



Universidad Autónoma  
de Madrid

**Biblos-e Archivo**  
Repositorio Institucional UAM

**Repositorio Institucional de la Universidad Autónoma de Madrid**  
<https://repositorio.uam.es>

Esta es la **versión de autor** del artículo publicado en:  
This is an **author produced version** of a paper published in:

The American Journal of Human Genetics 107.2 (2020): 234 – 250

**DOI:** <https://doi.org/10.1016/j.ajhg.2020.06.006>

**Copyright:** © 2020 Cell Press

El acceso a la versión del editor puede requerir la suscripción del recurso  
Access to the published version may require subscription

# The genetic landscape and epidemiology of phenylketonuria

Alicia Hillert<sup>1</sup>, Yair Anikster<sup>2</sup>, Amaya Belanger-Quintana<sup>3</sup>, Alberto Burlina<sup>4</sup>, Barbara K. Burton<sup>5</sup>, Carla Carducci<sup>6</sup>, Ana E Chiesa<sup>7</sup>, John Christodoulou<sup>8</sup>, Maja Đorđević<sup>9</sup>, Lourdes R. Desviat<sup>10</sup>, Aviva Eliyahu<sup>2</sup>, Roeland A.F. Evers<sup>11</sup>, Lena Fajkusova<sup>12</sup>, François Feillet<sup>13</sup>, Pedro Eduardo Bonfim Freitas<sup>14</sup>, Maria Gżewska<sup>15</sup>, Polina Gundorova<sup>16</sup>, Daniela Karall<sup>17</sup>, Katya Kneller<sup>2</sup>, Sergey I. Kutsev<sup>16</sup>, Vincenzo Leuzzi<sup>18</sup>, Harvey L. Levy<sup>19</sup>, Uta Lichter-Konecki<sup>20</sup>, Ania C. Muntau<sup>21</sup>, Fares Namour<sup>13</sup>, Mariusz Oltarzewski<sup>22</sup>, Andrea Paras<sup>5</sup>, Belen Perez<sup>10</sup>, Emil Polak<sup>23</sup>, Alexander V. Polyakov<sup>16</sup>, Francesco Porta<sup>24</sup>, Marianne Rohrbach<sup>25</sup>, Sabine Scholl-Bürgi<sup>17</sup>, Norma Spécola<sup>26</sup>, Maja Stojiljković<sup>27</sup>, Nan Shen<sup>28</sup>, Luiz Carlos Santana da Silva<sup>14</sup>, Anastasia Skouma<sup>29</sup>, Francjan van Spronsen<sup>11</sup>, Vera Stoppioni<sup>30</sup>, Beat Thöny<sup>25</sup>, Friedrich K. Trefz<sup>1</sup>, Jerry Vockley<sup>20</sup>, Youngguo Yu<sup>31</sup>, Johannes Zschocke<sup>32</sup>, Georg F. Hoffmann<sup>1</sup>, Sven Garbade<sup>1\*</sup>, Nenad Blau<sup>1,21\*</sup>

<sup>1</sup>Dietmar-Hopp Metabolic Center, University Children's Hospital, Heidelberg, Germany; <sup>2</sup>Metabolic Disease Unit, Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Tel-Aviv University, Tel-Aviv, Israel; <sup>3</sup>Unidad de Enfermedades Metabólicas, Servicio de Pediatría, Hospital Ramon y Cajal, Madrid, Spain; <sup>4</sup>Division of Inherited Metabolic Diseases, Department of Woman's and Child's Health, University Hospital, Padua, Italy; <sup>5</sup>Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA; <sup>6</sup>Department of Experimental Medicine, Sapienza University of Rome, Italy; <sup>7</sup>Fundación de Endocrinología Infantil (FEI). Buenos Aires, Argentina; <sup>8</sup>Murdoch Children's Research Institute and Department of Pediatrics, University of Melbourne, Melbourne, Australia; <sup>9</sup>Institute of Mother and Child Healthcare "Dr. Vukan Čupić", Belgrade, Serbia; <sup>10</sup>Centro de Diagnóstico de Enfermedades Moleculares, Centro de Biología Molecular CSIC-UAM, Universidad Autónoma de Madrid. CIBERER, IdiPAz, Madrid, Spain; <sup>11</sup>University of Groningen, University Medical Center Groningen, Beatrix Children's Hospital, Section of Metabolic Diseases, Groningen, the Netherlands; <sup>12</sup>Centre of Molecular Biology and Gene Therapy, University Hospital Brno, Brno, Czech Republic; <sup>13</sup>Reference Center for Inherited Metabolic Diseases, University Hospital of Nancy, Vandoeuvre-lès-Nancy, France; <sup>14</sup>Laboratory of Inborn Errors of Metabolism, Institute of Biological Sciences, Federal University of Pará, Belém, Brazil; <sup>15</sup>Department of Pediatrics, Endocrinology, Diabetology, Metabolic Diseases and Cardiology, Pomeranian Medical University, Szczecin, Poland; <sup>16</sup>Research Centre for Medical Genetics, Moscow, Russia; <sup>17</sup>Clinic of Pediatrics, Division of Inherited Metabolic Disorders, Medical University of Innsbruck, Innsbruck, Austria; <sup>18</sup>Department of Human Neuroscience, Sapienza

35 University of Rome, Italy; <sup>19</sup>Division of Genetics and Genomics, Boston Children's Hospital,  
36 Harvard Medical School, Boston, MA, USA; <sup>20</sup>University of Pittsburgh, Children's Hospital of  
37 Pittsburgh, Pittsburgh, PA, USA; <sup>21</sup>University Children's Hospital, University Medical Center  
38 Hamburg Eppendorf, Hamburg, Germany; <sup>22</sup>Department of Screening and Metabolic  
39 Diagnostics, Institute of Mother and Child, Warsaw, Poland; <sup>23</sup>Comenius University, Faculty  
40 of Natural Sciences, Department of Molecular Biology, Bratislava, Slovak Republic;  
41 <sup>24</sup>Department of Pediatrics, AOU Citta' della Salute e della Scienza di Torino, Torino, Italy;  
42 <sup>25</sup>Division of Metabolism, University Children's Hospital, Zürich, Switzerland; <sup>26</sup>Unidad de  
43 Metabolismo. Hospital de Niños "Sor Ludovica" de La Plata, Buenos Aires, Argentina;  
44 <sup>27</sup>Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Belgrade,  
45 Serbia; <sup>28</sup>Department of Infectious Diseases, Shanghai Children's Medical Center, Shanghai  
46 Jiao Tong University School of Medicine, Shanghai, China; <sup>29</sup>Institute of Child Health, Athens,  
47 Greece; <sup>30</sup>Centro Screening Neonatale Regione Marche, Azienda Ospedaliera Ospedali Riuniti  
48 Marche Nord, Fano, Italy; <sup>31</sup>Department of Pediatric Endocrinology/Genetics, Xinhua  
49 Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai Institute for Pediatric  
50 Research, Shanghai, China; <sup>32</sup>Institute of Human Genetics, Medical University Innsbruck,  
51 Innsbruck, Austria.

## Abstract

Phenylketonuria (PKU), caused by variants in the phenylalanine hydroxylase (PAH) gene, is the most common autosomal recessive Mendelian phenotype in amino acid metabolism. We estimated that globally 0.45 million have PKU, with global prevalence 1:23,930 live births (range 1:4,500 [Italy] – 1:125,000 [Japan]). Comparing genotypes and metabolic phenotypes from 16,092 patients revealed differences in diseases severity in 51 countries from 17 world regions, with the global phenotype distribution of 62% classic PKU, 22% mild PKU and 16% mild hyperphenylalaninemia. A gradient in genotype and phenotype distribution existed across Europe, from classic PKU in the east, to mild PKU in the southwest and mild hyperphenylalaninemia in the south. The p.Y414C-associated genotype can be traced from northern to western Europe, from Sweden via Norway, Denmark to central Europe. The frequency of classic PKU increases from Europe (56%) via Middle-East (71%) to Australia (80%). Of 758 *PAH* variants, p.R408W (22.2%), c.1066-11G>A (6.4%) and p.R261Q (5.5%) were most common and responsible for two prevalent genotypes: p.[R408W];[R408W] (11.4%) and c.[1066-11G>A];[1066-11G>A] (2.6%). Most genotypes (73%) were compound heterozygous, 27% were homozygous and 55% of 3,659 different genotypes occurred in a single patient only. *PAH* variants were scored using an allelic phenotype value and correlated with pre-treatment blood phenylalanine levels (n=6,796) and tetrahydrobiopterin test results (n=4,381), enabling a genotype-based phenotype (88%) and tetrahydrobiopterin responsiveness (83%) prediction. This study shows that large genotype databases enable accurate phenotype prediction, allowing appropriate targeting of therapies to optimize clinical outcome.

## Introduction

Phenylketonuria (PKU; MIM 261600) is the most frequent inborn errors of the amino acid metabolism. It is caused by more than 1,180 biallelic variations in the phenylalanine hydroxylase (PAH) gene located on chromosome 12q22-24.1.<sup>1</sup> These autosomal recessive inherited variations lead to a deficiency in the PAH enzyme, which normally hydroxylates phenylalanine into tyrosine, with the help of a cofactor (tetrahydrobiopterin; BH<sub>4</sub>), molecular oxygen, and non-heme iron.<sup>2</sup>

The clinical picture is highly heterogeneous as it depends on the degree of residual PAH activity and blood phenylalanine (Phe) levels. Usually, lower residual enzyme activity results in higher the blood phenylalanine levels and a more severe PKU phenotype.<sup>3</sup>

The severity of PKU is classified according to the pre-treatment blood phenylalanine concentration and daily dietary phenylalanine tolerance, ranging from the most severe classical PKU (cPKU) with pre-treatment blood phenylalanine levels of >1,200 µmol/L, to mild PKU (mPKU) with pre-treatment blood phenylalanine levels of 600–1,200 µmol/L and mild hyperphenylalaninemia (MHP) with pre-treatment phenylalanine blood levels of 120–600 µmol/L.<sup>1,4</sup>

Untreated PKU in newborns can result in global developmental delay or severe irreversible intellectual disabilities, as well as growth failure, hypopigmentation, motor deficits, and ataxia, and seizures.<sup>1</sup> The population of patients with PKU is also heterogeneous in terms of treatment history and diet compliance.<sup>4,5</sup> Early diagnosis and treatment with a low-Phe diet has enabled an almost normal life for the majority of PKU patients.<sup>6</sup> Pharmacological treatment with BH<sub>4</sub> (sapropterin) and enzyme substitution therapy with Phe ammonia lyase (PAL) provide alternative treatment options for some patients.<sup>7</sup>

PKU is one of the most frequent inherited disorders in Caucasians, with an incidence of roughly 1:10,000 live births in the USA,<sup>8</sup> although the prevalence of PKU varies significantly among ethnicities and geographic regions worldwide. In Europe, the incidence of PKU ranges from 1: 850 in the Karachay-Cherkess Republic (Russia)<sup>9</sup> to only 1:112,000 live births in Finland.<sup>10</sup> PKU occurs less often Japan, with an incidences of 1:120,000.<sup>11</sup>

A large number of *PAH* variations (>1,300) give rise to a wide scale of residual PAH enzyme activities that correspond to different PKU phenotypes.<sup>12,13</sup> Associations between genotypes

and *in vitro* residual PAH activity have been documented for many *PAH* variations.<sup>3</sup> Therefore, the molecular genetics of PKU and genotype-based phenotype prediction may be clinically useful, particularly where treatment recommendations are unclear (e.g. due to borderline blood phenylalanine levels) or for genetic counseling of patients' families.

The major goal of this study was the analysis of the largest phenotypes and genotypes database to elucidate the current distribution of PKU worldwide and to create an overview of the severity of *PAH* variants, genotypes and the resulting phenotypes worldwide, in various geographic regions and respective countries. Furthermore, this work improves the accuracy the genotypic phenotype prediction by the use of allelic phenotype value (APV) value and the genotypic BH<sub>4</sub> responsiveness prediction.

## **Subjects and methods**

### **LITERATURE SEARCH**

An electronic search using the databases MEDLINE (via Pub Med), the Cochrane library and Web of Science was carried out to compare articles covering the epidemiology and genetics of PKU in different world regions between January 1980 and October 2019. Key words included Phenylketonuria[MeSH Terms] OR "phenylalanine hydroxylase/deficiency"[MeSH Terms] OR hyperphenylalaninaemia[Title/Abstract] OR hyperphenylalaninemia[Title/Abstract] OR PAH deficiency[Title/Abstract] OR phenylalanine hydroxylase deficiency[Title/Abstract] OR phenylketonuri\*[Title/Abstract] OR pku[Title/Abstract]) AND ( "1980/01/01"[PDat] : "2019/12/31"[PDat] ) AND Humans[Mesh] Filters: Publication date from 1980/01/01 to 2019/12/31; Humans.

This literature search yielded 5,459 records without duplicates, of which 1,118 papers with an appropriate title and abstract were assessed. The final number of relevant records was 256. (Supplementary Figure 1)

## DATABASES

The *PAH* locus-specific database *PAHvdb*, ClinVar, HGMD and LOVD databases were searched for variants. *PAHvdb* is linked with the genotypes-phenotypes BIOPKU database and was used for analyses in this work. It follows the HGVS nomenclature recommendations.

The BIOPKU database encompasses information about more than 16,900 PKU patients from 51 countries, providing information on patients' genotype, corresponding metabolic phenotypes, BH<sub>4</sub> responsiveness (where reported) and highest blood Phe levels before starting the treatment (where reported). Individual information was collected from the published literature or anonymized records submitted online. Phenotype information was unknown for 690 patients. Table 1 shows the information included in the database.

## DEFINITION OF PHENOTYPES

We used the term "PKU" to refer to the entire spectrum of phenotypes due to *PAH* deficiency. Since not all countries use the same nomenclature for the severity of PKU and HPA, we used in this study the following 3 phenotype groups: classical PKU (cPKU; pre-treatment blood Phe >1200 µmol/L); mild PKU (mPKU; pre-treatment blood Phe 600–1200 µmol/L) and mild HPA (MHP; pre-treatment blood Phe 120–600). Patients with a moderate PKU (pre-treatment blood Phe 900–1,200 µmol/L) were included in the mPKU group and MHP included MHP-no treatment (blood Phe 120–360 µmol/L) and MHP-gray zone (360–600 µmol/L) categories.<sup>4</sup> Any classifications, not fitting into one of the above groups (due to different country-specific classifications), were reassigned on the basis of reported pre-treatment blood Phe levels and the genotype.

## DEFINITION OF THE ALLELIC (APV) AND GENOTYPIC PHENOTYPE VALUE (GPV)

APV is a value defining the association of a variant with the corresponding metabolic phenotype, thus defining its severity. APV was calculated for variants occurring in a functionally hemizygous constellation (i.e. in a combination with an inactive 0-allele) in at least 5 patients.<sup>14</sup> APVs range between 0 and 10, with following classification definitions cPKU (APV = 0-2.7), mPKU (APV = 2.8-6.6) and MHP (APVs 6.7-10).<sup>14</sup>

The genotypic phenotype value (GPV) was calculated from APVs of both alleles and was assigned to a higher APV (APV<sub>max</sub>). This calculation was based on the fact that the milder

variant (with a higher APV) is always dominant over the severe one.<sup>14,15</sup> Possible effects of interallelic complementation and epigenetic factors, which may influence the phenotype<sup>16,17</sup>, were not considered in this study.

#### **DEFINITION OF BH<sub>4</sub> RESPONSIVENESS**

BH<sub>4</sub> responsiveness was defined as a  $\geq 30\%$  reduction of blood Phe levels within 24-48 hours after the administration of BH<sub>4</sub> (20mg/kg body weight)<sup>18,19</sup>. A linear discriminant analysis was applied to predict BH<sub>4</sub> responsiveness based on the APV for untested patients.

#### **STATISTICAL ANALYSIS**

The statistical analysis was performed using R, an open source software and flexible programming language used for the statistical data analysis as well as graphic creations (<https://www.r-project.org/>). A total of 16,196 records with a complete genotype information (variant 1 and 2 known) were analyzed.

### **Results**

#### **PREVALENCE OF PKU**

Based on the literature search and reports from national screening centers, the prevalence of PKU was estimated for 64 countries. For parts of Africa, Asia, Southern America and Caribbean there was no information. The estimated total number of PKU patients (all phenotypes) from those 64 countries in 2018 was 360,466. For the remaining 257 countries we were unable to find credible PKU prevalence sources; we used the average regional prevalence of 64 countries, multiplied by their populations, resulting in an additional 94,114 PKU patients (total about 0.45 million PKU patients). The global PKU prevalence was estimated to be 1: 23,930 newborns (Figure 1, Supplementary Table 1).

The PKU prevalence was highest in European and certain Middle Eastern populations. Italy (1:4,000) and Ireland (1:4,545) had even higher prevalence than Iran and Jordan (both 1: 5,000) or Turkey (1:6,667). Saudi Arabia (1:14,245), Iraq (1:14,286) or United Arab Emirates (1:14,493) had lower PKU prevalence, however.



188 PKU prevalence was also high in central European countries PKU affect patients at high rates  
189 as well e.g. Germany (1:5,360), Czechia (1:5,521) and Slovakia (1:5,753). Slovenia (1:7,143),  
190 Austria (1:5,764) or Poland (1:8,309) had similar rates to eastern Europe, e.g. Estonia  
191 (1:7,143), Russia (1: 7,714) Belarus (1:7,692), and Croatia (1:8,333).

192 PKU occurred slightly less frequently in western Europe, e.g. in France (1:9,091), United  
193 Kingdom (1:10,000), Belgium (1:11,000) or the Netherlands (11,546), and in southern  
194 Europe, e.g. Spain (1:10,115) and Portugal (1:12,500).

195 Northern Europe presented with the lowest PKU rates in Europe, e.g. for Norway (1:11,457),  
196 Sweden (1: 12,681), Denmark (1: 13,434), or only 1:112,000 in Finland.

197 PKU occurred more frequently In Canada (1:15,000) than in the United States of America  
198 (1:25,000) or in Latin American countries, including Chile (1:19,231), Brazil (1:25,000),  
199 Mexico (1:27,778), Argentina (1:33,333) or Peru (1:46,970).

200 The lowest PKU prevalence was reported in Asian countries, such as Thailand (1:227,273),  
201 Japan (1:125,000), Philippines (1:116,006) or Singapore (1:83,333). One exception was China  
202 where the PKU prevalence was 1:15,924, which was comparable to Europe.

## 203 **DESCRIPTIVE ANALYSIS OF THE PHENYLALANINE HYDROXYLASE GENE LOCUS-SPECIFIC DATABASE**

204 Substitutions were by far the most frequent variant type In the *PAH*vdb (80.5%), followed by  
205 deletions (12.9%) and duplications (2.1%). Of all variants, 691 were missense variants  
206 (58.3%), followed by 165 frameshift variants (13.9%) and 155 splice site (13.1%) variants.  
207 Nonsense variants (6.9%), synonymous variants (4.9%) and in-frame variants (1.9%) were  
208 less frequent. Extension, complex and unknown variants, as well as polymorphisms,  
209 accounted for the remainder. Exon 6 contained the most variants (14.1%), followed by  
210 Exon 7 (12.2%) and Exon 3 (9.9%). Most variants (59.2%) were located in the central catalytic  
211 domain, 17.5% in the N-terminal regulatory domain and 5.4% in the C-terminal  
212 oligomerization domain of the PAH monomer. The remaining variants (17.9%) were either in  
213 the intronic or UTR regions. Only 7.7% of all variants were located in one of the 4 cofactor  
214 binding regions (Supplementary Figure 2).

215 The APV was known for 589/1,186 variants. Most variants (441) were defined as severe 0-  
216 alleles (APV=0), 32 variants as cPKU phenotype, 52 as mPKU and 64 as mild MHP alleles.

## DESCRIPTIVE ANALYSIS OF THE BIOPKU DATABASE

As of October 2019, the BIOPKU database contained anonymized data on more than 16,900 PKU patients. A total of 16,196 PKU patients with 3,659 different genotypes were analyzed. Information on the PKU phenotype was available for 16,902 patients. Maximum pretreatment Phe levels were recorded for only 6,796 patients and the BH<sub>4</sub> test information was available for 4,381 patients (Table 1).

## GLOBAL PKU PHENOTYPE AND GENOTYPE DISTRIBUTION

Information on the PKU phenotype was available for 16,092 out of 16,196 patients (99%). Of these, most had cPKU (9,923; 61.7%), 3,521 (21.9%) had mPKU, and 2,648 had MHP (16.4%). Information about the phenotype was unavailable for 104 patients

The comparison of pre-treatment Phe level with the phenotype (n=6,796) is illustrated in Figure 2A and with the GPVs (n=6,371) in Figure 2B. The interquartile range (n, median, 25<sup>th</sup>–75<sup>th</sup> percentile) was smallest and lowest for the MHP patient group: 1,283, 320 µmol/L, 242–432 µmol/L, and was larger for mPKU and cPKU (1,487, 793 µmol/L, 660–793 µmol/L, and 3,599, 1,550 µmol/L, 1,270–1936 µmol/L, respectively).

A total of 758 different alleles were identified in this study. The three most prevalent variants were p.R408W, with an allele frequency (AF) =22.2%, c.1066-11G>A (IVS10-11G>A; AF=6.4%) and p.R261Q; AF 5.5%). Figure 3 shows the most frequent phenotype-specific variants. A full list of PAH variants is shown in Supplementary Table 2.

Of all patients, 11,810 (72.9%) were compound heterozygotes and 4,386 (27.1%) were homozygotes. Of 3,659 genotypes, 3,446 (94.2%) were compound heterozygotes and 213 (5.8%) homozygotes. The three most prevalent genotypes were: p.[R408W];[R408W] with a genotype frequency (GF) =11.4%, followed by c.[1066-11G>A];[1066-11G>A] (GF=2.6) and c.[1222C>T];[1315+1G>A] (p.[R408W];IVS12+1G>A) (GF=1.6). A full list of genotypes is shown in Supplementary Table 3. Strikingly, 54.5% of all genotypes were specific for only one patient and were not used for the phenotype prediction.

## **BH<sub>4</sub> RESPONSIVENESS**

Information on BH<sub>4</sub> responsiveness was available for 4,381 patients. About half were classified as BH<sub>4</sub> responsive (2,053; 47%). Supplementary Table 4 shows the distribution of BH<sub>4</sub> responsiveness in different phenotype groups. As expected, milder forms of PKU appeared more likely to be BH<sub>4</sub> responsive, whereas most patients with cPKU were non-responders.

Pretreatment blood Phe levels (n, median, 25<sup>th</sup>-75<sup>th</sup> percentile) were much lower for BH<sub>4</sub> responsive patients (1,116, 620 µmol/L, 411–853 µmol/L), compared with non-responders (1,180, 1,361 µmol/L, 1039–1719 µmol/L (Figure 2C). The 11,584 patients not tested for BH<sub>4</sub> responsiveness were analyzed for potential responsiveness using the GPVs. Patients with an GPV>3.8 (n=3,023; 26%) were assigned as potential BH<sub>4</sub> responders ( $p<0.001$  vs. patients assigned as non-responders [n=8,561; 74%]) (Figure 2D).

## **PKU PHENOTYPES AND GENOTYPES IN WORLD REGIONS**

### ***Overview of global data***

Patient data originated from 51 countries from 17 world regions and a total of 15,357 patients from 33 countries with at least 35 reported cases, were analyzed (Supplementary Tables 5 and 6, Supplementary Figure 3). Most cases were reported from central (18.3%) and eastern (18.1%) Europe, eastern Asia (13.8%), western Europe (9.6%) and the Middle-East (9.2%). Very few patients were reported from the northern and southern parts of Africa.

cPKU was the most frequent phenotype in all world regions, with high rates reported in Australia and eastern Europe (81%), and rates of <50% reported only in Serbia, Argentina, Turkey, Netherlands, Sweden, Spain, Italy, Japan, Slovenia, Germany, and Taiwan (Figure 3A). Of 2,835 patients in Russia, 81% had cPKU, 9.9 % had mPKU and 9.1% had MHP. Estonia, representing the Baltic region, reported 93.5% of patients with cPKU. The ratio of mPKU:MHP patients were comparable, except for Eastern Asia, North America and South America, where mPKU was more common. The phenotype distribution in world regions is shown in Supplementary Figure 4.

Reports from individual regions are summarized in Table 2 and below.

## 274 **Europe**

275 Variant p.R408W was the most common (AF=63.7%), followed by c.1066-11G>A (AF=11%)  
276 and p.R261Q (AF=11%). Eastern Europe had the highest AF for p.R408W (54.6%), mostly with  
277 homozygous genotype p.[R408W];[R408W] (GF=32.7%); findings were similar for central  
278 Europe (AF=44.4% and GF=23.8%). Russia contributed the largest number of records. A  
279 single variant, p.R408W (AF=53.7%), was dominant in Russia, followed by p.R261Q  
280 (AF=5.6%) and p.P281L (AF=4.1%). The most common genotype was p.[R408W];[R408W]  
281 (30.6% of 458 different genotypes). (AF=82.6% for p.R408W). In southeastern Europe  
282 p.R408W was also the most frequent allele, but p.[L48S];[R408W] was the most prevalent  
283 genotype (GF=10.6%)

284 The distribution of patients with p.R408W on at least one allele (compound heterozygotes  
285 and homozygotes) decreased from 98% in Estonia to 89% in Poland, 76% in Russia, 69% in  
286 Slovakia, 65% in Czechia, 40% in Austria, 36% in Germany, 10% in France, 6% in Italy, and  
287 only 4% in Spain (Table 3, Figure 4). Genotypes with the IVS10-11G>A splice site variant,  
288 occurred commonly in Armenia (48%), Turkey (32%), Iran (26%), Israel (21%), Spain (20%)  
289 and Italy (19%) (Table 3, Figure 4). The prevalence of p.R261Q was 10–30% in different  
290 countries.

291 The distribution of PKU phenotypes was more even in southern Europe (39.5% cPKU, 36.6%  
292 MHP, 23.9% mPKU) and the milder phenotype was more predominant, e.g. 37% had MHP in  
293 Italy. This was consistent with a high frequency of variants with a substantial residual PAH  
294 activity, e.g. p.A403V (AF=8.4), p.A300S (AF=3.6), p.V245A (AF=2.4) or p.Y414C (AF=2.4).  
295 Variants c.1066-11G>A and p.R261Q, often called the “Mediterranean mutation” accounted  
296 for majority of mutant alleles, and the most frequent genotype was c.[1066-11G>A];[1066-  
297 11G>A] (GF=5.2%). In Portugal and Italy c.1066-11G>A and p.R261Q occurred at similar rates  
298 (AF=16.2% and AF=10.7%, respectively). Variant p.V388M was also common in Portugal  
299 (AF=11.3%) and Spain (AF=6.8%).

300 Low rates of cPKU were found in southwestern Europe (42.7%) and northern Europe  
301 (50.5%). In southeastern Europe, Croatia (68%) and Bulgaria (65%) had a higher prevalence  
302 of cPKU than Serbia (48%). p.R408W was the most frequent variant in Croatia (AF=31.4%)  
303 and Bulgaria (AF=32.4%), but was less common in Serbia (AF=14.7%). The p.L48S variant,

which was initially identified in Turkey,<sup>20</sup> has the highest reported AF (31.3%) in Serbia. The most frequent genotypes in other countries locally were p.[L48S];[L48S] (14.7 %) in Serbia; p.[L48S];[R408W] (14.3 %) in Croatia; and p.[R408W];[R408W] genotype (16.2 %) in Bulgaria.

In northern Europe c.1315+1G>A was the most common splice site variant (AF=25%). Most patients were compound heterozygous for p.[R408W];c.[1315+1G>A]. Denmark had the largest sample size in this study, with the highest number of cases of the mild p.Y414C variant in the world. Sweden and Norway had a relatively high rate of mPKU (>43%). In Denmark the most frequent variant was c.1315+1G>A with an AF= 27.3%, while p.G46S occurred more commonly in Sweden and Norway. Variants p.R408Q and p.F299C were specific for Norway.

The p.R408W variant was also the most frequent variant for all central European countries, except for Switzerland where p.R261Q (AF=15.3%) was more prevalent. Classical PKU was particularly frequent in Poland and Slovakia (>70%), but not in Czechia, consistent with a higher prevalence of the mild variant, p.A403V. There was a wide spectrum of PKU variants in Germany: of 102 distinct variants, p.R408W accounted for only 19.3%, and other variants such as p.Y414C (AF=9.7 %) as well as c.1315+1G>A (AF=8.8 %) were prominent. Almost half (44%) of patients in Germany had the mPKU phenotype, 32% had cPKU and 24% had MHP.

France represented the largest European sample (n=1,307): 59.1% had cPKU, and there were a total of 229 different variants, e.g. c.1066-11G>A (AF=7.4%), p.R261Q (AF=6.5%), p.R408W (AF=5.5%), c.1315+1G>A (AF=4.5%) and p.E280K (AF=3.7%). The p.R408W variant was also less common in the Netherlands (AF=2.2%). The predominant genotype in western Europe was p.[E280K];[E280K].

### ***Latin America***

The most frequent variants here, p.V388M (AF= 13.9%), p.R261Q (AF=10.7%) and c.1066-11G>A (AF=9.4%), were also prominent mutations in southern Europe (see above).

Homozygous p.[V388M];[V388M] (GF=4.4%) occurred most frequently. Of the three relevant South American countries Brazil had the highest numbers of cPKU (63%), followed by Mexico (57%) and Argentina (43%). In comparison to Brazil or Argentina, Mexico's most frequent variant was c.60+5G>T (IVS1+5G>T).

### 333 **North America**

334 The most prevalent variants in North America also resembled those in European  
335 populations: p.R408W (18.5%), c.1066-11G>A (7.9%), c.1315+1G>A (6.9%), as did the  
336 genotype distribution p.[R408W];[R408W] (GF=4.0%)

### 337 **Middle East**

338 The predominance of c.1066-11G>A variant (AF=20.1%) and it's homozygous genotype  
339 (GF=15.3%) was evident In Iran, Turkey, Israel and Saudi Arabia. Other frequent variants  
340 included p.R261Q, c.168+5G>C (IVS2+5G>C), p.P281L and p.R243\*. 81% of the Iranian  
341 patients had cPKU. In comparison to Iran, Turkey had a similar AF for c.1066-11G>A (22.9 %)  
342 and p.R261Q (11.8 %), but strongly different phenotypes: 42% had cPKU, 35% MHP and 23%  
343 mPKU. This may be related to the augmented presence of mild variants such as p.A300S  
344 (7,4%) or p.E390G (4.2%). In contrast to other Middle Eastern countries, the most frequent  
345 variant from Saudi Arabia was p.R252W (AF=27,6%), followed by p.R261\* (11.2 %) and  
346 p.V388M (10.4 %); 30% of all patients had a p.R252W-genotype (Table 3).

### 347 **Asia**

348 Most reported cases from Asia came from the east, with the most prevalent being the  
349 missense variant p.R243Q (AF= 21.8%), followed by Ex6-96A>G (9.8%) and p.R241C (8.7%).  
350 The most frequent genotype was homozygous p.[R243Q];[R243Q] (GF=5.6%).

351 In China, 62% of patients (n=2008) had a cPKU, 28% mPKU and 10% the MHP phenotype. A  
352 total of 234 different variants were reported, out of which five p.R243Q, Ex6-96A>G,  
353 p.R241C, p.R111\* and p.R413P, each with a frequency of >5%, which accounted for 52.5% of  
354 all alleles. Furthermore 679 different genotypes were reported, the most frequent being  
355 p.[R243Q];[R243Q]. Korea was the only country that like China has a high rate of cPKU (71%),  
356 whereas cPKU in Japan (37%) and Taiwan (25%) was less common. The p.R243Q, Ex6-96A>G,  
357 and p.R241C, each with allele frequencies >5%, were detected in Korea, Taiwan, and China.  
358 The p.R111\* and p. R413P variants were common in Japan, China and Taiwan, nut not Korea.  
359 The splice variant IVS4-1G>A was much more prevalent in Korea than in other Eastern Asian  
360 countries.

Genotypes with p.R243Q were common in China (40%) and Korea (27%) (Table 3), while those with p.R241C were prevalent in Taiwan (41%), Japan (27%) and China (15%) (Table 3). Patients with the splice site variant IVS4-1G>A were more common in Korea (18%), followed by China (8%) and Japan (7%). The splice variant Ex6-96A>G (c.611A>), which masquerades as missense variant p.Y204C<sup>21</sup>), was found 19% of patients in China, 13 in Taiwan and 12 in Korea (Table 3).

### ***Africa***

The low number of reports available from Africa were considered not sufficiently representative of this world region, and we did not analyze these data.

## **Discussion**

The aim of this retrospective study was to elucidate the PKU prevalence and distribution of causative *PAH* variants worldwide and in different countries. We provided a rough estimation of global PKU prevalence by calculating the number of affected patients for countries based on the provided PKU prevalence and the total population in 2018<sup>22</sup> and an average prevalence for countries not employing newborn screening for PKU. Overall, it seems that there are about 0.45 million PKU patients worldwide, of whom at least two thirds have PKU that requires treatment (most patients had a severe, cPKU phenotype). An improvement in early diagnosis via NBS in states without this should be an urgent priority.

Although PKU has been found to be most common in European populations,<sup>8</sup> its similar prevalence in certain Middle Eastern countries (particularly Turkey and Iran) was a remarkable finding.<sup>23,24</sup> A possible contributing factors may be a tendency to consanguinity in Arabic cultures, especially first cousin marriages, which favors the autosomal recessive inheritance of PKU.<sup>25</sup> Prevalence and marriage status was missing for many countries, making it more difficult to assess this proposed causal relationship for all states.<sup>26</sup>

We identified 758 different PKU alleles in this study, emphasizing the strong genetic heterogeneity of PKU. Information of about 10,000 genotypes and phenotypes from PKU patients in Europe enabled the a more detailed study of their distribution across different regions and countries than has been possible previously.<sup>27,28</sup> The severe p.R408W variant

was the most common overall, especially among Eastern European populations, in accordance with previous studies.<sup>29,30</sup> Furthermore, p.R408W dominates most Central European populations (Poland, Slovakia, Czechia, Austria, or Germany), also supporting previous studies.<sup>31-35</sup>

Previous research had suggested that some PKU variants appeared to have been carried by migration, e.g. p.R408W, IVS10-11G>A, IVS12+1G>A or p.F299C and R408Q.<sup>27</sup> Our study confirms and extends the knowledge on how certain variants appear to have spread within Europe and to different regions and countries. In particular, the distribution of PKU variants across Europe was consistent with successive waves of historical migration, and Figure 4 summarizes the likely geographical routes of transmission of these variants within Europe.

For example, Germany has occupied a crossroads location during several migration waves throughout history, and displays a broad spectrum of PAH variants, with similarities to some other populations, such as Northern European countries (Denmark, Sweden and Norway).<sup>35</sup> This implies a genetic connection of these regions, possibly during the Germanic settlement of Scandinavia. Norway's most frequent variant, p.G46S, did not occurring in the German group. Furthermore, the occurrence of IVS10-11G>A in Germany was surprisingly high in our analysis, which might be explained by immigration from Turkey. This variant, often described as the "Mediterranean mutation", may be of Italian origin.<sup>36</sup> and has been found mainly in Southern countries or the Middle-East.<sup>37,38</sup> Our study showed that IVS10-11G>A has the highest AF in Middle Eastern countries (Turkey, Iran and Israel), and in Portugal and Spain,

The p.R408W-associated genotype, the most common variant in people with Slavic roots, followed the east-central-southwest European axis, starting from Estonia, with the highest number of severe PKU patients, via Russia, Poland, Czechia, Slovakia, Germany, France, and Italy, down to Spain. The high prevalence of cPKU in Russia, and the dominance of the severe variant, p.R408W, were striking, given that Russia has a diverse ethnic composition, with >160 ethnic groups.<sup>39</sup>

Another well-documented axis exists for the p.Y414C-associated genotype, from northern to western Europe, i.e. from Sweden via Norway, Denmark and Germany to the Netherlands. Some variants, e.g., IVS12+1G>A, seem to have migrated within Scandinavian countries (Denmark >Sweden >Norway) at high frequency. The same was true for the p.E390G-associated genotype concentrating between Slovenia, Croatia, Serbia and Austria. A typical



421 east-west gradient can be seen for IVS10-11G>A, originating from Western Asia and the  
422 Middle-East, i.e., from Armenia (48%) via Turkey, Iran, Israel to Spain. Similar consequences  
423 of migration were evident for the very mild p.A300S-associated genotypes, migrating from  
424 Turkey via Israel, Italy to Spain. No trends for p.V245A and p.A403V distribution across  
425 Europe could be seen, however.

426 The influence of migration patterns during history can also be seen in the heterogenous  
427 spectrum of *PAH* variations in Latin America. These countries were strongly marked by  
428 immigration of Europeans, especially from southern Europe, during colonial times.<sup>40</sup> An  
429 example is p.V388M, which was especially prevalent in Brazil and Chile, once colonies of  
430 Portugal and Spain. The Argentina cohort was more prevalent for variations frequently  
431 occurring in eastern, central and northern Europe; e.g., as p.R408W, p.R261Q and IVS10-  
432 11G>A. In contrast to Brazil or Argentina, Mexico was characterized by a high percentage of  
433 indigenous population, which could be a reason for the exceptional high AF of the  
434 IVS1+5G>T splice variant there.<sup>41</sup>

435 The USA has one of the most multi-racial and ethnic populations in the world. The strongest  
436 ancestral influence is European, mainly from Germany, Ireland, England, Italy and France<sup>42</sup>,  
437 which explains a comparable distribution of PKU variants in the USA and Europe.<sup>43</sup> African  
438 Americans have a much lower PKU incidence than Caucasian Americans,<sup>44</sup> and observing  
439 changes in the growth of causative variants in USA, major ethnic groups there (Caucasian,  
440 African-American, Hispanic, Asian) grow at different rates would be an interesting area for  
441 future research.

442 While the overall PKU prevalence in China was 1:15,924, its distribution across the country  
443 varies significantly with higher rates in the north in comparison to the south.<sup>45</sup> In accordance  
444 with previous reports, p.R234Q accounts for 23% of all variants in China.<sup>46</sup> Three additional  
445 variants p.R243Q, Ex6-96A>G and p.R241C (AF>5%) were common in Korea, Taiwan as well,  
446 indicating commonalities of migration movements and evolution among those nationalities.  
447 Prevalent PKU variants for Europe, Middle-East, Latin America and the USA were uncommon  
448 in the Asian populations.

449 Preliminary data indicated a value of large patient databases with the genotype and  
450 phenotype information and introduction of the APV and GPV for the genotypic phenotype  
451 prediction.<sup>14</sup> This study extends this previous knowledge with additional information, and

has confirmed the power of genotyping in phenotype definition and BH<sub>4</sub> responsiveness in PKU, thus offering a powerful tool in the personalized medicine of this inherited metabolic disease.

This study also documents that functional mild variants (in a compound heterozygous constellation), with a substantial residual PAH activity, were always dominant over inactive severe variants (0-alleles) and gave rise to a milder phenotype and potential BH<sub>4</sub> responsiveness. Compared with severe classic PKU variants, they were much rarer in number, but determined the milder metabolic phenotype. Interestingly, homozygous mild variants have higher APV and thus milder phenotype than when in a compound heterozygous state with a 0-allele<sup>14</sup>. Two inactive severe variants are never BH<sub>4</sub>-responsive, as a rule, although anecdotal reports sometimes suggest this.<sup>47</sup>

In conclusion, this study provides an overview of the current distribution of the most important PAH variants and patients genotypes in various world regions. This information, together with the APV value, helps to predict the PKU phenotype and the possible treatment options. Despite immense progress in diagnosis and treatment of PKU in the last decade there are still too many areas of the world without adequate access to this care. Especially data from Africa, certain Asian and Southern American as well as Caribbean countries was missing.

## Description of Supplemental Data

**Supplementary Figure 1.** Flow-chart of the literature search.

**Supplementary Figure 2.** Physical structure of the *PAH* gene and variants tabulated in the locus-specific database (<http://www.biopku.org/pah/home.asp>). Regulatory domain: residues 1-142; catalytic domain: residues 143-410; oligomerization domain: residues 411-452. Cofactor (BH<sub>4</sub>)-binding regions (CBR): CBR1: residues 246-266; CBR2; residues 280-283; CBR3: residues 322-326; CBR4: residues 377-379. The reference accession number for the PAH sequence is ENSG00000171759; RefSeq NM\_000277.1.

**Supplementary Figure 3.** Heat maps of the distribution (%) of classic PKU (cPKU), mild PKU (mPKU) and mild hyperphenylalaninemia (MHP) A) in countries with at least 35 reported

cases and B) heat maps of overall genotypes severity, expressed as genotypic phenotype value (GPV), The lower the GPV, more severe the phenotype.

**Supplementary Figure 4.** Heat maps of the distribution (%) of classic PKU (cPKU), mild PKU (mPKU) and mild hyperphenylalaninemia (MHP) in **A)** different European regions and **C)** in 7 world regions. Heat maps of overall genotypes severity, expressed as genotypic phenotype value (GPV), in **B)** European regions and **D)** in 7 world regions. The lower the GPV, more severe the phenotype. F) Ten most common variants, their allele frequency (AF) and corresponding allelic phenotype values (APV) for the 3 PKU phenotype categories. The reference accession number for the *PAH* sequence is ENSG00000171759; RefSeq NM\_000277.1.

**Supplementary Table 1.** Prevalence of PKU (all phenotypes) in world regions and countries and estimated number of PKU patients.

**Supplementary Table 2.** PAH variations (n=758) found in the study cohort of PKU patients.

**Supplementary Table 3.** Genotypes (n=3659) found in the study cohort of PKU patients.

**Supplementary Table 4.** Distribution of BH<sub>4</sub> responsiveness in different phenotype groups. Those classified as “slow responders” are included in the non-responsive group and descriptive analysis of maximal pretreatment blood Phe levels and BH<sub>4</sub> responsiveness.

**Supplementary Table 5.** Definition of countries in the world regions.

Access to Supplementary files:

<https://drive.google.com/drive/folders/15DI4NQfczgjgojGizRj19colMrjvtPW0?usp=sharing>

## Declaration of interests

Authors: please enter any COI!

## Acknowledgements

We are indebted to Suha Daas, who works at the National Newborn screening Program, Ministry of Health, Tel Hashomer Israel for her help and analyses. Farrah Rajabi, Fran Rohr, Ann Wessel, Leslie Martell, Steven Dobrowolski, Per Guldberg, and the late Flemming Guttler were greatly responsible for the data from New England. This work was funded in part by the Fundación Isabel Gemio-Fundación La Caixa (LCF/PR/PR16/11110018), and the Regional Government of Madrid (CAM, B2017/BMD3721). The research conducted at the Murdoch Children's Research Institute was supported by the Victorian Government's Operational Infrastructure Support Program.

## Web Resources

BIOPKU: <http://www.biopku.org/biopku/home.asp>

ClinVar: <https://www.ncbi.nlm.nih.gov/clinvar/>

HGMD: <http://www.hgmd.cf.ac.uk/ac/index.php>

LOVD: <https://www.lovd.nl/>

HGVS nomenclature recommendations: <http://www.hgvs.org/content/guidelines>

PAHvdb: <http://www.biopku.org/pah/home.asp>

## References

1. Blau N, Van Spronsen FJ, Levy HL. Phenylketonuria. *Lancet* 2010;376:1417-27.
2. Flydal MI, Martinez A. Phenylalanine hydroxylase: function, structure, and regulation. *IUBMB life* 2013;65:341-9.
3. Himmelreich N, Shen N, Okun JG, Thiel C, Hoffmann GF, Blau N. Relationship between genotype, phenylalanine hydroxylase expression and in vitro activity and metabolic phenotype in phenylketonuria. *Mol Genet Metab* 2018;125:86-95.
4. Camp KM, Parisi MA, Acosta PB, et al. Phenylketonuria scientific review conference: State of the science and future research needs. *Mol Genet Metab* 2014;112:87-122.
5. van Spronsen FJ, van Wegberg AM, Ahring K, et al. Key European guidelines for the diagnosis and management of patients with phenylketonuria. *Lancet Diabetes Endocrinol* 2017;5:743-56.
6. Vockley J, Andersson HC, Antshel KM, et al. Phenylalanine hydroxylase deficiency: diagnosis and management guideline. *Genet Med* 2014;16:188-200.
7. Blau N, Longo N. Alternative therapies to address the unmet medical needs of patients with phenylketonuria. *Expert Opin Pharmacother* 2015;16:1-10.
8. Scriver CR. The PAH gene, phenylketonuria, and a paradigm shift. *Hum Mutat* 2007;28:831-45.
9. Gundorova P, Zinchenko RA, Kuznetsova IA, Bliznetz EA, Stepanova AA, Polyakov AV. Molecular-genetic causes for the high frequency of phenylketonuria in the population from the North Caucasus. *PLoS One* 2018;13:e0201489.
10. Guldborg P, Henriksen KF, Sipila I, Guttler F, de la Chapelle A. Phenylketonuria in a low incidence population: molecular characterisation of mutations in Finland. *J Med Genet* 1995;32:976-8.
11. Okano Y, Kudo S, Nishi Y, Sakaguchi T, Aso K. Molecular characterization of phenylketonuria and tetrahydrobiopterin-responsive phenylalanine hydroxylase deficiency in Japan. *Journal of human genetics* 2011;56:306-12.
12. Güttler F, Guldborg P, Henriksen KF, Mikkelsen I, Olsen B, Lou H. Molecular basis for the phenotypical diversity of phenylketonuria and related hyperphenylalaninaemias. *J Inher Metab Dis* 1993;16:602-4.
13. Blau N, Shen N, Carducci C. Molecular genetics and diagnosis of phenylketonuria: state of the art. *Expert review of molecular diagnostics* 2014;14:655-71.
14. Garbade SF, Shen N, Himmelreich N, et al. Allelic phenotype values: a model for genotype-based phenotype prediction in phenylketonuria. *Genet Med* 2019;21:580-90.

15. Zschocke J. Dominant versus recessive: molecular mechanisms in metabolic disease. *J Inherit Metab Dis* 2008;31:599-618.
16. Scriver CR, Waters PJ. Monogenic traits are not simple: lessons from phenylketonuria. *Trends Genet* 1999;15:267-72.
17. Shen N, Heintz C, Thiel C, Okun JG, Hoffmann GF, Blau N. Co-expression of phenylalanine hydroxylase variants and effects of interallelic complementation on in vitro enzyme activity and genotype-phenotype correlation. *Mol Genet Metab* 2015;117:328–35.
18. Evers RAFvW, Annemiek M J ; Anjema, Karen ; Lubout, Charlotte M A ; van Dam, Esther ; van Vliet, Danique ; Blau, Nenad ; van Spronsen, Francjan J. . The first European guidelines on phenylketonuria: its usefulness and implications for BH4 responsiveness testing. *Journal of inherited metabolic disease* 2019.
19. Muntau AC, Adams DJ, Belanger-Quintana A, et al. International best practice for the evaluation of responsiveness to sapropterin dihydrochloride in patients with phenylketonuria. *Mol Genet Metab* 2019;127:1-11.
20. Konecki DS, Schlotter M, Trefz FK, Lichter-Konecki U. The identification of two mis-sense mutations at the PAH gene locus in a Turkish patient with phenylketonuria. *Hum Genet* 1991;87:389-93.
21. Ellingsen S, Knappskog PM, Eiken HG. Phenylketonuria splice mutation (EXON6nt-96A-->g) masquerading as missense mutation (Y204C). *Hum Mutat* 1997;9:88-90.
22. United Nations -Population division DoEaSA. World population prospects, File Total population (both sexes combined) by region, subregion and country. 2019.
23. Matalon R, Michals K. Phenylketonuria: screening, treatment and maternal PKU. *Clin Biochem* 1991;24:337-42.
24. El-Metwally A, Al-Ahaidib LY, Sunqurah AA, et al. The prevalence of phenylketonuria in Arab countries, Turkey, and Iran: A systematic review. *BioMed Res Internat* 2018;doi.org/10.1155/2018/7697210.
25. Shoraka HR, Haghdoost AA, Baneshi MR, Bagherinezhad Z, Zolala F. Global prevalence of classic phenylketonuria based on neonatal screening programs: Systematic review and meta-analysis. *Clin Exp Pediatr* 2020;63:34-43.
26. Saadallah AA, Rashed MS. Newborn screening: experiences in the Middle East and North Africa. *J Inherit Metab Dis* 2007;30:482-9.
27. Zschocke J. Phenylketonuria mutations in Europe. *Hum Mutat* 2003;21:345-56.

- 591 28. Danecka MK, Woidy M, Zschocke J, Feillet F, Muntau AC, Gersting SW. Mapping the  
592 functional landscape of frequent phenylalanine hydroxylase (PAH) genotypes promotes personalised  
593 medicine in phenylketonuria. *J Med Genet* 2015;52:175-85.
- 594 29. Gundorova P, Stepanova AA, Kuznetsova IA, Kutsev SI, Polyakov AV. Genotypes of 2579  
595 patients with phenylketonuria reveal a high rate of BH4 non-responders in Russia. *PLoS One*  
596 2019;14:e0211048.
- 597 30. Lillevali H, Reinson K, Muru K, et al. Hyperphenylalaninaemias in Estonia: Genotype-  
598 Phenotype Correlation and Comparative Overview of the Patient Cohort Before and After Nation-  
599 Wide Neonatal Screening. *JIMD Rep* 2017;10.1007/8904\_2017\_61.
- 600 31. Bik-Multanowski M, Kaluzny L, Mozrzymas R, et al. Molecular genetics of PKU in Poland and  
601 potential impact of mutations on BH4 responsiveness. *Acta Biochim Pol* 2013;60:613-6.
- 602 32. Kadasi L, Polakova H, Ferakova E, et al. PKU in Slovakia: mutation screening and haplotype  
603 analysis. *Hum Genet* 1995;95:112-4.
- 604 33. Reblova K, Hrubá Z, Prochazkova D, Pazdirkova R, Pouchla S, Fajkusova L.  
605 Hyperphenylalaninemia in the Czech Republic: Genotype-phenotype correlations and in silico  
606 analysis of novel missense mutations. *Clinica chimica acta; international journal of clinical chemistry*  
607 2013;419:1-10.
- 608 34. Sterl E, Paul K, Paschke E, et al. Prevalence of tetrahydrobiopterine (BH4)-responsive alleles  
609 among Austrian patients with PAH deficiency: comprehensive results from molecular analysis in 147  
610 patients. *Journal of inherited metabolic disease* 2013;36:7-13.
- 611 35. Zschocke J, Hoffmann GF. Phenylketonuria mutations in Germany. *Hum Genet* 1999;104:390-  
612 8.
- 613 36. Eisensmith RC, Okano Y, Dasovich M, et al. Multiple origins for phenylketonuria in Europe.  
614 *Am J Hum Genet* 1992;51:1355-65.
- 615 37. Okano Y, Wang T, Eisensmith RC, et al. Phenylketonuria missense mutations in the  
616 Mediterranean. *Genomics* 1991;9:96-103.
- 617 38. Kleiman S, Avigad S, Vanagaite L, et al. Origins of hyperphenylalaninemia in Israel. *Eur J Hum*  
618 *Genet* 1994;2:24-34.
- 619 39. Statistics FOo. "On the results of the All-Russia Population Census 2010".
- 620 40. Desviat LR, Perez B, De Lucca M, Cornejo V, Schmidt B, Ugarte M. Evidence in Latin America  
621 of recurrence of V388M, a phenylketonuria mutation with high in vitro residual activity. *Am J Hum*  
622 *Genet* 1995;57:337-42.
- 623 41. Vela-Amieva M, Abreu-Gonzalez M, Gonzalez-Del Angel A, et al. Phenylalanine hydroxylase  
624 deficiency in Mexico: genotype-phenotype correlations, BH4 responsiveness and evidence of a  
625 founder effect. *Clin Genet* 2015;88:62-7.

- 626 42. Ramirez KRHNAJRR. Overview of Race and Hispanic Origin: 2010. United States Census  
627 Bureau 2011.
- 628 43. Rajabi F, Rohr F, Wessel A, et al. Phenylalanine hydroxylase genotype-phenotype associations  
629 in the United States: a single center study. Mol Genet Med 2019;submitted.
- 630 44. Hofman KJ, Steel G, Kazazian HH, Valle D. Phenylketonuria in U.S. blacks: molecular analysis  
631 of the phenylalanine hydroxylase gene. American journal of human genetics 1991;48:791-8.
- 632 45. Xiang L, Tao J, Deng K, et al. Phenylketonuria incidence in China between 2013 and 2017  
633 based on data from the Chinese newborn screening information system: a descriptive study. BMJ  
634 Open 2019;9:e031474.
- 635 46. Li N, Jia H, Liu Z, et al. Molecular characterisation of phenylketonuria in a Chinese mainland  
636 population using next-generation sequencing. Sci Rep 2015;5:15769.
- 637 47. Hennermann JB, Vetter B, Kulozik AE, Mönch E. Partial und total tetrahydrobiopterin-  
638 responsiveness in classical and mild phenylketonuria (PKU). J Inherit Metab Dis 2002;25:21.



## Legends to figures

**Figure 1.** Prevalence (A: 1:<30,000; B: 1:>30,000 live births) of PKU (all phenotypes) in countries from 6 world regions.

**Figure 2. A:** Boxplot (median, 25<sup>th</sup>-75<sup>th</sup> percentile, 1.5) of the maximal pretreatment blood Phe levels for three metabolic phenotypes in 6,796 PKU patients. The circles in the cPKU bar represent ordinary high blood Phe levels, since cPKU doesn't have an upper Phe limit for its classification; **B:** Contour plot of two-dimensional densities of pretreatment blood Phe levels and corresponding genotypic phenotype values (GPV) for 6,796 PKU patients; **C:** Boxplot (median, 25<sup>th</sup>-75<sup>th</sup> percentile, 1.5) of GPV in 2,246 BH<sub>4</sub> non-responder and 1,755 responder PKU patients; **D:** Boxplot (median, 25<sup>th</sup>-75<sup>th</sup> percentile, 1.5) of GPV (APV<sub>max</sub>) for 11,584 PKU patients with a known genotype, but not tested for BH<sub>4</sub> responsiveness. Horizontal grey bar: separation area between GPVs for predicted BH<sub>4</sub> responsiveness (3.8-10) and non-responsiveness (0-3.3).

**Figure 3.** The world map with the relative frequency (%) of PKU and the corresponding most common variants for classic PKU (cPKU), mild PKU (mPKU) and mild hyperphenylalaninemia (MHP). AF: allele frequency; APV: allelic phenotype value (cPKU=0-2.6; mPKU=2.7-6.6; MHP=6.7-10).

**Figure 4.** Likely routes of transmission of common *PAH* variants associated with migration in Europe.