


RESEARCH

Open Access



Identification of deep intronic variants of *PAH* in phenylketonuria using full-length gene sequencing

Chuan Zhang^{1,2,3†}, Yousheng Yan^{4†}, Bingbo Zhou¹, Yupei Wang¹, Xinyuan Tian¹, Shengju Hao¹, Panpan Ma¹, Lei Zheng¹, Qinghua Zhang¹, Ling Hui¹, Yan Wang¹, Zongfu Cao^{2*}  and Xu Ma^{2,3*}

Abstract

Background Phenylketonuria (PKU) is an autosomal recessive congenital metabolic disorder caused by *PAH* variants. Previously, approximately 5% of PKU patients remained undiagnosed after Sanger sequencing and multiplex ligation-dependent probe amplification. To date, increasing numbers of pathogenic deep intronic variants have been reported in more than 100 disease-associated genes.

Methods In this study, we performed full-length sequencing of *PAH* to investigate the deep intronic variants in *PAH* of PKU patients without definite genetic diagnosis.

Results We identified five deep intronic variants (c.1199+502A>T, c.1065+241C>A, c.706+368T>C, c.706+531>C, and c.706+608A>C). Of these, the c.1199+502A>T variant was found at high frequency and may be a hotspot *PAH* variant in Chinese PKU. c.706+531T>C and c.706+608A>C are two novel variants that extend the deep intronic variant spectrum of *PAH*.

Conclusion Deep intronic variant pathogenicity analysis can further improve the genetic diagnosis of PKU patients. In silico prediction and minigene analysis are powerful approaches for studying the functions and effects of deep intronic variants. Targeted sequencing after full-length gene amplification is an economical and effective tool for the detection of deep intron variation in genes with small fragments.

Keywords PKU, *PAH*, Deep intronic variant, RNA splicing, Minigene

Introduction

Hyperphenylalaninemia (HPA) is the most common hereditary disorder of amino acid metabolism worldwide. The main type of HPA is phenylketonuria (PKU), an autosomal recessive disease caused by variants of *PAH*, the gene encoding phenylalanine hydroxylase [24]. As of November 2, 2022, the PAHvdb database (<http://www.biopku.org/home/pah.asp>) has collected 1583 types of *PAH* variants, including missense, frameshift, synonymous, and splicing variants, UTR variation, and large-scale deletion. These variations are mainly concentrated in exons and exon–intron boundaries. Existing molecular detection technology can accurately detect the

[†]Chuan Zhang and Yousheng Yan contributed equally to this work.

*Correspondence:

Zongfu Cao
zongfu_cao@163.com

Xu Ma
maxubioinfo@163.com

¹ Gansu Province Medical Genetics Center, Gansu Provincial Maternity and Child-Care Hospital, Lanzhou, China

² National Research Institute for Health and Family Planning, National Human Genetic Resources Center, Beijing, China

³ Graduate School of Peking, Union Medical College, Beijing, China

⁴ Prenatal Diagnostic Center, Beijing Obstetrics and Gynecology Hospital, Capital Medical University; Beijing Maternal and Child Health Care Hospital, Beijing, China



above *PAH* variants; however, in different regions of the world, regardless of the detection technology applied, the genetic diagnosis rate of PKU by researchers is 70.6–96%, which fails to reach 100% [3, 4, 8, 14–16, 20, 22, 27, 28, 30]. Therefore, sequence information limited to exons and exon–intron boundaries cannot identify the entire genetic information of PKU.

Human protein-coding genes consist, on average, of short coding fragments containing 8–10 exons that are interrupted by noncoding sequences or introns that are approximately 20 times longer [27]. Introns have been critical for the evolution of eukaryotes. The intron–exon structure of eukaryotic genes plays an important role in the generation of new genes through exon shuffling [5, 19], and the ability to alternately select different exon combinations is crucial to the gene expression diversity of complex organisms [13]. With the clinical application of Whole Genome Sequencing (WGS) technology, more and more pathogenic deep intron variants (more than 100 bp from the exon–intron boundary) have been discovered. Variants in deep intronic regions exist in a variety of diseases, and hundreds of studies on the pathogenesis of deep introns have been published. Deep intron mutations are located at least 100 bp away from the nearest typical splice site [27]). Therefore, the pathogenicity analysis of deep intron variants of *PAH* may be a new strategy to improve the diagnostic rate of PKU when only one pathogenic variant is found in exons and flanking sequences of *PAH*.

We previously identified three *PAH* deep intron variants (c.706+368T>C, c.1065+241C>A, and c.1199+502A>T) in ten patients with PKU in northwest China using WGS [12]. Later, Gao et al. [10] also found deep intronic *PAH* variants in patients with PKU without definite genetic diagnosis, which indicated that the study of deep intronic variation of *PAH* can improve the PKU genetic diagnosis rate. Therefore, in this study, we performed full-length sequencing of *PAH* for PKU patients with unknown genotypes to investigate whether they were carrying pathogenic deep intronic *PAH* variants, and to expand the pathogenic variation spectrum of deep intronic variation of *PAH*.

Methods

DNA samples

Based on clinical features and newborn screening, 967 cases of PKU were diagnosed in the Medical Genetics Center of Gansu Provincial Maternity and Child-Care Hospital between January 2012 and December 2021. From these patients, we selected 45 patients with only one or no variants detected in *PAH*, and screened for *PAH* variants by Sanger sequencing, whole exon sequencing, and MLPA on diagnosis (Fig. 1). Of these 45

patients, according to the standard classification criteria [29], 11 were classic PKU (cPKU, ≥ 1200 $\mu\text{mol/L}$), 17 were mild PKU (mPKU, 360–1200 $\mu\text{mol/L}$), and 17 were mild hyperphenylalaninemia (MHP, 120–360 $\mu\text{mol/L}$). This study was undertaken according to the tenets of the Declaration of Helsinki 1975 and its later amendments. The study protocol was approved by the Ethics Committee of the Gansu Provincial Maternity and Child-Care Hospital (2021GSFY[65]). Written informed consent was obtained from all study participants or their legal guardians.

Genomic DNA preparation

Genomic DNA was extracted from peripheral blood samples (2–3 mL) of the probands and their parents using the Tiangen DNA extraction kit (Tiangen Biotech, China). DNA quality was quantified with a NanoDrop 2000 (Thermo, USA).

Full-length sequencing of *PAH*

We designed 13 pairs of primers to cover the whole *PAH* gene using Primer3 online software (v.0.4.0, <https://bioinformatics.ut.ee/primer3-0.4.0/>), the primers and fragment length of Full-length amplification of *PAH* were showed in Table 1. The PCR-amplified fragments were tested using 1% agarose gel electrophoresis. After successful amplification, the products were mixed and sent to Beijing Nuo Zhiyuan Technology Co., Ltd for targeted sequencing. The library for genome sequencing was generated from *PAH* full-length amplified products using the Illumina TruSeq DNA PCR-Free Library Prep Kit (Illumina). Sequencing was performed on the Illumina HiSeq 2500 System, the sequencing range covered the entire *PAH* gene, and the sequencing depth was 10,000X. Data analysis and variant curation were performed using Pgenomics software (<https://www.pgenomics.cn/>). Single-nucleotide variants and small insertions and deletions were identified using MedGAP v.2.0, a pipeline based on GATK best practices for data preprocessing and variant discovery with the GATK Haplotype Caller (<https://gatk.broadinstitute.org/hc/en-us, v.3.1.1>) [12].

Variants were described according to the nomenclature recommended by the Human Genome Variation Society (www.hgvs.org/). Variant frequencies were searched in the GnomAD (<http://gnomad.broadinstitute.org/>), Exome Sequencing Project (ESP, <http://evs.gs.washington.edu>), and dbSNP (<http://www.ncbi.nlm.nih.gov/projects/snp>) databases. Suspicious variants were verified by Sanger sequencing, and the PCR products were bidirectionally sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA) on an ABI 3500DX Genetic Analyzer (Applied Biosystems) after purification on 2% agarose gels.

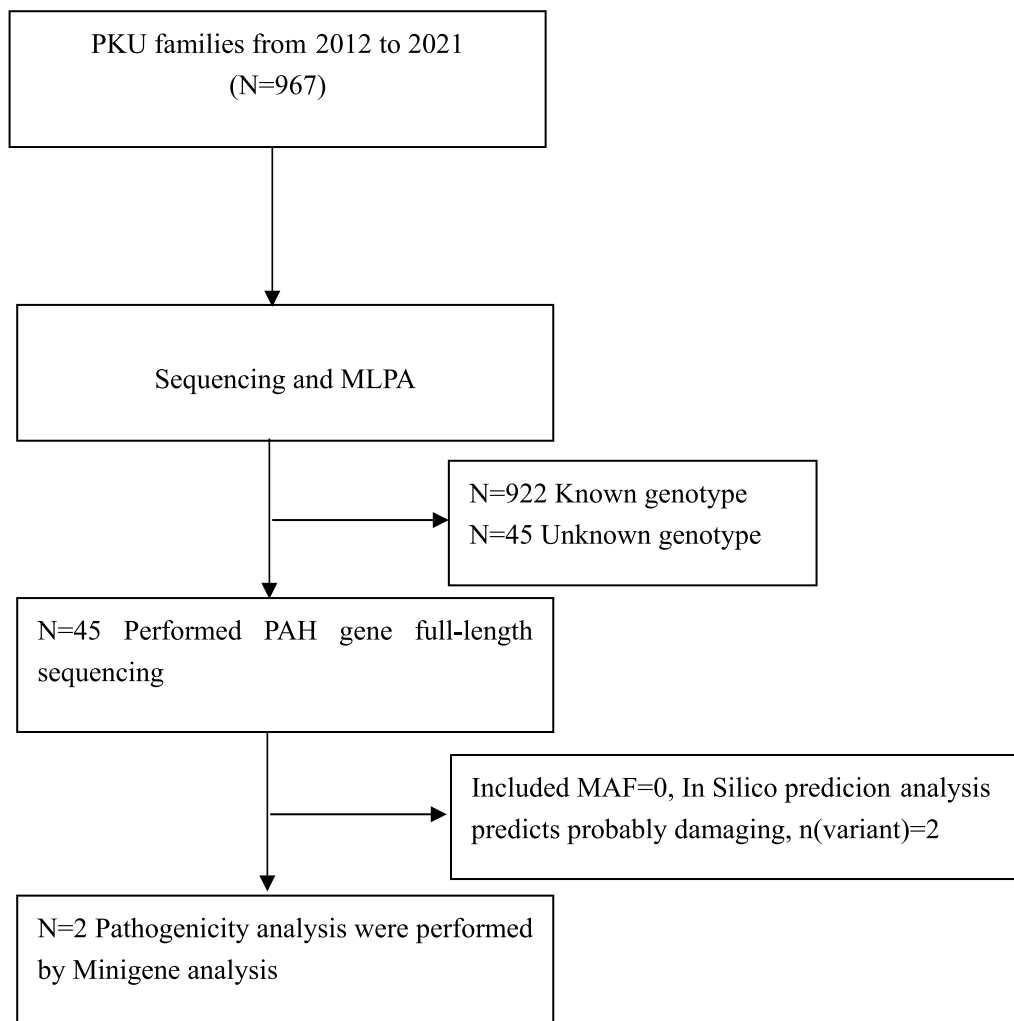


Fig. 1 Flow diagram of the recruitment and exclusion process for the PKU patients

In silico prediction of deep intronic variants

Alamat Visual v.2.11 (Interactive Biosoftware) software was used to predict the influence of variation sites on splicing site selection. The process included a splicing module that integrates a number of prediction algorithms and splicing prediction data using Splice Site Finder-like (SSFL), MaxEntScan (MES), NNSplice, and GeneSplicer to analyze splicing signals. Exonic splicing enhancer (ESE) finder 3.0 and RESCUE-ESE were used for ESE binding site prediction and for high confidence branch point prediction [12]. Individual tools were deemed to predict altered splicing where the change in the splice site score was $\geq 10\%$ (based on MES and GeneSplicer) or $\geq 5\%$ (based on NNSplice and SSFL). SVM-BPfinder (http://regulatorygenomics.upf.edu/Software/SVM_BP/) and RNABP: Branch Point Selection in RNA Splicing Using Deep Learning (<http://nscbio.jbnu.ac.kr/tools/RNABP/>) were used to predict the RNA splicing branch points.

Minigene analysis

To evaluate the in vitro splicing of two deep intronic variants (c.706+531T>C and c.706+608A>C) in *PAH*, we constructed minigenes using the pMini-CopGFP vector (Invitrogen). In the minigene pMini-CopGFP (+) vector, *_In6* for c.706+531T>C and c.706+608A>C (a fragment of the human *PAH* gene including full-length exon 6, intron 6, and exon 7) were amplified from the patient's genomic DNA using primers located in introns 5 and 7. The gene fragments and their flanking regions were cloned into pMini-CopGFP (+) using BamHI/XhoI restriction endonucleases. The mutant and wildtype minigene constructs were thus prepared. The positive cloning screening/identification primers for the two sites were the same: β -globin intron-F, 5'-GATATACACTGTTTGGAGATGAGGA-3'; PAH-E6-E7-R, 5'-TAGATATGCTACTAATCCCC-3'.

Table 1 The primers and fragment length of Full-length amplification of *PAH*

Fragment	Primers	Primer sequence 5'–3'	Genomic locus	Fragment length
1	PAH-L1F	ACTCTCTTCTCCTCCTAGTGCG	chr12:103311000	11,916 bp
	PAH-L1R	CCAAATAGCTCCCTGATTACCC	chr12:103299084	
2	PAH-L2F	AAGGTAGACAAGGTGGTAGGACT	chr12:103299364	3975 bp
	PAH-L2R	GAGAGCACACTTCAAAAAGAAGG	chr12:103295389	
3	PAH-L3F	TGCTTTTCCCACTGTACTACTCC	chr12:103295660	4029 bp
	PAH-L3R	GGCTGGATGACTCAAGAGTTTTA	chr12:103291631	
4	PAH-L4F	ACGAATGTGGGAGTGGGATGCTT	chr12:103291906	3144 bp
	PAH-L4R	ACCAGAACAGGAAAACCTAACGC	chr12:103288762	
5	PAH-L5F	GTCTGACCCCTATTCAAGCAGA	chr12:103289098	11,693 bp
	PAH-L5R	ACCTCTGAGCTCTGCACCTTGTC	chr12:103277405	
6	PAH-L6F	CCCATCAACCCCTGAAGGACGT	chr12:103277697	10,531 bp
	PAH-L6R	TGAAGGAGGTACAGCCATTGGT	chr12:103267166	
7	PAH-L7F	TCGTGAGTGGTAGTTTTCCATGG	chr12:103267491	5388 bp
	PAH-L7R	AAGGAAGGGAGGGAAGATAGGAG	chr12:103262103	
8	PAH-L8F	TAGAAACGAGGCACAACAGTAGT	chr12:103262383	2337 bp
	PAH-L8R	AGGCTGTTTTATTTCAGGACCGAG	chr12:103,260,046	
9	PAH-L9F	CCTGTGTACCGTCAAGACGGAA	chr12:103260407	8101 bp
	PAH-L9R	TCTACCACATTGCACTCATTCC	chr12:103,252,306	
10	PAH-L10F	CCTTACACACTGCCTTAAATGT	chr12:103252590	4048 bp
	PAH-L10R	ACACAACACACACTCCTAACTC	chr12:103248542	
11	PAH-L11F	GGAAAGACAAGTATGTGGAGCAA	chr12:103,248,822	4049 bp
	PAH-L11R	GATTGTTTGAGCCAGAAAGTCTG	chr12:103244773	
12	PAH-L12F	CCTTGTGAGGCAGTTATTTGTGT	chr12:103245076	3913 bp
	PAH-L12R	ATAGCTGGAGAGTCTAACACATC	chr12:103,241,163	
13	PAH-L13F	CCCTGGAGCTCTAGTCCCTCTGT	chr12:103241472	10,642 bp
	PAH-L13R	AGCAAGATCATCTGTGTAAGACTG	chr12:103230830	

For the minigene assays[12], 293T cells were seeded in 35 cm² wells (density, 2–3 × 10⁵) in 2 mL of 10% minimum essential medium (MEM) and then grown overnight. Cells were transfected with DNA (4 µg per well) using Lipofectamine 3000 Transfection Reagent (Thermo Fisher). Cells were harvested by trypsinization after 48 h. Total RNA was isolated using TRIzol Reagent (Thermo Fisher) and phenol–chloroform extraction. Complementary DNA (cDNA) synthesis was performed using the HiScript II 1st Strand cDNA Synthesis Kit (+gDNA wiper) (Vazyme). Splicing analysis of the two intronic variants was performed by PCR amplification with FastStart Taq Polymerase (Vazyme) using the specific primers F: 5'-GGCTAACTAGAGAACCCACTGCTTA-3'; and R: 5'-GGTTCGGGGTATACATGGGCTT-3' for the pMini-CopGFP (+) minigene. PCR products were confirmed by Sanger sequencing after purification on 2% agarose gels.

Results

Screening for deep intronic variants

Of the 967 PKU patients, only 45 patients showed a heterozygous variant in the *PAH* gene after exon–intron junction detection of the *PAH* gene. All the 45 patients got successfully full-length sequencing, sequencing quality shows an average sequencing depth of 9019× (ranged 3731× to 15,097×). After full-length sequencing of *PAH*, 24 of 45 PKU patients showed pathogenic and suspected pathogenic intronic variants, including 11 cPKU, 11 mPKU, and two MHP (Table 2). No suspected disease-causing variants had been identified for six mPKU and 15 MHP patients (Table 2). Among the 24 PKU patients with definite genotyping, 18 patients had the c.1199+502A>T variant, three patients had c.1065+241C>A, one patient had c.706+368T>C, one patient carried c.706+531T>C, and one carried c.706+608A>C. The variants c.1199+502A>T,

Table 2 Genotypes of 45 undiagnosis PKU families

ID	Classification	Paternal variant	Maternal variant
1	cPKU	c.728G>A	c.1199+502A>T
2	cPKU	c.1197A>T	c.706+368T>C
3	cPKU	c.1199+502A>T	c.728G>A
4	cPKU	c.1199+502A>T	c.728G>A
5	cPKU	c.1301C>A	c.1065+241C>A
6	cPKU	c.1199+502A>T	c.1238G>C
7	cPKU	c.1238G>C	c.1199+502A>T
8	cPKU	c.1199G>A	c.1199+502A>T
9	cPKU	c.1199+502A>T	c.728G>A
10	cPKU	c.1199+502A>T	c.782G>A
11	cPKU	c.526C>T	c.1199+502A>T
12	mPKU	c.1199+502A>T	c.1199G>A
13	mPKU	c.208_210delTCT	c.1065+241C>A
14	mPKU	c.1199+502A>T	c.694C>T
15	mPKU	c.842+2T>A	c.1199+502A>T
16	mPKU	c.1068C>A	c.1199+502A>T
17	mPKU	c.1199+502A>T	c.194T>C
18	mPKU	c.1199+502A>T	c.331C>T
19	mPKU	c.87C>A	c.1199+502A>T
20	mPKU	c.1065+241C>A	c.728G>A
21	mPKU	EX1-2del	c.706+531T>C
22	mPKU	c.706+608A>C	c.694C>T
23	mPKU	?	c.782G>A
24	mPKU	c.1301C>A	?
25	mPKU	c.782G>A	?
26	mPKU	c.1289T>C	?
27	mPKU	?	c.473G>A
28	mPKU	c.1301C>A	?
29	MHP	c.1199+502A>T	c.158G>A
30	MHP	c.532G>A	c.1199+502A>T
31	MHP	Exon3del	?
32	MHP	EX1-Updel	?
33	MHP	c.1256A>G	?
34	MHP	?	c.770G>T
35	MHP	c.1114A>T	?
36	MHP	c.611A>G	?
37	MHP	c.728G>A	?
38	MHP	c.194T>C	?
39	MHP	?	c.1200-1G>C
40	MHP	?	c.1068C>A
41	MHP	?	c.1068C>A
42	MHP	c.1315+6T>A	?
43	MHP	c.688G>A	?
44	MHP	?	c.782G>A
45	MHP	c.208_210delTCT	?

c.1065+241C>A, and c.706+368T>C were classified as likely pathogenic or pathogenic in our previous study [16], and c.706+531T>C and c.706+608A>C were two novel variants that had not previously been reported.

In silico prediction analysis

The novel deep intronic variant c.706+531T>C was identified in a patient with cPKU, where this variant formed a compound heterozygous mutation with EX1-2del in the proband. The novel deep intronic variant c.706+608A>C was also identified in a patient with cPKU, this variant formed a compound heterozygous mutation with c.694C>T in the proband. Both variants were absent in the ClinVar (www.ncbi.nlm.nih.gov/clinvar/), HGMD (www.hgmd.cf.ac.uk/), and gnomAD (<http://gnomad.broadinstitute.org/>) databases.

In silico analysis by Alamut Visual predicted that the two variants probably impact splice site selection (Table 3 and Fig. 2). Using ESEfinder, c.706+531T>C was predicted to create an SRSF1 (SF2/ASF(IgM- BRCA1)) binding site and c.706+531T>C was predicted to strengthen the SRSF2 (SC35) binding site.

Minigene analysis

The RT-PCR results of the c.706+531T>C variant showed that there were two amplified bands in the WT (wild type) group and only one in MT (mutant type), among which only 495 bp of the expected sequence was amplified in the WT group. In addition, both groups had 518 bp bands, but the WT group had relatively weak bands (Fig. 3). Sanger sequencing showed that the WT-A band was consistent with the expected transcript sequence and the normal splicing mode, and was the main product of the WT group (Fig. 3A). The WT-B and MT bands resulted from the inclusion of a 114 bp intronic sequence (pseudo-exon) in intron 6. The splicing mode of the mRNA was NM_000277.3: c.706+534_706+647ins 114 bp (Fig. 3A). Therefore, it was classified as a likely pathogenic (PS1, PM3, PP3, and PP4) variant.

The RT-PCR results of the c.706+608A>C variant showed that there were two amplified bands in the WT group and four amplified bands in the MT group, where both 404 bp and 518 bp fragments were amplified in both the WT and MT groups. In addition, the WT group had a relatively weak 518 bp band while the MT group had a relatively weak 404 bp band (Fig. 3B). The MT group also had two relatively weak bands (MT-C and MT-D, Fig. 3B). Sanger sequencing showed that the WT-A band was consistent with the expected transcript sequence and the normal splice mode, and was the main product of the WT group (Fig. 3B). The WT-B and MT-B bands resulted from the inclusion of a 114 bp intronic sequence

Table 3 In silico prediction scores for the splice sites of two deep intronic variants assessed in this study by Alamut® Visual

Variant	Position	3'/5'	SSFL (0–100)		Max EntScan (0–12)		NNSPLICE (0–1)		GeneSplicer (0–21)		Branch points (0–100)	
			WT	Mut	WT	Mut	WT	Mut	WT	Mut	WT	Mut
c.706+531T>C	c.706+534	3'	–	–	7.6	8.1	1.0	1.0	–	–	–	–
	c.706+530	3'	–	–	–	–	–	–	–	–	67.7	70.9
	c.706+532	3'	–	–	–	–	–	–	–	–	52.7	59.3
c.706+608A>C	c.706+601	5'	–	–	–	–	0.7	0.8	–	–	–	–
	c.706+613	3'	–	–	–	–	–	–	–	–	66.5	69.7
	c.706+614	3'	–	–	–	–	–	–	–	–	46.3	48.8

The higher the score, the higher the credibility, “–” no predictive splicing sites



Fig. 2 In silico prediction of deep intronic variants. **A** ESEfinder predicted that the c.706+531T>C might create an SRSF1 (SF2/ASF(IgM-RRCA1)) binding sites; **B** ESEfinder predicted that the c.706+608A>C variant may strengthen the SRSF2 (SC35) binding sites

(pseudo-exon) in intron 6. The mRNA splicing mode was NM_000277.3: c.706+534_706+647ins 114 bp (Fig. 3B). The MT-C band resulted from the inclusion of

a 313 bp intronic sequence (pseudo-exon) in intron 6, and the splicing mode of the mRNA was NM_000277.3: c.706+335_706+647ins 313 bp (Fig. 3B). The MT-D

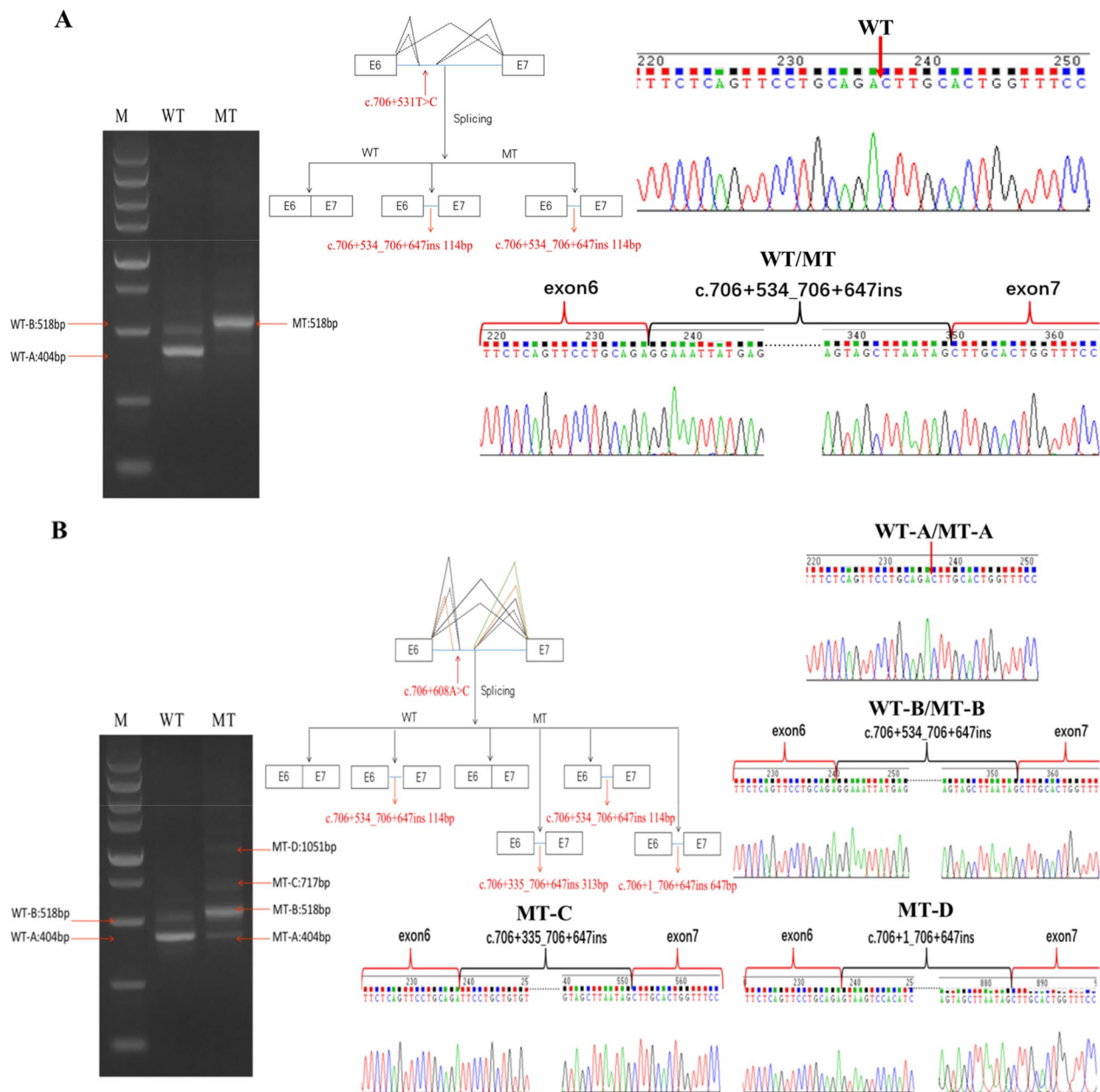


Fig. 3 Variant splice effects seen by minigene RT-PCR analysis. **A** The c.706+531T>C wild type minigene produced a major fragment (WT-A) and one low abundance fragments (WT-B), mutant minigene only produced a fragment (MT). Sanger sequencing showed that WT-B and MT include a 114-nt pseudo-exon upstream of exon 7. **B** The c.706+608A>C wild type minigene produced a major fragment (WT-A) and one low abundance fragments (WT-B), mutant minigene produced a major fragment (MT-B) and three low abundance fragments (MT-A, MT-C, MT-D). Sanger sequencing showed that WT-B and MT-B include a 114-nt pseudo-exon, MT-C includes a 313-nt pseudo-exon and MT-D includes a 647-nt pseudo-exon upstream of exon 7. RT-PCR, reverse transcription polymerase chain reaction

band resulted from the inclusion of a 647 bp intronic sequence (pseudo-exon) in intron 6, and the mRNA splicing mode was NM_000277.3: c.706+1_706+647ins 647 bp (Fig. 3B). Therefore, the pathogenicity of c.706+608A>C was classified as likely pathogenic (PS1, PM3, PP3, and PP4).

Discussion

PKU is an autosomal recessive genetic disease, which is mainly caused by variation of the *PAH* gene, encoding phenylalanine hydroxylase. Early, rapid, and accurate genetic etiological analysis is very important for subsequent patient treatment, genetic counseling, and

fertility guidance [10]. Previous studies on *PAH* gene variation have mainly focused on exon and flanking intron sequences using Sanger sequencing and multiplex ligation-dependent probe amplification (MLPA), resulting in an unsatisfactory genetic diagnosis rate for PKU [3, 4, 8, 14–16, 20, 22, 27, 28, 30]. Approximately 5% of PKU patients with typical clinical symptoms do not receive a definitive genetic diagnosis [18].

Introns are very important for eukaryotic evolution. The intron–exon structure of eukaryotic genes plays an important role in the generation of new genes through exon changes [5, 19], and the ability to alternately select different exon combinations is crucial for the gene expression diversity of complex organisms [13]. Nowadays, NGS has greatly improved the diagnosis of genetic diseases, and increasing numbers of pathogenic deep intronic variants have been discovered. Pathogenic deep intronic variants have been reported in more than 100 disease-associated genes. These pathogenic deep intronic variants commonly result in pseudo-exon inclusion via activation of atypical splice sites or changes in splicing regulatory elements [12, 27].

In this study, we performed pathogenicity analysis of deep intronic variants of the *PAH* gene in 45 patients with PKU, and we identified five different deep intron variants, c.1199+502A>T, c.1065+241C>A, c.706+368T>C, c.706+531T>C, and c.706+608A>C. Of these, c.1199+502A>T was detected in 18 patients, c.1065+241C>A was detected in three patients, and c.706+368T>C, c.706+531T>C, and c.706+608A>C were detected in one patient each. The c.1199+502A>T variant might be a hotspot of deep intronic variants for PKU patients in China. These results suggest that deep intronic variants in *PAH* can result in PKU, and *PAH* full-length sequencing can be used to identify deep intronic variants as well as WGS, with a lower detection cost. c.706+531T>C and c.706+608A>C in *PAH* are two novel variants that have not been reported previously. In silico prediction analyses showed that the two deep intronic variants may impact splice site selection, resulting in pseudo-exon inclusion in PAH. It has been reported that the deep intron variants can cause

pseudo-exon inclusion [12]. The mechanism may involve the intronic variant creating a new donor splice site and activating a pre-existing atypical acceptor splice site [7, 21, 26]. Some studies have also reported that deep intronic variants can create a new acceptor splice site or interfere with the splicing of regulatory elements, which results in pseudo-exon inclusion [1, 9, 25].

The Alamut Visual software indicated that c.706+531T>C might impact splice site selection (Table 3) and affect the branch point (Table 4). This variant was predicted by ESEfinder to create an SRSF1 (SF2/ASF(IgM-BRCA1))-binding ESE in the variant sequence, leading to activation, probably by assisting with recognition of the weak splice donor site (SDS) in a similar way as previously reported [11, 23]. We identified a 114-nt pseudo-exon in the mutant type by RT-PCR analysis of the minigene. A deep intronic variant has previously been shown to create an SRSF1-binding ESE, leading to pseudo-exon activation [11, 23]. The wildtype intron 6 also produced 114-nt pseudo-exons, however, the bands were weak. We speculated that c.706+531T>C led to a stronger SRSF1 binding site (score: 4.76, Table 4), thus activating ESE. Another reason could be that c.706+531T>C affects the splice branching point sequence, but further experimental evidence is needed to verify this.

Alamut Visual software indicated that c.706+608A>C might impact splice site selection (Table 3) and affect the branch point (Table 4). Although ESEfinder predicted that c.706+608A>C would not change the ESE (Table 4), it might strengthen the SRSF2 (SC35) binding site. We identified 114 nt, 313 nt, and 647 nt pseudo-exons in the mutant type by RT-PCR analysis of the minigene. Although the wildtype intron 6 also produced 114 nt pseudo-exons, the bands were also weak. Similar to c.706+368T>C, we hypothesized that although the c.706+608A>C variant creates a stronger SDS that is inhibited by TDP-43 binding, the c.706+608A>C variant might also activate a cryptic SDS and splice acceptor site (SAS) by strengthening an SRSF2 (SC35) binding site, resulting in pseudo-exon generation [12]. However, as before, whether this is because it affects the splicing

Table 4 In silico prediction scores for the c.706+531T>C and c.706+608A>C variants, as assessed by ESEfinder3.0

Variant	Position	SF2/ASF		SF2/ASF (IgM-BRCA1)		SC35		SRp40		SRp55	
		WT	mut	WT	mut	WT	mut	WT	mut	WT	mut
c.706+531T>C	CATAGGG	2.68	5.29	2.69	4.76	–	–	–	–	–	–
	CCATAGG	–	–	–	–	–	–	3.58	5.97	–	–
c.706+608A>C	–	–	–	–	–	–	–	–	–	–	–

WT: wild type; mut: mutant type; The threshold: SF2/ASF, 1.956; SF2/ASF (IgM-BRCA1), 1.867; SC35, 2.383; SRp40, 2.67; SRp55, 2.676. “–” indicates an unrecognized ESE sites

branch point sequence still requires further experimental evidence for verification.

Jin et al. [12] performed WGS in 10 undiagnosed PKU patients and identified three pathogenic deep intronic variants. In this study, we used 13 pairs of primers to amplify the whole *PAH* gene, and we found two pathogenic deep intronic variants through targeted full-length sequencing of *PAH*. WGS can effectively detect deep intronic variant, and has been widely used in the diagnosis of genetic diseases. However, WGS also has certain defects, such as low sequencing depth and high price. The full-length sequencing of *PAH* used in this study to detect deep intronic variant achieves the same effect as WGS, with an average sequencing depth of 9019×, and the detection cost is only 20% of WGS. Therefore, targeted sequencing after amplification of the entire gene can be an economical and effective strategy to study the deep intronic variation of the target gene.

Pathogenicity analysis of deep intronic variant in *PAH* can improve the genetic diagnosis rate of PKU. However, three mPKU and 15 MHP cases in our study did not have a clear genetic diagnosis. Gao et al. also found three mPKU cases and one cPKU case without a definitive genetic diagnosis after detecting deep intron variants in the *PAH* gene [10]. This could be because some non-*PAH* genes may affect *PAH* function and subsequently increase blood phenylalanine concentrations [2], and may also be related to some epigenetic factors [6]. The lncRNA *Pair* and the human *HULC* gene are associated with *PAH* and modulate enzymatic activities by facilitating *PAH* substrate and *PAH*–cofactor interactions [10, 17]. In addition, existing studies on deep intronic variants mainly analyzed that missense variant, whether indel variant or rearrangement of deep intron region also affect the function of *PAH* needs further investigation.

Conclusions

In summary, we analyzed the pathogenicity of deep intron variants in the *PAH* gene, and pathogenic deep intronic variants were identified in 24 of 45 PKU patients with unknown genotypes. We identified two novel pathogenic deep intronic variants in *PAH*, extending the deep intronic variant spectrum of *PAH*. This study has shown that deep intronic variant pathogenicity analysis can further improve the genetic diagnosis of PKU patients, and suggests that pathogenic deep intron variation may not be uncommon in *PAH*. In silico prediction and minigene analysis can be powerful approaches for studying the functions and effects of deep intronic variants. Analysis of deep intron variation of disease-specific genes can be conducted by targeted sequencing after full-length gene amplification, and so this study provides a new avenue for

the detection of deep intron variation in genes with small fragments.

Abbreviations

cPKU	Classic PKU
ESE	Exonic splicing enhancer
HPA	Hyperphenylalaninemia
MHP	Mild hyperphenylalaninemia
MLPA	Multiplex ligation-dependent probe amplification
mPKU	Mild PKU
MT	Mutant type
PKU	Phenylketonuria
SAS	Splice acceptor site
SDS	Splice donor site
WGS	Whole genome sequencing
WT	Wild type

Acknowledgements

The authors are grateful to the patients and the pedigree members who participated in this study.

Author contributions

ZC, YYS, CZF and MX: designed the research; ZC, YYS, ZBB, WYP and TXY performed the molecular experiments and bioinformatics analysis; ZC, HSJ, MPP, ZL, ZQH, HL and WY: analyzed the data; ZC, CZF and MX: wrote the paper; and all authors: read, critically revised and approved the final manuscript.

Funding

This work was supported by the National Key Research and Development Program of China (2016YFC1000307); Gansu Provincial Science and Technology Plan Funding Project (22YF7FA094); Lanzhou Science and Technology Plan Project (2021-1-182); National Human Genetic Resources Sharing Service Platform (2005DKA21300); National Science and Technology Resource Sharing Service Platform Project (YCZYPT[2020]05-03); Gansu Provincial Clinical Research Center for Birth Defects and Rare Diseases (21JR7RA680); Gansu Natural Science Foundation (21JR1RA047, 21JR1RA045).

Availability of data and materials

The data that support the findings of this study is available upon reasonable request from corresponding authors. The two novel variants have been submitted to the Clinvar (<https://www.ncbi.nlm.nih.gov/clinvar/>). The accession numbers of c.706+531T>C was SCV002599130 and c.706+608A>C was SCV002599425.

Declarations

Ethics approval and consent to participate

The study protocol was approved by the Ethics Committee of the Gansu Provincial Maternity and Child-Care Hospital (2021GSFY[65]). Written informed consent was obtained from all study participants or their legal guardians.

Consent for publication

All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Received: 4 January 2023 Accepted: 18 May 2023

Published online: 26 May 2023

References

- Albert S, Garanto A, Sangermano R, Khan M, Bax NM, Hoyng CB, Zernant J, Lee W, Allikmets R, Collin RWJ, Cremers FPM. Identification and rescue of splice defects caused by two neighboring deep-intronic ABCA4 mutations underlying stargardt disease. *Am J Hum Genet.* 2018;102:517–27.

2. Anikster Y, Haack TB, Vilboux T, Pode-Shakked B, Thöny B, Shen N, Guarani V, Meissner T, Mayatepek E, Trefz FK, Marek-Yagel D, Martinez A, Huttlin EL, Paulo JA, Berutti R, Benoist JF, Imbard A, Dorboz I, Heimer G, Landau Y, Ziv-Strasser L, Malicdan MCV, Gemperle-Britschgi C, Cremer K, Engels H, Meili D, Keller I, Bruggmann R, Strom TM, Meitinger T, Mullikin JC, Schwartz G, Ben-Zeev B, Gahl WA, Harper JW, Blau N, Hoffmann GF, Prokisch H, Opladen T, Schiff M. Biallelic mutations in DNAJC12 cause hyperphenylalaninemia, dystonia, and intellectual disability. *Am J Hum Genet.* 2017;100:257–66.
3. Aulehla-Scholz C, Heilbronner H. Mutational spectrum in German patients with phenylalanine hydroxylase deficiency. *Hum Mutat.* 2003;21:399–400.
4. Bashyam MD, Chaudhary AK, Kiran M, Nagarajaram HA, Devi RR, Ranganath P, Dalal A, Bashyam L, Gupta N, Kabra M, Muranjan M, Puri RD, Verma IC, Nampoothiri S, Kadandale JS. Splice, insertion-deletion and nonsense mutations that perturb the phenylalanine hydroxylase transcript cause phenylketonuria in India. *J Cell Biochem.* 2014;115:566–74.
5. Becker PW, Sacilotto N, Normes S, Neal A, Thomas MO, Liu K, Preece C, Ratnayaka I, Davies B, Bou-Gharios G, De Val S. An Intronic Flk1 enhancer directs arterial-specific expression via RBPJ-mediated venous repression. *Arterioscler Thromb Vasc Biol.* 2016;36:1209–19.
6. Cederbaum S. Phenylketonuria: an update. *Curr Opin Pediatr.* 2002;14:702–6.
7. Chang CY, Perng CL, Cheng SN, Hu SH, Wu TY, Lin SY, Chen YC. Deep intronic variant c.5999-277G>A of F8 gene may be a hot spot mutation for mild hemophilia A patients without mutation in exonic DNA. *Eur J Haematol.* 2019;103:47–55.
8. Chen C, Zhao Z, Ren Y, Kong X. Characteristics of PAH gene variants among 113 phenylketonuria patients from Henan Province. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi.* 2018;35:791–5.
9. Fadaie Z, Khan M, Del Pozo-Valero M, Cornelis SS, Ayuso C, Cremers FPM, Roosing S, The Abca Study Group. Identification of splice defects due to noncanonical splice site or deep-intronic variants in ABCA4. *Hum Mutat.* 2019;40:2365–76.
10. Gao J, Li X, Guo Y, Yu H, Song L, Fang Y, Yuan E, Shi Q, Zhao D, Yuan E, Zhang L. Identification of phenylketonuria patient genotypes using single-gene full-length sequencing. *Hum Genom.* 2022;16:23.
11. Homolova K, Zavadakova P, Doktor TK, Schroeder LD, Kozich V, Andresen BS. The deep intronic c.903+469T>C mutation in the MTRR gene creates an SF2/ASF binding exonic splicing enhancer, which leads to pseudo-exon activation and causes the cblE type of homocystinuria. *Hum Mutat.* 2010;31:437–44.
12. Jin X, Yan Y, Zhang C, Tai Y, An L, Yu X, Zhang L, Hao S, Cao X, Yin C, Ma X. Identification of novel deep intronic PAH gene variants in patients diagnosed with phenylketonuria. *Hum Mutat.* 2022;43:56–66.
13. Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet.* 2010;11:345–55.
14. Kostandyan N, Britschgi C, Matevosyan A, Oganezova A, Davtyan A, Blau N, Steinmann B, Thöny B. The spectrum of phenylketonuria genotypes in the Armenian population: identification of three novel mutant PAH alleles. *Mol Genet Metab.* 2011;104:593–6.
15. Lee YW, Lee DH, Kim ND, Lee ST, Ahn JY, Choi TY, Lee YK, Kim SH, Kim JW, Ki CS. Mutation analysis of PAH gene and characterization of a recurrent deletion mutation in Korean patients with phenylketonuria. *Exp Mol Med.* 2008;40:533–40.
16. Li N, Jia H, Liu Z, Tao J, Chen S, Li X, Deng Y, Jin X, Song J, Zhang L, Liang Y, Wang W, Zhu J. Molecular characterisation of phenylketonuria in a Chinese mainland population using next-generation sequencing. *Sci Rep.* 2015;5:15769.
17. Li Y, Tan Z, Zhang Y, Zhang Z, Hu Q, Liang K, Jun Y, Ye Y, Li YC, Li C, Liao L, Xu J, Xing Z, Pan Y, Chatterjee SS, Nguyen TK, Hsiao H, Egranov SD, Putluri N, Coarfa C, Hawke DH, Gunaratne PH, Tsai KL, Han L, Hung MC, Calin GA, Namour F, Guéant JL, Muntau AC, Blau N, Sutton VR, Schiff M, Feillet F, Zhang S, Lin C, Yang L. A noncoding RNA modulator potentiates phenylalanine metabolism in mice. *Science.* 2021;373:662–73.
18. Liu N, Huang Q, Li Q, Zhao D, Li X, Cui L, Bai Y, Feng Y, Kong X. Spectrum of PAH gene variants among a population of Han Chinese patients with phenylketonuria from northern China. *BMC Med Genet.* 2017;18:108.
19. Long M, Betrán E, Thornton K, Wang W. The origin of new genes: glimpses from the young and old. *Nat Rev Genet.* 2003;4:865–75.
20. Lu Q, Liu Y, Yang B, Xie K, Zou Y, Lu W, Wang F. Study of the phenylalanine hydroxylase gene variants in patients with phenylketonuria from Jiangxi province. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi.* 2019;36:1057–61.
21. Malekkou A, Sevastou I, Mavrikiou G, Georgiou T, Vilageliu L, Moraitou M, Michelakakis H, Prokopiou C, Drousiotou A. A novel mutation deep within intron 7 of the GBA gene causes Gaucher disease. *Mol Genet Genomic Med.* 2020;8:e1090.
22. Okano Y, Kudo S, Nishi Y, Sakaguchi T, Aso K. Molecular characterization of phenylketonuria and tetrahydrobiopterin-responsive phenylalanine hydroxylase deficiency in Japan. *J Hum Genet.* 2011;56:306–12.
23. Palhais B, Præstegaard VS, Sabaratnam R, Doktor TK, Lutz S, Burda P, Suormala T, Baumgartner M, Fowler B, Bruun GH, Andersen HS, Kožich V, Andresen BS. Splice-shifting oligonucleotide (SSO) mediated blocking of an exonic splicing enhancer (ESE) created by the prevalent c.903+469T>C MTRR mutation corrects splicing and restores enzyme activity in patient cells. *Nucleic Acids Res.* 2015;43:4627–39.
24. Scriver CR. Phenylketonuria—genotypes and phenotypes. *N Engl J Med.* 1991;324:1280–1.
25. Sun W, Xiao X, Li S, Jia X, Zhang Q. A novel deep intronic COL2A1 mutation in a family with early-onset high myopia/ocular-only Stickler syndrome. *Ophthalmic Physiol Opt.* 2020;40:281–8.
26. Tozawa Y, Abdroub SSMA, Nogawa-Chida N, Nishiuchi R, Ishida T, Suzuki Y, Sano H, Kobayashi R, Kishimoto K, Ohara O, Imai K, Naruto T, Kobayashi K, Ariga T, Yamada M. A deep intronic mutation of c.1166–285T>G in SLC46A1 is shared by four unrelated Japanese patients with hereditary folate malabsorption (HFM). *Clin Immunol.* 2019;208:108256.
27. Vaz-Drago R, Custódio N, Carmo-Fonseca M. Deep intronic mutations and human disease. *Hum Genet.* 2017;136:1093–111.
28. Wang R, Shen N, Ye J, Han L, Qiu W, Zhang H, Liang L, Sun Y, Fan Y, Wang L, Wang Y, Gong Z, Liu H, Wang J, Yan H, Blau N, Gu X, Yu Y. Mutation spectrum of hyperphenylalaninemia candidate genes and the genotype-phenotype correlation in the Chinese population. *Clin Chim Acta.* 2018;481:132–8.
29. Writing Group For Practice Guidelines For Diagnosis And Treatment of Genetic Diseases Medical Genetics Branch Of Chinese Medical Association, Huang S, Song F. Clinical practice guidelines for phenylketonuria. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi.* 2020;37:226–34.
30. Xiao Y, Gu Q, Wu HR, Wang ST, Pei P, Zheng XF, Pan H, Ma YN. Molecular diagnosis of phenylketonuria in 157 Chinese families and the results of prenatal diagnosis in these families. *Chin Med J.* 2021;134:1626–8.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

