=) / IVS12+1G>A	c.1315+1G>A		1	0	4.4	Not defined	Not defined	Not defined	Not defined
p.Ile65Thr	c.194T>C	33	4	1	4	Not defined	deleterious	probably damaging	destabilizing
p.Pro281Leu	c.842C>T	2	1	0	3.1	destabilizing	deleterious	probably damaging	destabilizing
p.Arg158Gln	c.473G>A	10	1	0	2.9	destabilizing	deleterious	probably damaging	destabilizing
p.Leu48Ser	c.143T>C	39	4	2	2.7	Not defined	deleterious	probably damaging	destabilizing
p.Arg243Gln	c.728G>A	14	1	0	2.7	destabilizing	deleterious	probably damaging	destabilizing
p.Val388Met	c.1162G>A	28	4	1.9	1.8	neutral	deleterious	probably damaging	destabilizing
mPKU									
p.(Gly352Cys)	c.1054G>T			5	5	destabilizing	deleterious	probably damaging	destabilizing
p.Tyr414Cys	c.1241A>G	57	4	5.1	3.1	destabilizing	deleterious	probably damaging	destabilizing
p.Arg241Cys	c.721C>T	57	8	5.5	1.1	destabilizing	deleterious	probably damaging	destabilizing
p.Arg408Gln	c.1223G>A	41	4	5.2	0.63	destabilizing	deleterious	probably damaging	destabilizing
p.Arg68Ser	c.204A>T	25	4	5.4	0.6	Not defined	deleterious	probably damaging	destabilizing
p.Arg241His	c.722G>A	23	4	5.2	0.46	destabilizing	deleterious	probably damaging	tolerated
p.Ala309Val	c.926C>T	42	4	3.3	0.22	destabilizing	deleterious	probably damaging	destabilizing
p.(Arg261Pro)	c.782G>C	10	4	2.8	0.2	destabilizing	deleterious	probably damaging	destabilizing
p.(=) / IVS4+5G>A	c.441+5G>A		1	4.4	0.14	Not defined	Not defined	Not defined	Not defined
p.(Ala434Asp)	c.1301C>A	9	1	3.1	0.14	Not defined	deleterious	probably damaging	destabilizing
HPA									
p.Ala403Val	c.1208C>T	66	8	9.7	2.4	destabilizing	deleterious	possibly damaging	tolerated
p.Ala300Ser	c.898G>T	65	8	9.7	1.5	neutral	deleterious	probably damaging	destabilizing
p.Glu390Gly	c.1169A>G	62	8	6.9	1.3	neutral	tolerated	probably damaging	destabilizing
p.Ile306Val	c.916A>G	39	8	9.8	0.6	destabilizing	tolerated	possibly damaging	destabilizing
p.Val245Ala	c.734T>C	50	8	9.9	0.49	destabilizing	deleterious	probably damaging	destabilizing
p.Asp415Asn	c.1243G>A	72	8	10	0.44	destabilizing	tolerated	benign	tolerated
p.Glu178Gly	c.533A>G	39	8	7.6	0.3	neutral	deleterious	possibly damaging	destabilizing
p.(Thr380Met)	c.1139C>T	28	8	10	0.3	neutral	deleterious	probably damaging	destabilizing
p.Val230Ile	c.688G>A	63	8	10	0.27	neutral	tolerated	benign	tolerated
p.(Pro211Leu)	c.632C>T	72	8	10	0.21	destabilizing	deleterious	probably damaging	tolerated

TABLE 2. The most frequent variants in each metabolic phenotype

Information related to enzymatic activity is available at 140 variants from the 1285. The enzymatic activity is positively correlated with AV and APV ($p=0,004$).

AV was reported in the case of 549 variants of the 1285 and APV in the case of 587 variants. Based on AV, 421 variants (32.8% from the 1285 total) are with AV = 1, representing a severe phenotype – cPKU, 65 variants (5.1%) have an AV = 4, representing a moderate severity – mPKU and in 63 cases (4.9%) the AV = 8, defining a mild phenotype – HPA; in 736 cases (57.3%) AV wasn't assigned. Based on the APV, 488

variants (38.5%) have a severe phenotype, with 456 of these having an APV = 0, indicating the most severe form of disease; 57 variants (4.9%) are with a moderate phenotype and 76 (6%) have a mild phenotype. The APV wasn't reported in 666 variants (50,6%). AV and APV correlate positively ($p<0.0001$).

DISCUSSIONS

Despite the progress made in the last decades in the diagnosis and the treatment of phenylketonuria, the complete picture is not known yet. Globally pro-

gress is made in the identification of new PAH variants, and in the establishment of a genotype – phenotype correlation, all this with the objective of the improvement of disease evolution prediction. With the increase in molecular testing accessibility, the new PAH gene reports have also increased. The existence of a database with a large number of information related to genotype and phenotype is a great benefit in the understanding of the relationship between them. This study enlarges the sphere of knowledge regarding phenylketonuria by adding updated information of the existing PAH variants, their localization on the gene and their impact on the metabolic phenotype.

The majority of variants are substitutions, missense, in the catalytic domain, on the 6th exon. The values obtained in this study, about the distribution of the variants' localization and the type are similar with the ones in previous studies [2,14,15].

The genotype isn't the only factor that influences the phenotype. It can vary according to the promptitude of dietary treatment initiations, the access of the patients to the treatment, treatment compliance, and other environmental and epigenetic factors. This is the reason why some variants, with a low frequency, don't have an APV attributed. The APV in the case of the variants with the frequency under 0,01% is not reliable [6]. Surprisingly, about 50% of the allele are found in single patients and can't be used for phenotypic prediction [2].

Conflict of interest: none declared

Financial support: none declared

REFERENCES

- Blau N, van Spronsen FJ, Levy HL. Phenylketonuria. *Lancet*. 2010;376:1417–27.
- Hillert A, Anikster Y, Belanger-Quintana A, et al. The Genetic Landscape and Epidemiology of Phenylketonuria. *Am J Hum Genet*. 2020 Aug 6;107(2):234–50.
- Bembea M. Genetica în pediatrie – compendiu clinic. Cluj-Napoca: Editura Risoprint, 2016:210-11.
- Van Wegberg AMJ, MacDonald A, Ahring K, et al. The complete European guidelines on phenylketonuria: diagnosis and treatment. *Orphanet J Rare Dis*. 2017;12:162.
- Stone WL, Basit H, Los E. Phenylketonuria. [Updated 2021 Jan 24]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing, 2021 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK535378/>.
- Blau N, Yue W, Perez B. PAH International Database of Variations in Phenylalanine Hydroxylase Gene. Accessed 6.02.2022. <http://www.biopku.org>.
- Blau N. BIOPKUdb: Database of Patients and Genotypes Causing HPA/PKU incl. BH4-Responsive Phenotype. Accessed 6.02.2022. <http://www.biopku.org>.
- Guldberg P, Rey F, Zschocke J et al. A European multicenter study of phenylalanine hydroxylase deficiency: classification of 105 mutations and a general system for genotype-based prediction of metabolic phenotype [published correction appears in *Am J Hum Genet* 1998 Oct;63(4):1252-3]. *Am J Hum Genet*. 1998;63(1):71-9.
- Garbade SF, Shen N, Himmelreich N et al. Allelic phenotype values: a model for genotype-based phenotype prediction in phenylketonuria. *Genet Med*. 2019 Mar;21(3):580-90.
- Schymkowitz J, Borg J, Stricher F et al. The FoldX web server: an online force field. *Nucleic Acids Res*. 2005 Jul 1;33 (Web Server issue):W382-8.
- Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4:1073–1081.
- Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7:248–9.
- Yue P, Melamud E, Moulton J. SNPs3D: candidate gene and SNP selection for association studies. *BMC Bioinformatics*. 2006;7:166.
- Blau N. Genetics of Phenylketonuria: Then and Now. *Hum Mutat*. 2016;37(6):508-15.
- Vieira Neto E, Laranjeira F, Quelhas D et al. Genotype-phenotype correlations and BH4 estimated responsiveness in patients with phenylketonuria from Rio de Janeiro, Southeast Brazil. *Mol Genet Genomic Med*. 2019;7(5):610.
- Wettstein S, Underhaug J, Perez B et al. Linking genotypes database with locus-specific database and genotype–phenotype correlation in phenylketonuria. *Eur J Hum Genet*. 2015;23:302–9.
- Muntau AC, Leandro J, Staudigl M et al. Innovative strategies to treat protein misfolding in inborn errors of metabolism: pharmacological chaperones and proteostasis regulators. *J Inher Metab Dis*. 2014;37:505–23.