

目标序列捕获二代测序技术在苯丙酮尿症 基因诊断中的应用

陈 瑛^{1Δ} 魏晓明² 王本敬³ 易 鑫² 毛 君⁴

(¹ 南京医科大学附属苏州医院新生儿科-中心实验室,³ 妇科,⁴ 产科 苏州 215002;² 深圳华大基因研究院 深圳 518083)

【摘要】 目的 评估目标序列捕获二代测序技术在苯丙酮尿症(phenylketonuria, PKU)基因诊断中的应用价值。**方法** 采用目标区序列捕获及第二代测序技术对9例经典型PKU患者的苯丙氨酸羟化酶(phenylalanine hydroxylase, PAH)基因进行检测;采用Sanger测序技术对患者及其父母基因突变型进行验证。**结果** 9例患者检测到PAH基因的12种致病性突变,包括7种错义突变(p. I65T、p. F161S、p. Q204C、p. R241C、p. L242F、p. R243Q和p. Q375E),2种无义突变(p. R111X和p. Y356X),3种剪接突变[c. 442-1G>A(IVS4-1G>A)、c. 1315+6T>A(IVS12+6T>A)和c. 1316-2A>C(IVS12-2A>C)],以及3种非致病性的变异(p. Q232Q、p. V245V和p. L385L),致病性突变均来自患者父母。其中,2个致病性突变未见报道,分别为c. 1316-2A>C和p. Q375E(CAA->GAA)。**结论** 目标序列捕获二代测序可以准确检测出PAH基因突变,对遗传病因明确的疾病具有一定的临床应用价值。

【关键词】 苯丙酮尿症; 苯丙氨酸羟化酶; 序列捕获; 第二代测序技术

【中图分类号】 R 596. 2; R 394 **【文献标志码】** A **doi:** 10. 3969/j. issn. 1672-8467. 2015. 03. 021

Application of captured high throughput sequencing in the genetic diagnosis of phenylketonuria

CHEN Ying^{1Δ}, WEI Xiao-Ming², WANG Ben-Jing³, YI Xin², MAO Jun⁴

(¹ Department of Neonatology-Central Lab, ³ Department of Gynecology, ⁴ Department of Obstetrics,
Nanjing Medical University Affiliated Suzhou Hospital, Suzhou 215002, Jiangsu Province, China;

² Beijing Genomics Institute at Shenzhen, Shenzhen 518083, Guangzhou Province, China)

【Abstract】 Objective Assess the performance of captured high throughput sequencing in the genetic diagnosis of phenylketonuria (PKU). **Methods** Target region sequence capture and next-generation sequencing were used to detect phenylalanine hydroxylase (PAH) gene of 9 children affected with typical PKU. Sanger sequencing was used to confirm the results of the patients and their parents.

Results Nine children carried 12 pathogenicity mutations in PAH gene, including 7 missense mutation (p. I65T, p. F161S, p. Q204C, p. R241C, p. L242F, p. R243Q and p. Q375E), 2 nonsense mutation (p. R111X and p. Y356X), 3 splicing mutation [c. 442-1G>A(IVS4-1G>A), c. 1315+6T>A(IVS12+6T>A) and c. 1316-2A>C(IVS12-2A>C)] and 3 unpathologic silence mutations (p. Q232Q, p. V245V and p. L385L). All the mutations were from their parents, in which 2 novel pathogenicity mutations were not reported before, including c. 1316-2A>C and p. Q375E (CAA->GAA). **Conclusions** Next generation sequencing technology is a useful tool for the detection of PAH gene, which is valuable for

江苏省“十二五”科教兴卫工程医学重点人才资助项目(RC2011036);苏州市生殖与遗传转化医学重点实验室建设项目(SZS201206)

ΔCorresponding author E-mail: cyandzh@sohu. com

diagnosing diseases of genetic etiology confirmed in clinical application.

【Key words】 phenylketonuria; phenylalanine hydroxylase; target sequence capture sequencing; next-generation sequencing

* This work was supported by the "Twelfth Five-Year Plan" Key Medical Talents Project in Science and Education of Jiangsu Province (RC2011036) and Suzhou Key Lab of Translational Medicine Project (SZS201206).

序列捕获测序是通过一系列寡核苷酸探针与全基因组 DNA 杂交,从而将基因组上的特定区段富集起来,并通过第二代测序技术进行测序^[1]。将人目标区基因全部外显子捕获下来,只需针对目标区中外显子区域的 DNA 即可,包含合成蛋白质所需要的重要信息,涵盖了与个体表型相关的大部分功能性变异,因此目标区序列检测远比进行全基因组序列测序更简便、经济,已成为现阶段第二代测序工作的重心。采用该技术的报道还不多,仅在假肥大型肌营养不良、眼-耳-肾综合征(Alport's syndrome)、Meckel 综合征、II 型黏多糖贮积症、尿素循环代谢障碍等少数疾病检测中有报道^[2-6]。在苯丙酮尿症(phenylketonuria, PKU)基因诊断中的应用尚无报道。本试验采用目标区基因外显子定点捕获的方式结合高通量测序技术,对 9 例 PKU 患者进行检测,探索目标区序列捕获及第二代测序技术在 PKU 基因诊断方面的临床应用价值。

资料和方法

研究对象 2009 年至 2012 年在苏州市立医院新生儿疾病筛查中心进行新生儿遗传代谢病筛查后经串联质谱确诊的 PKU 患儿 9 例,年龄 1 个月至 1 岁,男 4 例,女 5 例。

杂交文库制备 家属签署知情同意后,取患者外周血 200 μ L,用 QIAamp DNA Blood Mini Kit 试剂盒(德国 Qiagen 公司)提取外周血中的基因组 DNA。取 1 μ g DNA,采用 Covaris S2 超声打断仪打断至 200~300 bp。打断后的 DNA 片段按照美国 Illumina 公司提供的标准建库流程进行文库制备。取 3 μ g 制备好的 DNA 文库及 Cot-1 DNA 经 SpeedVac(美国 Thermo 公司)蒸干后与 2.1 mol/L 液相捕获芯片由瑞士 Nimblegen 公司商业个性化订制,包含苯丙氨酸羟化酶(phenylalanine hydroxylase, PAH)基因在内的多个基因捕获探针。杂交捕获目的 DNA 片段,72 h 后洗脱、纯化,通过

富集度检测合格(>60 nt)的文库使用 Hiseq 2500 平台上机测序。

信息分析 使用 Illumina Pipeline (V1.3.4)对图像分析和错误评估处理,获取初始碱基数据,采用 BWA 软件包将过滤后的读序比对至 NCBI 数据库(Build 37)参考序列上,使用 SOAPsnp 和 GATK Indel 软件分别对单核苷酸多态性(SNP)位点和小的插入及缺失(Indel)进行数据分析。为了评估新发现错义突变的致病性,我们使用 164 个中国人的 PAH 基因 SNP 频率信息,将频率<0.05 的变异视为可疑突变,为了进一步确认该变异引起的氨基酸替换对蛋白结构和功能的影响,我们采用 SIFT 在线预测工具(<http://sift.jcvi.org/>)以及 PolyPhen-2 软件(Polymorphism Phenotyping v. 2, <http://genetics.bwh.harvard.edu/pph2>)对其致病性进行预测。蛋白序列从 NCBI (<http://www.ncbi.nlm.nih.gov/protein>)调取用于比较分析。

Sanger 测序与家系成员验证 采用 Sanger 法对 PAH 基因 13 个外显子测序为本实验室常规技术,引物由上海捷瑞公司合成,PCR 反应条件:95 $^{\circ}$ C 预变性 10 min;95 $^{\circ}$ C 变性 30 s,57 $^{\circ}$ C 或 58 $^{\circ}$ C 退火 30 s,72 $^{\circ}$ C 延伸 30 s,35 个循环;72 $^{\circ}$ C 灭活 7 min;4 $^{\circ}$ C 下保存;此过程在 ABI 9700 PCR 仪上完成。测序过程在 ABI 3130 遗传分析仪上完成。对 9 例经典型 PKU 患者中的 3 例进行了 13 个外显子的测序,其余 6 例患者及 9 例患者的父母仅针对特定突变位点进行测序验证。

SNaPshot 验证 为了进一步确认特定突变在正常人群中的发生频率,采用混合样本单碱基延伸 SNP 基因分型技术(SNaPshot)进行验证。扩增引物为 5'-GGGCTGTGATGTAGAAGG-3' 和 5'-C-ACCTTTGTCACCACCTC-3'。延伸引物为 5'-TT-TTTCCTGGAGCTGGAGAAGACAGCCATC-3'。检测 500 个正常对照样本。PCR 产物纯化:取 1.0 μ L 扩增产物,加入 2.0 U 虾碱性磷酸酶(SAP 酶)、2.0 U Fermentas[®] Exon I 酶,反应总体积为 6 μ L,

混匀后 37 ℃ 反应 80 min,80 ℃ 灭活 15 min。PCR 反应体系为 6.0 μL,含纯化产物 1 μL、5 × Seq Buffer 1.2 μL、SNaPshot Multiplex Mix (美国 ABI 公司) 1.0 μL、延伸引物 1 μL。PCR 反应条件:96 ℃ 预变性 1 min;96 ℃ 变性 10 s,52 ℃ 退火 5 s,60 ℃ 延伸 30 s,28 个循环。延伸产物纯化:延伸反应产物中加入 1 U SAP 酶,37 ℃ 反应 60 min,75 ℃ 灭活 15 min。取 1 μL 纯化后产物进行毛细管电泳。

结 果

二代测序 外显子区域测序覆盖 13 个外显子,每个外显子碱基的平均测序深度在 131 × ~184 × ,覆盖度 100%。高通量测序结果显示,在 9 例 PKU 患者中 PAH 基因编码区及其侧翼区发现 7 种错义突变、2 种无义突变、3 种剪接突变和 3 种非致病性的变异(图 1)。其中 10 种是曾报道的致病性突变,2 种为首次发现的可疑致病性突变。

突变分析 突变分析使用 SOAPsnp 和 GATK Indel 软件分别对单核苷酸多态性(SNP)位点和小的插入及缺失(Indel)进行数据分析(表 1)。采用

SIFT 和 PolyPhen-2 软件针对 7 种错义突变对蛋白结构和功能的影响进行预测(表 1)。p. I65T、p. F161S、p. Q204C、p. R241C、p. L242F、p. R243Q 为致病性突变。在 3 号患者中检测到的 p. Q375E 为新发突变,NCBI 及 PAHdb 都未曾报道过这个突变,在 164 个中国人的 SNP 数据库中也未发现,然而软件分析提示该突变为耐受性的和良性的突变。我们还检测到 c. 442-1G>A(IVS4-1G>A)、c. 1315 + 6T>A(IVS12 + 6T>A)和 c. 1316-2A>C(IVS12-2A>C)3 种内含子剪切部位的突变,其中 c. 1316-2A>C 为首次报道。

Sanger 法测序验证 采用 Sanger 法对 3 例患者的 13 个外显子进行测序验证,结果与第二代测序一致,对 6 例患者及全部患者父母针对可疑或致病性突变相应的外显子或内含子序列进行双向测序验证,与第二代测序结果一致。患者的复合杂合突变均来自双亲(表 1)。

新突变发生频率检测 在 164 个中国人的 SNP 数据库中未发现 p. Q375E 突变,采用 SNaPshot 技术在 500 个正常人群中发现 1 例为 p. Q375E 杂合突变,整体检出率为 1.5‰(1/664)。

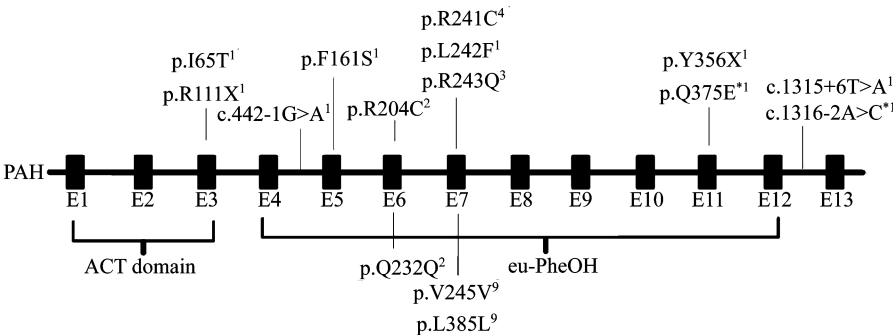


图 1 9 例 PKU 患者 PAH 基因的 12 种致病性突变和 3 种变异分布图

Fig 1 Distributions of 12 pathogenicity mutations and 3 unpathologic mutations in PAH gene of 9 patients with PKU

Pathogenicity mutations were above the line, and unpathologic mutations were under the line. p. Q375E and c. 1316-2A>C were novel mutations. The sample numbers detected were upper right the mutations.

讨 论

经典型 PKU 的分子病因是 PAH 基因突变,涉及 13 个外显子和内含子剪接序列突变,包括错义突变、无义突变、插入或缺失突变和剪接部位突变以及大片段缺失等^[7]。目标序列捕获二代测序结果显示,在 9 个 PKU 患者中发现 15 种 PAH 基因的突变,其中 7 种为错义突变,6 种错义突变经过软件预

测和以往文献报道均为致病性突变。本研究检测到的突变分布于外显子 3、5、6 和 7,其中外显子 7 上的突变最多,这与以往的外显子 7 为热区之一的报道一致^[8]。东方人 PKU 患者常见的突变位于外显子 12 和 7,分别为 p. R413P 和 p. R243Q,而欧美白种人 PKU 患者的常见突变位于外显子 12 和内含子 12,其中 p. R408Q 和 IVS12 + 1G>A 突变发生率较高。另外一个 p. Q375E(CAA->GAA)错义突变,导致谷氨酰胺被谷氨酸替代,NCBI 及 PAHdb

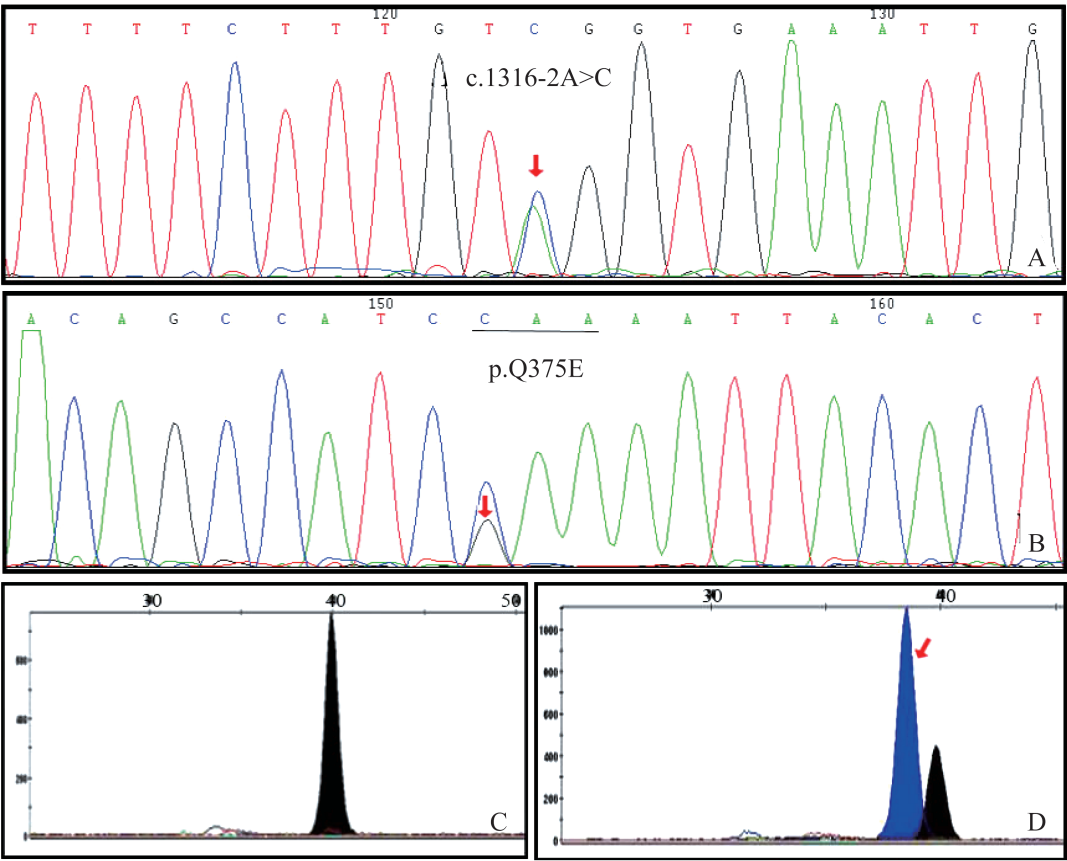


图 2 Sanger 测序及 SNaPshot 验证结果图

Fig 2 Verification results by Sanger sequencing and SNaPshot

A: Mutation of c. 1316-2A>C in P2; B: Mutation of p. Q375E (CAA-GAA) in P3; C and D: Results of p. Q375E detected by SNaPshot technology (C: Control; D: Heterozygous genotype of p. Q375E). Red arrow: The peak of mutation allele.

数据库都未曾报道过这个突变,164 个正常中国人 SNP 数据库中该位点没有变异,低频率携带率提示为可疑突变,但通过 SIFT 和 PolyPhen-2 软件对该突变进行致病性预测均显示为良性突变,为了进一步确定这两种矛盾的推断结果,我们对 500 个正常人群该位点进行基因分型显示该位点仅在 1 例正常人中携带,说明该变异在中国人群中属罕见型别,由于缺乏功能验证,其致病性有待进一步证实。

本研究发现 3 种剪接突变 c. 442-1G>A、c. 1315+6T>A 和 c. 1316-2A>C,其中 c. 442-1G>A 位于内含子 4,另两个位于内含子 12。剪接错误造成移码突变,将产生截断的 PAH 蛋白。此外,在 9 例患者中都存在 p. V245V 和 p. L385L 多态,通过与 NCBI 的 SNP 数据库比较,p. V245V 频率在不同、人种不同地域存在很大的差异,欧洲人群绝大部分为野生型,而亚洲人突变型占大部分,但不同地区有差异,p. L385L 频率在欧洲和亚洲人群中差

异不大,以突变型占绝大多数。在 2 例患者中检测到 p. Q232Q,而周永安等^[9]报道的山西省人群 p. Q232Q 突变型频率高达 96.2%,p. V245V 突变型频率达 76.1%,该突变在不同人种、不同地区的发生频率存在差异。目前,这些同义突变是否致病尚不清楚。

第二代测序技术是一种高通量测序技术,是对传统测序一次革命性的改变,具有高准确性、高通量、高灵敏度和低运行成本等突出优势^[10],尤其在全基因组突变位点筛查方面,第二代测序技术与常规 Sanger 测序技术相比,更加省时^[11]。相对于第一代测序技术,第二代测序不采用电泳,使用了反应信号实时阅读,即在测序反应的同时,就将反应信号收集起来,一次试验可以检测 40 万~2600 万条序列,可以读取 1~28 G 的碱基数,缩短了检测时间,并可将测序费用降低几个数量级。并且,第二代测序技术适合多样本、多外显子的同步筛查,多个样本

表 1 9 例 PKU 患者 PAH 基因的 15 种变异分析

Tab 1 Analysis of 15 mutations in PAH gene in 9 patients with PKU

Sample No.	Mutations	rs-ID	Exons	Fr. 164	SIFT	PolyPhen	Father	Mother
P1	p. R243Q	rs62508588	7	0	damaging	Probably damaging	p. Y356X	p. R243Q
	p. Y356X	rs62516095	11	0. 0061	—	—		
	p. V245V	rs1042503	11	0. 4695	—	—		
	p. L385L	rs772897	11	0. 3171	—	—		
P2	p. R241C	rs76687508	7	0	damaging	probably damaging	c. 1316 - 2A>C	p. R241C
	c. 1316-2A>C	novel	12	0	—	—		
	p. V245V	rs1042503	7	0. 4695	—	—		
	p. L385L	rs772897	11	0. 3171	—	—		
P3	c. 442 - 1G>A	rs62514907	4	0	—	—	p. Q375E	c. 442 - 1G>A
	p. Q375E	novel	11	0	tolerated	benign		
	p. Q232Q	rs1126758	7	0. 2134	—	—		
	p. V245V	rs1042503	7	0. 4695	—	—		
P4	p. L385L	rs772897	11	0. 3171	—	—	p. R241C	p. Y204C
	p. Y204C	rs62514927	6	0. 0061	damaging	probably damaging		
	p. R241C	rs76687508	7	0	damaging	probably damaging		
	p. V245V	rs1042503	7	0. 4695	—	—		
P5	p. L385L	rs772897	11	0. 3171	—	—	p. I65T	p. R241C
	p. I65T	rs75193786	3	0	damaging	probably damaging		
	p. R241C	rs76687508	7	0	damaging	probably damaging		
	p. V245V	rs1042503	7	0. 4695	—	—		
P6	p. L385L	rs772897	11	0. 3171	—	—	p. R243Q	p. Y204C
	p. Y204C	rs62514927	6	0. 0061	damaging	probably damaging		
	p. R243Q	rs62508588	7	0	damaging	probably damaging		
	p. V245V	rs1042503	7	0. 4695	—	—		
P7	p. V245V	rs1042503	11	0. 3171	—	—	c. 1315 + 6T>A	p. R243Q
	c. 1315 + 6T>A	rs62508650	7	0	—	—		
	p. R243Q	rs62508588	7	0	damaging	probably damaging		
	p. V245V	rs1042503	7	0. 4695	—	—		
P8	p. L385L	rs772897	11	0. 3171	—	—	p. F161S	p. L242F
	p. F161S	rs79635844	5	0	damaging	probably damaging		
	p. L242F	novel	7	0	damaging	probably damaging		
	p. V245V	rs104250	7	0. 4695	—	—		
P9	p. L385L	rs772897	11	0. 3171	—	—	p. R241C	p. R111X
	P. Q232Q	rs1126758	6	0. 2134	—	—		
	p. R111X	rs76296470	3	0	—	—		
	p. R241C	rs76687508	7	0	damaging	probably damaging		
	p. V245V	rs1042503	7	0. 4695	—	—		
	p. L385L	rs772897	11	0. 3171	—	—		

标记后与一个序列文库结合,在一个反应中进行目标区域富集,大大地降低了富集反应的成本^[12]。本试验初步尝试将 9 个样本用不同标签标记后,在一个反应孔中上机测序,相当于同时检测 117 个外显子的信息。通过该技术也可以实现 90 多个样本、数十万到上千万个片段的侧序分析,为基于人群范围的遗传病筛查奠定了基础,因此具有较大的临床应用潜能。

目前,第二代测序技术在临床应用方面所面临的主要问题是:第二代测序是基于短序列比对的原则,所以对于那些重复数差异所致的变异(如三核苷酸重复突变)难以准确检出;SNP 的识别结合了概率统计的原理,可能会导致少量假阳性结果的产生;同第一代测序技术一样,不能解决染色体结构变异、CNV 以及基因倒位重排等问题。因此,对于上述遗

传背景的疾病要寻找替代方案,如多重连接依赖的探针扩增技术和定量 PCR 技术解决拷贝数变异问题;凝胶电泳或 TA 克隆测序解决重复数变异问题,对于可疑的突变,一般要进行 Sanger 法测序验证,所以两代测序技术结合使用,可以发挥各自优势,满足临床准确性的需求。

我们尝试采用目标区序列捕获结合第二代高通量测序技术对 PKU 患者进行 PAH 基因突变检测,抽取了 9 例经典型 PKU 病例,运用二代测序技术对其 PAH 基因进行测序,成功找到了遗传病因,检测结果与 Sanger 测序法验证结果一致。同时对其中 3 例进行了 Sanger 全基因测序验证,也得到相同的结果,证明外显子捕获测序技术准确可靠,具有一定的临床应用前景^[6,13-15]。第二代测序技术要根据基因突变的特点与其他传统的检测技术结合使用,相互取长补短,可以满足临床上更经济、更准确、更快速的要求。

参 考 文 献

- [1] Mamanova L, Coffey AJ, Scott CE, *et al.* Target-enrichment strategies for next-generation sequencing[J]. *Nat Methods*, 2010, 7(2): 111 - 118.
- [2] Lim BC, Lee S, Shin JY, *et al.* Genetic diagnosis of Duchenne and Becker muscular dystrophy using next-generation sequencing technology: comprehensive mutational search in a single platform[J]. *J Med Genet*, 2011, 48(11): 731 - 736.
- [3] Artuso R, Fallerini C, Dosa L, *et al.* Advances in Alport syndrome diagnosis using next-generation sequencing[J]. *Eur J Hum Genet*, 2012, 1(20): 50 - 57.
- [4] Hopp K, Heyer CM, Hommerding CJ, *et al.* B9D1 is revealed as a novel Meckel syndrome (MKS) gene by targeted exon-enriched next-generation sequencing and deletion analysis[J]. *Hum Mol Genet*, 2011, 20(13): 2524 - 2534.
- [5] Wei XM, Jin F, Ye Y, *et al.* A novel mutation of IDS gene in a Chinese patient with mucopolysaccharidosis II by next-generation sequencing[J]. *Clin Chim Acta*, 2011, 412(23 - 24): 2340 - 2342.
- [6] Amstutz U, Andrey-Zürcher G, Suci D, *et al.* Sequence capture and next-generation resequencing of multiple tagged nucleic acid samples for mutation screening of urea cycle disorders[J]. *Clin Chem*, 2011, 57(1): 102 - 111.
- [7] Desviat LR, Pérez B, Ugarte M. Identification of exonic deletions in the PAH gene causing phenylketonuria by MLPA analysis[J]. *Clin Chim Acta*, 2006, 373(1 - 2): 164 - 167.
- [8] 宋昉, 金煜炜, 王红, 等. 苯丙氨酸羟化酶(PAH)基因外显子 7 及其两侧内含子的突变研究[J]. *遗传*, 2005, 27(1): 53 - 56.
- [9] 周永安, 李素云, 张改秀, 等. 苯丙酮尿症患者苯丙氨酸羟化酶 exon6 基因、exon7 基因突变的研究[J]. *中华临床医师杂志*, 2011, 5(1): 27 - 30.
- [10] Ansorge WJ. Next-generation DNA sequencing techniques[J]. *N Biotechnol*, 2009, 25(4): 195 - 203.
- [11] Smith DR, Quinlan AR, Peckham HE, *et al.* Rapid whole-genome mutational profiling using next-generation sequencing technologies[J]. *Genome Res*, 2008, 18(10): 1638 - 1642.
- [12] Kenny EM, Cormican P, Gilks WP, *et al.* Multiplex target enrichment using DNA indexing for ultra-high throughput SNP detection[J]. *DNA Res*, 2011, 18(1): 31 - 38.
- [13] Voelkerding KV, Dames S, Durtschi JD. Next generation sequencing for clinical diagnostics-principles and application to targeted resequencing for hypertrophic cardiomyopathy: a paper from the 2009 William Beaumont Hospital Symposium on Molecular Pathology[J]. *J Mol Diagn*, 2010, 12(5): 539 - 551.
- [14] Raca G, Jackson C, Warman B, *et al.* Next generation sequencing in research and diagnostics of ocular birth defects[J]. *Mol Genet Metab*, 2010, 100(2): 184 - 192.
- [15] Chou LS, Liu CS, Boese B, *et al.* DNA sequence capture and enrichment by microarray followed by next-generation sequencing for targeted resequencing: neurofibromatosis type 1 gene as a model[J]. *Clin Chem*, 2010, 56(1): 62 - 72.

(收稿日期: 2013-11-18; 编辑: 段佳)