ORIGINAL ARTICLE

Mutation spectrum of *PAH* gene in phenylketonuria patients in Northwest China: identification of twenty novel variants

Yousheng Yan [1,2,3] · Chuan Zhang [3] · Xiaohua Jin [1,2] · Qinhua Zhang [3] · Lei Zheng [3] · Xuan Feng [3] · Shengju Hao [3] · Huafang Gao [1,2] · Xu Ma [1,2]

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

This study was performed to analyze the mutational spectrum of the phenylalanine hydroxylase (*PAH*) gene in phenylketonuria (PKU) patients in Northwest China, to identify mutational hot spots, and to determine the correlation between variants and clinical phenotypes of PKU. A large cohort of 475 PKU families in Northwest China was enrolled to analyze *PAH* gene variants using Sanger sequencing, Multiplex ligation-dependent probe amplification (MLPA), and gap-PCR. Bioinformatics software was used to predict the pathogenicity of novel variants and analyze the correlations between *PAH* gene variants and phenotypes of PKU patients. A total of 895 variants were detected in the 950 alleles of 475 patients with PKU (detection rate: 94.21%), 20 of which were novel variants. Other 108, previously known variants, were also identified, with the three most frequent variants being p.Arg243Gln (14.00%), c.611A > G (5.58%), and p.Tyr356* (4.95%). Seven different large deletion/duplication variants were identified by the MLPA method, including the large deletion c.-4163_-406del3758 with high frequency. A correlation analysis between patient phenotype and gene variant frequency showed that p.Arg53His and p.Gln419Arg were correlated with mild hyperphenylalaninemia (MHP). In conclusion, the mutational spectrum underlying PKU in Northwest China was established for the first time. Functional analysis of 20 novel *PAH* gene variants enriched the *PAH* gene mutational spectrum. Correlation analysis between variants frequencies in compound heterozygous patients and phenotype severity is helpful for phenotypic prediction.

Keywords Phenylketonuria · Phenylalanine hydroxylase · Sanger sequencing · MLPA · Gap-PCR · Gene variant

Introduction

Phenylketonuria (PKU) is one of the most common autosomal recessive metabolic diseases. Phenylalanine (Phe) metabolic disorders in the liver can be caused by phenylalanine hydroxylase (*PAH*) gene variant (Blau et al. 2010). The incidence of PKU in the northwest regions of China is significantly higher than the average national level (1/11,188) (Gu and Wang 2004) reaching its highest level in Gansu (1/3420) (Wang et al. 2015). This may be related to the unique geographic location and ethnic composition of the northwest region. The low rates of PKU screening and treatment affect the quality of life of the population and the infant mortality rate. Therefore, early diagnosis and treatment of PKU is particularly important.

The human *PAH* gene is located on chromosome 12q23.2 and contains a total of 13 exons and 12 introns (Woo et al. 1983). More than 1000 variants have been identified in *PAH,* recorded in the locus-specific database (LSD) known as *PAH*vdb (http://www.biopku.org/pah/). The detection of *PAH* gene variants and correlation analysis between PKU genotype and

✉ Huafang Gao
   gaohuafang@nrifp.org.cn

✉ Xu Ma
   genetic@263.net.cn

[1] Graduate School of Peking Union Medical College, Beijing 100730, China

[2] National Research Institute for Health and Family Planning, No. 12 Dahuisi Road, Beijing 100081, China

[3] Gansu Province Medical Genetics Center, Gansu Province Maternal and Child Health Care Hospital, Lanzhou 730050, China

clinical phenotype may be of great value for PKU genetic counseling and prenatal diagnosis. As an important link on the Silk Road, gene mixing is frequent in the northwest regions and *PAH* gene variants may have a unique distribution in this area. Thus, although it is expected that some overlap exists with various other provinces and autonomous regions of China (He et al. 2015; Qiang et al. 2014; Yan et al. 2009; Zhang et al. 2015), a large-scale analysis of the *PAH* gene mutational spectrum of PKU patients is of great significance and importance for the prevention and treatment of PKU in this region.

In this study, an analysis of *PAH* gene variants was performed in a large cohort of 475 PKU patients and their families in the Northwest of China by Sanger sequencing and the MLPA method. This is helpful for the establishment of the *PAH* gene mutation spectrum and clarification of *PAH* gene mutation distribution characteristics of PKU patients in Northwest China. These results provide the basic data for prenatal diagnosis and prevention of PKU. Further correlation exploration between genotype and phenotype of PKU patients also provides some theoretical basis for the study of gene function.

Materials and methods

Patients and phenotypes

A large cohort of 475 patients with PKU, diagnosed at the Newborn Screening Center and Medical Genetics Center of Gansu Province Maternal and Child Health Hospital from January of 2007 to January of 2017, were included in this study. All patients were from Gansu, Shaanxi, Ningxia, Qinghai and Xinjiang. This study was approved by the Ethics Committee of Gansu Province Maternal and Child Health Care Hospital, and the subjects included in this study have provided written informed consent.

These patients were identified through a neonatal screening program. The plasma Phe concentration were measured by tandem mass spectrometry from dried blood spot (DBS) samples before starting treatment. According to classification criteria (Yang and Ye 2014), the 475 PKU patients were divided into three types: 253 cases of classic PKU (Phe ≥ 1200 μmol/L), 150 cases of moderate PKU (Phe: 360 ~ 1200 μmol/L), and 74 cases of mild hyperphenylalaninemia (MHP) (Phe: 120 ~ 360 μmol/L). All patients were excluded from tetrahydrobiopterin deficiency through a BH4 loading test, a urinary pterin analysis, and a DHPR activity assay on DBS samples.

Sanger sequencing

Genomic DNA (gDNA) was extracted from the peripheral blood mononuclear cell or DBS samples of patients using Whole Blood Genomic DNA Extraction Kits (Tiangen, Beijing, China) or the Chelex-100 (Bio-rad) method according to the product descriptions. The *PAH* gene reference sequence (NG_008690.1) and the cDNA sequence (NM_000277.1) were obtained from the UCSC database (Giardine et al. 2007). A total of 13 PCR primer pairs, targeting exons 1–13 and the flanking sequences of the *PAH* gene were designed using Primer Premier 5.0 software (Premier Biosoft International, Palo Alto, CA, US). The BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) was used to generate sequencing products from purified PCR products. Sequencing products were subjected to capillary electrophoresis on an ABI 3500DX DNA analyzer (Applied Biosystems, Foster City, CA, USA). Sequencing results were analyzed using Chromas 2.33 software (Technelysium Pty Ltd., South Brisbane, Australia), and variants were determined using Seqman in the DNASTAR® Lasergene v7.1 software package (DNASTAR, Inc., Madison, USA) by comparing sample sequences with the normal reference sequence.

MLPA analysis

The detection of large-scale deletion/duplication variants of *PAH* gene was performed in patients with only one variant allele of *PAH* or in those with no variant found, using the SALSA MLPA kit P055 *PAH* (Cat. P055-100R, MRC Holland, Amsterdam, Netherlands). The assay operation was performed according to the manufacturer's protocol. Copy number variations of exons of *PAH* were analyzed by the Coffalyser data analysis software (MRC-Holland b.v., Amsterdam, Netherlands). Copy number was indicated by ratio value. The ratios range of normal reference is 0.7–1.3; the ratios indicate 3 copies in the range of 1.3–1.7, which is judged to be a heterozygous duplication mutation; the ratios indicate a single copy from 0.3 to 0.7, which is judged to be a heterozygous deletion mutation; the ratios indicate homozygous at 0 to 0.3 and is judged to be a homozygous deletion mutation. Four normal control DNA samples from healthy individuals were used in each run. Copy number variations of exons of *PAH* detected by MLPA were identified by gap-PCR analysis. Gap-PCR primers for detection of type D1 (Ex1del3758) and type D2 (Ex1del5269ins56) in exon 1 and deletion of exon 3 were designed based on previous reports (Desviat

et al. 2006; Lu et al. 2011). The primers for other deletion sites were designed using Primer 5.0 (Supplementary Table 1). All aberrantly amplified fragments detected by gap-PCR were sequenced to identify the deletion sites.

Analysis of novel variants

All detected *PAH* gene variants were screened against the 1000 Genomes Project (Sudmant et al. 2015), dbSNP (Hubbard et al. 2016), and ExAC databases (Song et al. 2016) to exclude polymorphic sites in the population, and against the ClinVar (Zastrow et al. 2018), HGMD (Qiang et al. 2014), and *PAH*vdb databases (Polak et al. 2013) to identify potential unreported variants (novel variants). The nomenclature of novel variants was based on the sequence variant nomenclature of the Human Genome Variation Society (den Dunnen et al. 2016). The novel missense variants and splice variants were identified in accordance with the criteria for interpretation of sequence variants (Richards et al. 2015).

Statistical analysis

SPSS 16.0 software was used for statistical analysis. Comparisons of the distribution frequencies of different variants in different types of patients were performed with the Chi-square test. A P value <0.05 was considered statistically significant.

Results

Mutation spectrum of *PAH* gene

A total of 895 variant alleles were detected in the 950 alleles of 475 PKU patients, and 128 different pathogenic variants were identified (Table 1). All detected variants were derived from parents and no de novo mutation was found. Among the 895 detected mutant alleles, the 10 most frequent variants represented 50.21% of the total, which were p.Arg243Gln (14.00%), EX6-96A > G (c.611A > G) (5.58%), p.Tyr356* (4.95%), p.Arg413Pro (4.74%), IVS4-1G > A (c.442-1G > A) (4.32%), p.Val399 = (4.11%), p.Arg241Cys (4.11%), p.Arg111* (2.95%), IVS7 + 2 T > A (c.842 + 2 T > A) (2.95%), and p.Ile65Thr (2.53%). The highest frequencies of variants were found in exon 7 (33.04%), exon 11 (15.29%), exon 12 (12.7%), exon 5 (10.4%), exon 3 (9.15%) and exon 6 (8.37%), respectively.

The 895 variants were found at 128 sites and involved 8 variant types, including 83 missense variants, 20 splice site variants, 12 non-sense variants, 7 large deletion variants, 4 minor deletion variants, 1 large-scale duplication variant, 1 silence/splicing variant, and 1 5'UTR variant.

Large-scale deletion/duplication analysis

A total of 74 cases without two known pathogenic variants were identified by exon sequencing. *PAH* gene deletion/duplication variant was analyzed in those 74 patients using the MLPA method. Seven large deletion/duplication variants of *PAH* gene were detected in 25 patients. Two of the seven deletion variants were located in the 5'UTR of exon 1 of the *PAH* gene. Four other deletion variants were located in exon 3 to exon 7, and the last one was a duplication variant in exon 12 (Table 1, Fig. 1).

Gap-PCR products were sequenced for large-scale deletions involving exons 1 and 3, and abnormally amplified fragments were detected (c.1-4163_1-406del3758, c.1-1932_60 + 3402del5269ins56, and c.169-4940_352 + 1459del6604ins8). Gap-PCR products using primers IVS5-8F and IVS-1R revealed that the length of the amplified fragment of exon 6 deletion variant was 804 bp, while the normal amplified fragment was 8597 bp. The PCR product was identified by sequencing and compared with the reference gene sequence to confirm that a 2 G bases replaced a fragment of 7793 bp, which was from the nucleotide −6708 of intron 5 to the nucleotide +888 of intron 6. According to the HGVS naming rules (den Dunnen et al. 2016), this variant was named c.510-6708_706 + 888del7793insGG (abbreviation: Exon6del7793insGG) (Fig. 2).

Novel variants in *PAH* gene

A total of 20 of the 128 variants were not reported in the ClinVar, HGMD, or *PAH*vdb databases (Fig. 3). In addition, 3 of the 128 variants were recorded with extremely low frequency and unknown pathogenicity in 1000 Genomes, dbSNP and ExAC databases. Pathogenicity analysis showed that 6 missense variants were likely pathogenic, and that 4 missense variants were variant with the unknown clinical significance. In addition, abnormal splicing was predicted for the IVS9 + 4A > T (c.969 + 4A > T) variant in exon 9, where the original splice site was removed (Table 2).

Table 1 PAH gene mutational spectrum in PKU patients in Northwest China

Systematic name (DNA level)	Common name (protein effect)	Location	Mutation type	Classical PKU (n = 253)		Moderate PKU (n = 150)		MHP (n = 72)		P value
				No. of alleles	RF %	No. of alleles	RF %	No. of alleles	RF %	
c.728G>A	p.Arg243Gln	E7	Missense	81	16.01	44	14.67	8	5.56	0.0057[#]
c.611A>G	p.?/EX6-96A>G	E6	Splice	32	6.32	15	5.00	6	4.17	0.5301
c.1068C>A	p.Tyr356*	E11	Nonsense	31	6.13	8	2.67	8	5.56	0.0851
c.1238G>C	p.Arg413Pro	E12	Missense	23	4.55	15	5.00	7	4.86	0.9550
c.442-1G>A	p.?/IVS4-1G>A	I4	Splice	30	5.93	7	2.33	4	2.78	0.0323[#]
c.1197A>T	p.Val399=	E11	Silence/Splice	27	5.34	10	3.33	2	1.39	0.0781
c.721C>T	p.Arg241Cys	E7	Missense	10	1.98	25	8.33	4	2.78	0.0000[#]
c.331C>T	p.Arg111*	E3	Nonsense	22	4.35	5	1.67	1	0.69	0.0208[#]
c.842+2T>A	p.?/IVS7+2T>A	I7	Splice	15	2.96	10	3.33	3	2.08	0.7662
c.194T>C	p.Ile65Thr	E3	Missense	15	2.96	6	2.00	3	2.08	0.6549
c.158G>A	p.Arg53His	E2	Missense	0	0.00	1	0.33	20	13.89	0.0000[#]
c.782G>A	p.Arg261Gln	E7	Missense	11	2.17	7	2.33	2	1.39	0.8002
c.1199G>A	p.Arg400Lys	E11	Missense	10	1.98	9	3.00	0	0.00	0.1069
c.208_210delTCT	p.Ser70del	E3	Deletion	8	1.58	8	2.67	2	1.39	0.4897
c.764T>C	p.Leu255Ser	E7	Missense	12	2.37	4	1.33	0	0.00	0.1266
c.-4163_-406del3758	p.?/Ex1del3758	5'UTR_E1	Large Deletion/Insertion	11	2.17	4	1.33	1	0.69	0.4050
c.1301C>A	p.Ala434Asp	E12	Missense	6	1.19	8	2.67	0	0.00	0.0678
c.526C>T	p.Arg176*	E6	Nonsense	9	1.78	2	0.67	0	0.00	0.1338
c.1252A>C	p.Thr418Pro	E12	Missense	0	0.00	9	3.00	1	0.69	0.0003[#]
c.1256A>G	p.Gln419Arg	E12	Missense	1	0.20	2	0.67	7	4.86	0.0000[#]
c.1315+6T>A	p.?/IVS12+6T>A	I12	Splice	2	0.40	5	1.67	3	2.08	0.0976
c.498C>G	p.Tyr166*	E5	Nonsense	4	0.79	6	2.00	0	0.00	0.1080
c.1222C>T	p.Arg408Trp	E12	Missense	7	1.38	2	0.67	0	0.00	0.2652
c.694C>T	p.Gln232*	E6	Nonsense	6	1.19	2	0.67	1	0.69	0.7202
c.740G>T	p.Gly247Val	E7	Missense	3	0.59	5	1.67	0	0.00	0.1325
c.781C>T	p.Arg261*	E7	Nonsense	7	1.38	1	0.33	0	0.00	0.1403
c.472C>T	p.Arg158Trp	E5	Missense	4	0.79	2	0.67	0	0.00	0.5698
c.671T>C	p.Ile224Thr	E6	Missense	4	0.79	1	0.33	1	0.69	0.7269
c.739G>C	p.Gly247Arg	E7	Missense	4	0.79	2	0.67	0	0.00	0.5698
c.838G>A	p.Glu280Lys	E7	Missense	4	0.79	1	0.33	1	0.69	0.7269
c.1114A>T	p.Thr372Ser	E11	Missense	0	0.00	3	1.00	2	1.39	0.0496[#]
c.1174T>A	p.Phe392Ile	E11	Missense	0	0.00	0	0.00	5	3.47	0.0000[#]
c.1199G>C	p.Arg400Thr	E11	Missense	4	0.79	0	0.00	1	0.69	0.3104
c.320A>G	p.His107Arg	E3	Missense	0	0.00	2	0.67	3	2.08	0.0088[#]

Table 1 (continued)

| Systematic name (DNA level)	Common name (protein effect)	Location	Mutation type	Classical PKU ($n = 253$)		Moderate PKU ($n = 150$)		MHP ($n = 72$)		P value
				No. of alleles	RF %	No. of alleles	RF %	No. of alleles	RF %	
c.722delG	p. Arg241Profs*100	E7	Deletion	4	0.79	1	0.33	0	0.00	0.4383
c.1045 T > C	p.Ser349Ala	E10	Missense	2	0.40	1	0.33	1	0.69	0.8522
c.1066-11G > A	p.?/IVS10−11	I10	Splice	3	0.59	1	0.33	0	0.00	0.6004
c.168 + 5G > C	p.?/IVS2 + 5G > C	I2	Splice	3	0.59	1	0.33	0	0.00	0.6004
c.47_48delCT	p.Ser16*	E1	Deletion	4	0.79	0	0.00	0	0.00	0.1716
c.473G > A	p.Arg158Gln	E5	Missense	4	0.79	0	0.00	0	0.00	0.1716
c.505C > A	p.Arg169Ser	E5	Missense	1	0.20	3	1.00	0	0.00	0.1645
c.827 T > A	p.Met276Lys	E7	Missense	2	0.40	1	0.33	1	0.69	0.8522
c.910C > A	p.Gln304Lys	E8	Missense	3	0.59	1	0.33	0	0.00	0.6004
c.1147C > G	p.Gln383Glu	E11	Missense	2	0.40	0	0.00	1	0.69	0.4258
c.1159 T > G	p.Tyr387Asp	E11	Missense	3	0.59	0	0.00	0	0.00	0.2670
c.1289 T > C	p.Leu430Pro	E12	Missense	0	0.00	3	1.00	0	0.00	0.0384[#]
c.1315 + 5G > C	p.?/IVS12 + 5G > C	I12	Splice	2	0.40	1	0.33	0	0.00	0.7555
c.470G > A	p.Arg157Lys	E5	Missense	3	0.59	0	0.00	0	0.00	0.2670
c.475A > T	p.Lys159*	E5	Nonsense	2	0.40	1	0.33	0	0.00	0.7555
c.482 T > C	p.Phe161Ser	E5	Missense	2	0.40	0	0.00	1	0.69	0.4258
c.59_60AG > CC	p.Gln20Pro	E1	Missense	2	0.40	1	0.33	0	0.00	0.7555
c.688G > A	p.Val230Ile	E6	Missense	0	0.00	0	0.00	3	2.08	0.0002[#]
c.755G > A	p.Arg252Gln	E7	Missense	3	0.59	0	0.00	0	0.00	0.2670
c.770G > T	p.Gly257Val	E7	Missense	2	0.40	0	0.00	1	0.69	0.4258
c.842C > T	p.Pro281Leu	E7	Missense	1	0.20	2	0.67	0	0.00	0.3958
c.971 T > A	p.Ile324Asn	E10	Missense	1	0.20	1	0.33	1	0.69	0.6430
c.-1932_60 + 3402del 5269ins65	p.?/Ex1del5269 ins26	5'UTR	Large Deletion	3	0.59	0	0.00	0	0.00	0.2670
c.1024G > A	p.Ala342Thr	E10	Missense	0	0.00	1	0.33	1	0.69	0.2360
c.1045 T > G	p.Ser349Pro	E10	Missense	0	0.00	1	0.33	1	0.69	0.2360
c.1054G > C	p.Gly352Arg	E10	Missense	1	0.20	0	0.00	1	0.69	0.3259
c.1162G > A	p.Val388Met	E11	Missense	1	0.20	1	0.33	0	0.00	0.7698
c.284_286delTCA	p.Ile95del	E3	Deletion	1	0.20	1	0.33	0	0.00	0.7698
c.454C > A	p.Pro152Thr	E5	Missense	0	0.00	1	0.33	1	0.69	0.2360
c.665A > G	p.Asp222Gly	E6	Missense	2	0.40	0	0.00	0	0.00	0.4151
c.727C > T	p.Arg243*	E7	Nonsense	1	0.20	1	0.33	0	0.00	0.7698
c.775G > A	p.Ala259Ser	E7	Missense	0	0.00	2	0.67	0	0.00	0.1140
c.833C > T	p.Thr278Ile	E7	Missense	2	0.40	0	0.00	0	0.00	0.4151
c.913-7A > G	p.?/IVS8-7A > G	I8	Splice	0	0.00	1	0.33	1	0.69	0.2360

Table 1 (continued)

Systematic name (DNA level)	Common name (protein effect)	Location	Mutation type	Classical PKU (n = 253)		Moderate PKU (n = 150)		MHP (n = 72)		P value
				No. of alleles	RF %	No. of alleles	RF %	No. of alleles	RF %	
c.926C > A	p.Ala309Asp	E9	Missense	0	0.00	2	0.67	0	0.00	0.1140
c.977G > A	p.Trp326*	E10	Nonsense	2	0.40	0	0.00	0	0.00	0.4151
c.353-?_509 +?del	p.?/Ex4_5del	E4_5	Large Deletion	1	0.20	1	0.33	0	0.00	0.7698
c.169-4940_352 + 1459 del6604ins8	p.?/Ex3del6604ins8	E3	Large Deletion/Insertion	2	0.40	0	0.00	0	0.00	0.4151
c.353-?_842 +?del	p.?/Exon4_7del	E4_7	Large Deletion	2	0.40	0	0.00	0	0.00	0.4151
c.1024G > T	p.Ala342Ser	E10	Missense	0	0.00	0	0.00	1	0.69	0.0607
c.1123C > G	p.Gln375Glu	E11	Missense	0	0.00	0	0.00	1	0.69	0.0607
c.1139C > T	p.Thr380Met	E11	Missense	0	0.00	0	0.00	1	0.69	0.0607
c.1172C > A	p.Ser391Thr	E11	Missense	0	0.00	0	0.00	1	0.69	0.0607
c.1199 + 1G > A	p.?/IVS11 + 1G > A	I11	Splice	1	0.20	0	0.00	0	0.00	0.6446
c.1199 + 1G > C	p.?/IVS11 + 1G > C	I11	Splice	0	0.00	1	0.33	0	0.00	0.3381
c.1200-1G > C	p.?/IVS11-1G > C	I11	Splice	0	0.00	0	0.00	1	0.69	0.0607
c.1208C > T	p.Ala403Val	E12	Missense	0	0.00	0	0.00	1	0.69	0.0607
c.1216A > G	p.Ile406Val	E12	Missense	0	0.00	0	0.00	1	0.69	0.0607
c.121C > T	p.Leu41Phe	E2	Missense	0	0.00	0	0.00	1	0.69	0.0607
c.1223G > A	p.Arg408Gln	E12	Missense	0	0.00	1	0.33	0	0.00	0.3381
c.1241A > G	p.Tyr414Cys	E12	Missense	0	0.00	1	0.33	0	0.00	0.3381
c.1242C > A	p.Tyr414*	E12	Nonsense	1	0.20	0	0.00	0	0.00	0.6446
c.1262 T > C	p.Ile421Thr	E12	Missense	0	0.00	1	0.33	0	0.00	0.3381
c.1285C > A	p.Gln429Lys	E12	Missense	0	0.00	1	0.33	0	0.00	0.3381
c.1303G > A	p.Asp435Asn	E12	Missense	0	0.00	1	0.33	0	0.00	0.3381
c.1315 + 4A > G	IVS12 + 4A > G	I12	Splice	0	0.00	1	0.33	0	0.00	0.3381
c.1331 T > C	p.Leu444Pro	E13	Missense	1	0.20	0	0.00	0	0.00	0.6446
c.151G > C	p.Val51Leu	E2	Missense	0	0.00	1	0.33	0	0.00	0.3381
c.168 + 5G > A	p.?/IVS2 + 5G > A	I2	Splice	0	0.00	1	0.33	0	0.00	0.3381
c.168 + 5G > T	p.?/IVS2 + 5G > T	I2	Splice	1	0.20	0	0.00	0	0.00	0.6446
c.194 T > A	p.Ile65Asn	E3	Missense	0	0.00	1	0.33	0	0.00	0.3381
c.205C > A	p.Pro69Thr	E3	Missense	0	0.00	1	0.33	0	0.00	0.3381
c.212G > A	p.Arg71His	E3	Missense	0	0.00	1	0.33	0	0.00	0.3381
c.280A > G/688G > A	p.Ile94Val	E3	Missense	0	0.00	0	0.00	1	0.69	0.0607
c.332G > A + c.1054G > C	p.Arg111Gln	E3	Missense	0	0.00	1	0.33	0	0.00	0.3381
c.464G > A	p.Arg155His	E5	Missense	0	0.00	0	0.00	1	0.69	0.0607
c.502 T > A	p.Tyr168Asn	E5	Missense	1	0.20	0	0.00	0	0.00	0.6446

Table 1 (continued)

Systematic name (DNA level)	Common name (protein effect)	Location	Mutation type	Classical PKU (n = 253)		Moderate PKU (n = 150)		MHP (n = 72)		P value
				No. of alleles	RF %	No. of alleles	RF %	No. of alleles	RF %	
c.505C > T	p.Arg169Cys	E5	Missense	0	0.00	0	0.00	1	0.69	0.0607
c.516G > T	p.Gln172His	E6	Missense	0	0.00	0	0.00	1	0.69	0.0607
c.-52G > C	p.?	5'UTR	5'UTR	1	0.20	0	0.00	0	0.00	0.6446
c.689 T > C	p.Val230Ala	E6	Missense	1	0.20	0	0.00	0	0.00	0.6446
c.707-1G > A	IVS6-1G > A	I6	Splice	1	0.20	0	0.00	0	0.00	0.6446
c.724C > T	p.Leu242Phe	E7	Missense	1	0.20	0	0.00	0	0.00	0.6446
c.737C > T	p.Ala246Val	E7	Missense	0	0.00	0	0.00	1	0.69	0.0607
c.773 T > G	p.Leu258Arg	E7	Missense	0	0.00	1	0.33	0	0.00	0.3381
c.813 T > G	p.His271Gln	E7	Missense	1	0.20	0	0.00	0	0.00	0.6446
c.837C > T	p.Pro279=	E7	Missense	0	0.00	0	0.00	1	0.69	0.0607
c.842 + 1G > A	p.?/IVS7 + 1G > A	I7	Splice	0	0.00	1	0.33	0	0.00	0.3381
c.842 + 4A > T	p.?/IVS7 + 4A > T	I7	Splice	0	0.00	0	0.00	1	0.69	0.0607
c.843-6 T > C	p.?/IVS7-6 T > C	I7	Splice	1	0.20	0	0.00	0	0.00	0.6446
c.865G > A	p.Gly289Arg	E8	Missense	1	0.20	0	0.00	0	0.00	0.6446
c.871G > T	p.Val291Leu	E8	Missense	0	0.00	0	0.00	1	0.69	0.0607
c.87C > A	p.Cys29*	E8	Nonsense	0	0.00	1	0.33	0	0.00	0.3381
c.890G > A	p.Arg297His	E8	Missense	1	0.20	0	0.00	0	0.00	0.6446
c.907 T > C	p.Ser303Pro	E8	Missense	1	0.20	0	0.00	0	0.00	0.6446
c.910C > T	p.Gln304*	E8	Nonsense	0	0.00	1	0.33	0	0.00	0.3381
c.911A > G	p.Gln304Arg	E8	Missense	1	0.20	0	0.00	0	0.00	0.6446
c.925G > A	p.Ala309Thr	E9	Missense	0	0.00	1	0.33	0	0.00	0.3381
c.965C > A	p.Ala322Asp	E9	Missense	0	0.00	1	0.33	0	0.00	0.3381
c.969 + 4A > T	p.?/IVS9 + 4A > T	I9	Splice	1	0.20	0	0.00	0	0.00	0.6446
c.992 T > C	p.Phe331Ser	E10	Missense	1	0.20	0	0.00	0	0.00	0.6446
c.1200-?_1315 +?dup	p.?/EX12dup	E12	Large Duplication	0	0.00	0	0.00	1	0.69	0.0607
c. 510-6708_706 + 888 del7793insGG	p.?/EX6del7793 insGG	E6	Large Deletion	0	0.00	0	0.00	1	0.69	0.0607
c.169-?_c.352 +?del	p.?/EX3del	E3	Large Deletion	0	0.00	1	0.33	0	0.00	0.3381
Undetected mutations				19	3.75	19	6.33	16	11.11	0.0029

#, significant difference among the three PKU phenotypes;?, Uncertain protein change
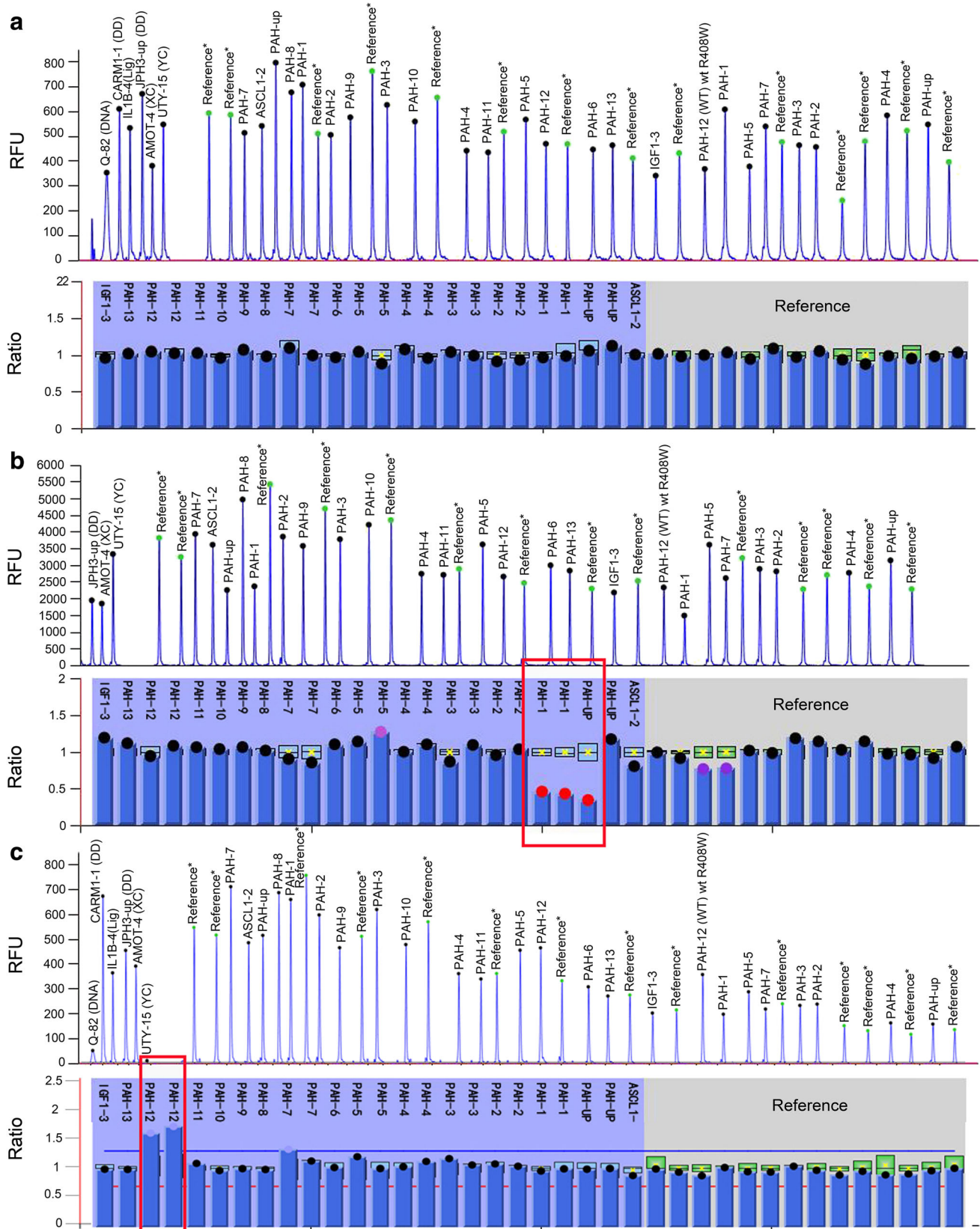
Fig. 1 Analysis of deletions and duplications detected by MLPA. a, normal sample; **b,** deletion in exon1 upstream region (D2); **c,** exon 12 duplication mutation (Ex12dup). The ratios range of normal reference is 0.7–1.3; the ratios indicate 3 copies in the range of 1.3–1.7, which is judged to be a heterozygous duplication mutation; the ratios indicate a single copy from 0.3 to 0.7, which is judged to be a heterozygous deletion mutation; the ratios indicate homozygous at 0 to 0.3 and is judged to be a homozygous deletion mutation

Correlation between variants and phenotype in patients with PKU

The detection rates of variants in the three types of patients (classic PKU, intermediate PKU, and MHP) were 96.25%, 94.79%, and 88.89%, respectively ($P < 0.05$). The frequencies of each *PAH* variants were compared among the three phenotype patients. We used the variants with frequency more than 1% for this comparison. The distributions of the 7 variant sites with the highest frequencies were significantly different among the three types of patients: p.Arg243Gln ($P = 0.0057$), IVS4-1G > A ($P = 0.0323$), p.Arg241Cys ($P < 0.0000$), p.Arg111* ($P = 0.0208$), p.Arg53His ($P < 0.0000$), p.Thr418Pro ($P = 0.0003$), and p.Gln419Arg ($P < 0.0000$). The p.Phe392Ile and p.Val230Ile variants were only detected in patients with MHP. The p.Leu430Pro variant was only detected in moderate PKU patients. The p.His107Arg variant was detected only in MHP and moderate PKU patients. Finally, the frequencies of the p.Arg53His and p.Gln419Arg variants were clearly higher in MHP patients than in the other two types of PKU patients (Table 1).

Discussion

A large cohort of 475 PKU patients detected by newborn screening in the past 10 years were recruited, a variant detection rate of 94.21% (95% CI, 92.72%–95.70%) confirms the pervasiveness of genetic mutations in the *PAH* gene. Variants in six exons accounted for 88.61% of the total, which are also variantal hot spots in the other regions of China (Li et al. 2015; Song et al. 2005). Among the 128 variants detected in this study, those with the highest frequencies are consistent with the results from other large-scale studies in Chinese populations, with the p.Arg243Gln and EX6-96A > G variants being the most common (Gao et al. 2011; Guo et al. 2011). A novel finding of our study is the p.Ile65Thr variant, which had a relatively high frequency of occurrence. The similarity of our results with those found in previous reports from Shaanxi (Qiang et al. 2014), Ningxia (Zhang et al. 2015), Qinghai (He et al. 2015), and Gansu (Yan et al. 2009) regions lay a theoretical foundation for the next stage of early screening.

One deletion variant (c.1-4163_c.1-406del3758) was found in compound heterozygous patients, having the same previously reported breakpoints (Lu et al. 2011). Chen et al. (Chen et al. 2002) found

that a liver-specific enhancer is located in this deleted region. *PAH* transcriptional activity is severely damaged because of this deletion and results in the classic PKU phenotype. Our results showed that this harmful variant accounted for 57.14% (16/27) of the large-scale deletion variants and is the most common large-scale deletion variant in the Northwest Chinese population. A potentially unique type of variant in the Northwest of China was identified in this study: a deletion variant (Exon6del7793insGG). This variant results in a codon deletion in exon 6 of the *PAH* gene and results in a truncated protein without the key catalytic region and directly impacts *PAH* function.

Our predictions showed that the original splice mutation in exon 9 was disrupted by IVS9 + 4A > T resulting in abnormal splicing and classical PKU, but other two variants (c.843-6 T > C and c.970-3C > T) predicted not to affect splicing occurred in classical PKU and MHP patients, respectively. Because differences existed between the results predicted by pathogenicity scoring system and the phenotype, and given that splicing occurs in vivo, functional studies regarding splice site variants are still necessary. Although this study has further demonstrated the different detection rates of *PAH* gene variants among the three types of PKU patients ($P < 0.05$) (Bercovich et al. 2008; Zhao et al. 2016). Most of the MHP patients were compound heterozygous patients with one mutation in *PAH* gene. Transient hyperphenylalaninemia caused by a slight decrease in enzyme activity due to a single polymorphism, young age, liver immaturity, or other unknown cause remains to be confirmed by further follow-up results.

The p.Arg243Gln variant, a severe type of variant with 13% residual enzyme activity (Pey et al. 2007), was found as the main variant in classical PKU with 16.01% of variant frequency in this study. We also found that the frequency of the p.Arg53His variant in patients with MHP was 13.89%, which was the variant with the highest frequency in this type of patient as same as the results from another large sample study (Zhao et al. 2016). The results of in vitro expression experiments showed that the residual enzyme activity of p.Arg53His variant was 79%, and a high phenylalanine metabolism ability was obtained in patients with p.Arg53His and other mutations (Polak et al. 2013). It is reported that multiple patients have p.Arg53His in cis with other pathogenic variants (Dobrowolski et al. 2011), and some researchers believe that the variant may be a polymorphism in the normal population (Lu et al. 2011).
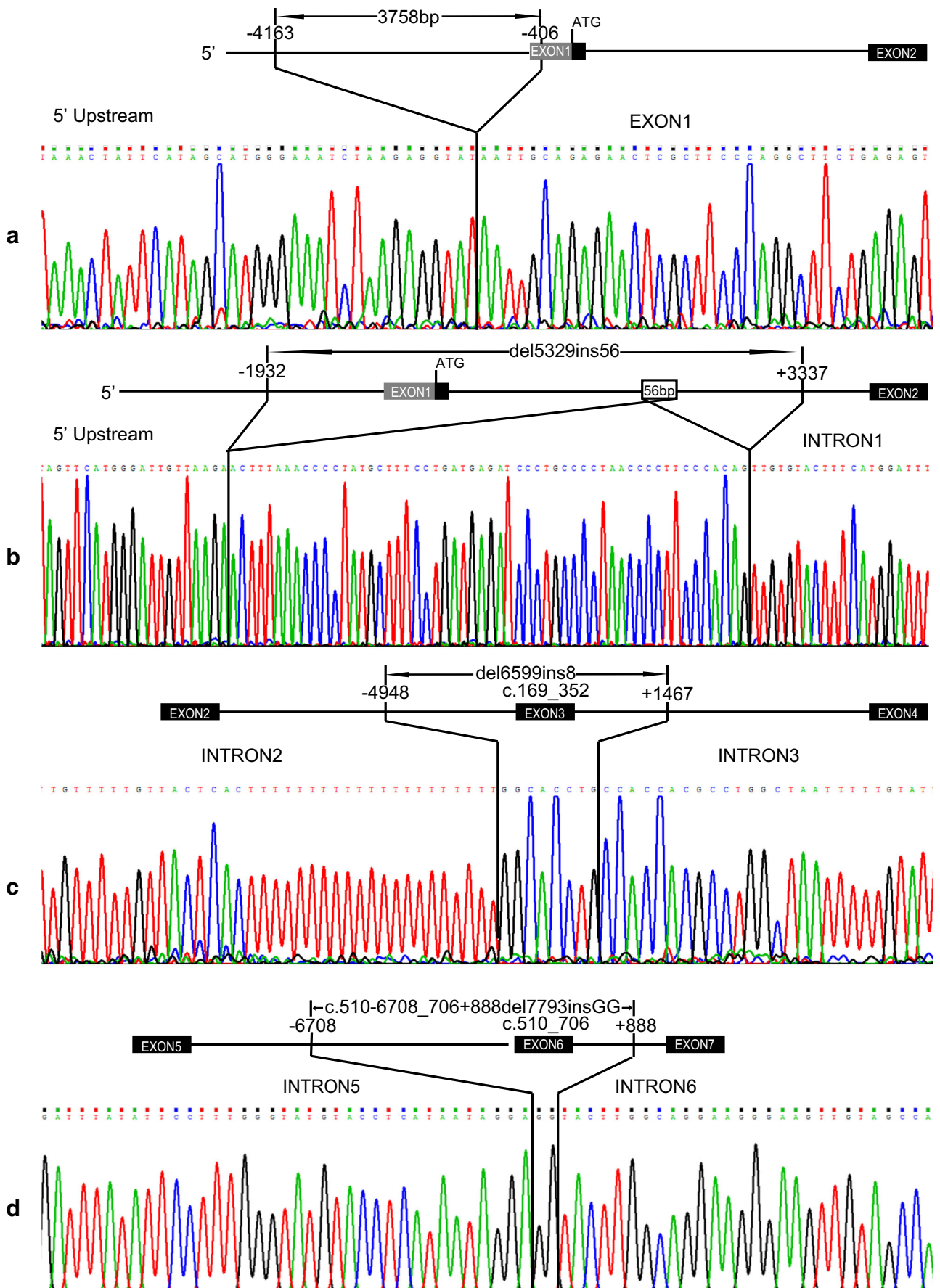
◄ **Fig. 2 Large-scale mutation breakpoint analysis. a**, deletion variant in Exon 1 region (Ex1del3758); **b**, deletion variant in Exon 1 region (Ex1 del5329ins56); **c**, deletion variant in Exon 3 region (Ex3 del6599ins8); **d**, deletion variant in Exon 6 region (Exon6del7793insGG)

In summary, the Sanger sequencing combined with MLPA detection technology demonstrated the PAH gene mutation spectrum in Northwest China, enriched the PAH gene mutation library, and laid a theoretical foundation for the next step of population carrier screening. At the same time, the comparison of variant frequencies among patients with phenotypes of different severity further clarified the correlation between compound heterozygosity and phenotype, and provided the basis for phenotypic prediction.
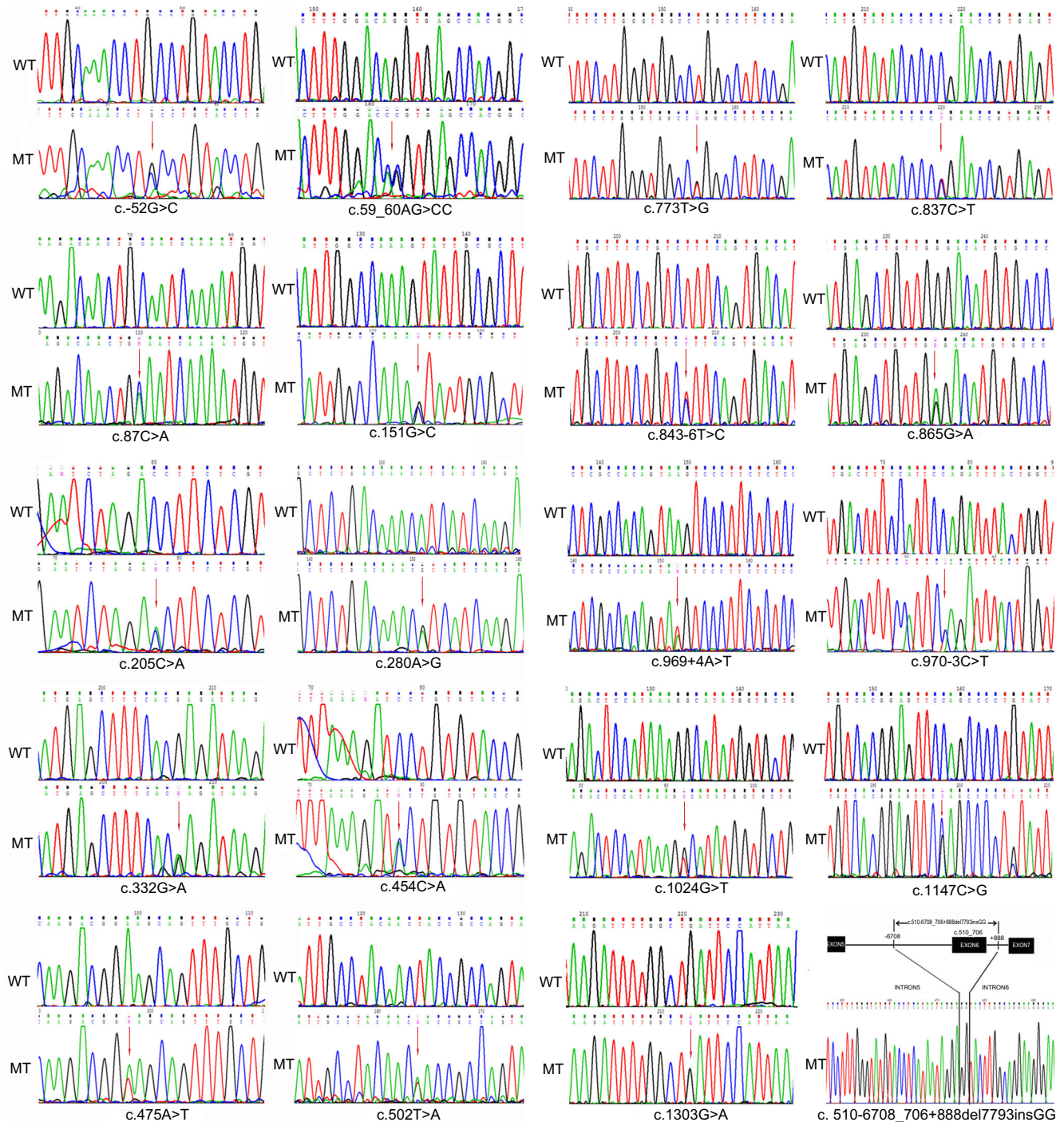
Fig. 3 Sequencing results for novel variants. WT: wild type; MT: mutant type

Table 2 Functional predictions of novel mutations detected in the *PAH* gene

Number	Mutation	Amino acid change	Classification[#]
1	c.-52G > C	p.?	Uncertain significance(PM1 + PM2)
2	c.59_60AG > CC	p.Gln20Pro	Pathogenic (PVS1 + PM2 + PP4)
3	c.87C > A	p.Cys29*	Likely pathogenic(PVS1 + PM2)
4	c.151G > C	p.Val51Leu	Uncertain significance(PM1 + PM2 + PP3)
5	c.205C > A	p.Pro69Trp	Uncertain significance(PM1 + PM2 + PP3)
6	c.280A > G	p.Ile94Val	Uncertain significance(PM1 + PM2 + PP3)
7	c.332G > A	p.Arg111Gln	Uncertain significance(PM1 + PM2 + PP3)
8	c.454C > A	p.Pro152Thr	Uncertain significance(PM1 + PM2 + PP3)
9	c.475A > T	p.Lys159*	Pathogenic(PVS1 + PM2 + PP3)
10	c.502 T > A	p.Tyr168Asn	Likely pathogenic(PM1 + PM2 + PM5 + PP3)
11	c. 510-6708_706 + 888 del7793insGG	p.?/ EX6del7793 insGG	Pathogenic(PVS1 + PM2 + PM4)
12	c.773 T > G	p.Leu258Arg	Uncertain significance(PM1 + PM2 + PP3)
13	c.837C > T	p.Pro279=	Likely benign(PM2 + BP4 + BP7)
14	c.843-6 T > C	p.?/ IVS7-6 T > C	Uncertain significance(PM2 + BP4)
15	c.865G > A	p.Gly289Arg	Uncertain significance(PM1 + PM2 + PP3)
16	c.969 + 4A > T	p.?/ IVS9 + 4A > T	Likely pathogenic (PM1 + PM2 + PM3 + PP3)
17	c.970-3C > T	p.?	Uncertain significance(PM2 + BP4)
18	c.1024G > T	p.Ala342Ser	Uncertain significance(PM1 + PM2 + PP3)
19	c.1147C > G	p.Gln383Glu	Uncertain significance(PM1 + PM2 + PP3)
20	c.1303G > A	p.Asp435Asn	Uncertain significance(PM1 + PM2 + PP3)

[#] Classification of novel mutations according to ACMG guideline (Richards et al. 2015);?, Uncertain protein change

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in this study.

References

Bercovich D, Elimelech A, Zlotogora J, Korem S, Yardeni T, Gal N, Goldstein N, Vilensky B, Segev R, Avraham S, Loewenthal R, Schwartz G, Anikster Y (2008) Genotype-phenotype correlations analysis of mutations in the phenylalanine hydroxylase (PAH) gene. J Hum Genet 53(5):407–418

Blau N, van Spronsen FJ, Levy HL (2010) Phenylketonuria. Lancet 376(9750):1417–1427

Chen KJ, Chao HK, Hsiao KJ, Su TS (2002) Identification and characterization of a novel liver-specific enhancer of the human phenylalanine hydroxylase gene. Hum Genet 110(3):235–243

den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, Roux AF, Smith T, Antonarakis SE, Taschner PEM, on behalf of the Human Genome Variation Society (HGVS), the Human Variome Project (HVP), and the Human Genome Organisation (HUGO) (2016) HGVS recommendations for the description of sequence variants: 2016 update. Hum Mutat 37(6):564–569

Desviat LR, Perez B, Ugarte M (2006) Identification of exonic deletions in the PAH gene causing phenylketonuria by MLPA analysis. Clin Chim Acta 373(1–2):164–167

Dobrowolski SF, Heintz C, Miller T, Ellingson C, Ellingson C, Özer I, Gökçay G, Baykal T, Thöny B, Demirkol M, Blau N (2011) Molecular genetics and impact of residual in vitro phenylalanine hydroxylase activity on tetrahydrobiopterin responsiveness in Turkish PKU population. Mol Genet Metab 102(2):116–121

Gao WH et al (2011) Study on the mutations of phenylalanine hydroxylase gene in patients with phenylketonuria in Shanxi Province. Chin J Med Genet 28(4):393–396

Giardine B, Riemer C, Hefferon T, Thomas D, Hsu F, Zielenski J, Sang Y, Elnitski L, Cutting G, Trumbower H, Kern A, Kuhn R, Patrinos GP, Hughes J, Higgs D, Chui D, Scriver C, Phommarinh M, Patnaik SK, Blumenfeld O, Gottlieb B, Vihinen M, Väliaho J, Kent J, Miller W, Hardison RC (2007) PhenCode: connecting ENCODE data with mutations and phenotype. Human Mutat 28(6):554–562

Gu XF, Wang ZG (2004) Screening for phenylketonuria and congenital hypothyroidism in 5.8 million neonates in China. Chin J Prev Med 38(2):99–102

Guo HJ, Zhao ZH, Jiang M, Shi HR, Kong XD (2011) Mutation analysis of phenylalanine hydroxylase gene in patients with phenylketonuria in Henan Province. Chin J Med Genet 28(2):142–146s

He J, Xu FL, Wang HZ, Yang X, Wang R, Zou HY, Yu WZ (2015) Mutation analysis of the phenylalanine hydroxylase gene in patients with phenylketonuria in Han ethnic group of Qinghai Province, China. Chin J Reprod Heal 26(6):512–517

Hubbard TD, Murray IA, Bisson WH, Sullivan AP, Sebastian A, Perry GH, Jablonski NG, Perdew GH (2016) Divergent ah receptor ligand selectivity during hominin evolution. Mol Biol Evol 33(10):2648–2658

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 (2015) Molecular characterisation of phenylketonuria in a Chinese mainland population using next-generation sequencing. Sci Rep 5:15769

Lu CX et al (2011) Mutation analysis of phenylalanine hydroxylase gene in 55 patients with phenylketonuria from Hebei Province. Natl Med J China 91(42):2971–2976

Pey AL, Stricher F, Serrano L, Martinez A (2007) Predicted effects of missense mutations on native-state stability account for phenotypic outcome in phenylketonuria, a paradigm of misfolding diseases. Am J Hum Genet 81(5):1006–1024

Polak E et al (2013) Phenylalanine hydroxylase deficiency in the Slovak population: genotype-phenotype correlations and genotype-based predictions of BH4-responsiveness. Gene 526(2):347–355

Qiang R et al (2014) Mutations of the phenylalanine hydroxylase gene in phenylketonuria patients from Shaanxi. Chin J Med Genet 31(1):74–77

Richards S et al (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17(5):405–424

Song F, Qu YJ, Zhang T, Jin YW, Wang H, Zheng XY (2005) Phenylketonuria mutations in northern China. Mol Genet Metab 86(Suppl 1):S107–S118

Song W, Gardner SA, Hovhannisyan H, Natalizio A, Weymouth KS, Chen W, Thibodeau I, Bogdanova E, Letovsky S, Willis A, Nagan N (2016) Exploring the landscape of pathogenic genetic variation in the ExAC population database: insights of relevance to variant classification. Genet Med 18(8):850–854

Sudmant PH et al (2015) An integrated map of structural variation in 2, 504 human genomes. Nature 526(7571):75–81

Wang X, Hao SJ, Cheng PL, Feng X, Yan YS (2015) Analysis on screening results of phenylketonuria among 567691 neonates in Gansu Province. Int J Lab Med 36(24):3588–3590

Woo SL, Lidsky AS, Guttler F, Chandra T, Robson KJ (1983) Cloned human phenylalanine hydroxylase gene allows prenatal diagnosis and carrier detection of classical phenylketonuria. Nature 306(5939):151–155

Yang YL, Ye Y (2014) Consensus about the diagnosis and treatment of hyperphenylalaninemia. Chin J Pediatr 52(6):420–425

Yan YS, Wang Z, Hao SJ, Meng Y, Zheng L, Huang SZ (2009) Mutation analysis of the PAH gene in patients with phenylketonuria in Gansu province. Chin J Med Genet 26(4):419–422

Zastrow DB, Baudet H, Shen W, Thomas A, Si Y, Weaver MA, Lager AM, Liu J, Mangels R, Dwight SS, Wright MW, Dobrowolski SF, Eilbeck K, Enns GM, Feigenbaum A, Lichter-Konecki U, Lyon E, Pasquali M, Watson M, Blau N, Steiner RD, Craigen WJ, Mao R, ClinGen Inborn Errors of Metabolism Working Group (2018) Unique aspects of sequence variant interpretation for inborn errors of metabolism (IEM): the ClinGen IEM working group and the phenylalanine hydroxylase gene. Hum Mutat 39(11):1569–1580

Zhang NJ, Yan YS, Dai NL, Wang DF, Wang G, Jiao HY (2015) Mutational spectrum of phenylalanine hydroxylase gene in Ningxia patients with phenylketonuria. J Appl Clin Pediatr 30(20):1557–1560

Zhao DH, Li XL, Jia CL, Ni M, Kong XD (2016) Genotype and phenotype correlation of phenylalanine hydroxylase deficiency among patients from Henan. Chin J Med Genet 33(3):300–305