The Journal of Clinical Endocrinology & Metabolism, 2022, **XX**, 1–9
https://doi.org/10.1210/clinem/dgac187
Advance access publication 1 April 2022
Clinical Research Article

Long-read Amplicon Sequencing of the *CYP21A2* in 48 Thai Patients with Steroid 21-Hydroxylase Deficiency

Nithiphut Tantirukdham,[1,2,*] Taninee Sahakitrungruang,[3,*] Ratikorn Chaisiwamongkol,[3] Monnat Pongpanich,[4,5] Chalurmpon Srichomthong,[6,7] Adjima Assawapitaksakul,[6,7] Aayalida Buasong,[6,7] Siraprapa Tongkobpetch,[6,7] Patra Yeetong,[8] and Vorasuk Shotelersuk[6,7]

[1]Genetics Program, Division of Human Genetics, Department of Botany, Faculty of Science, Chulalongkorn University, Bangkok, Thailand
[2]Molecular and Genomics Research Laboratory, Centre of Learning and Research in Celebration of HRH Princess Chulabhorn's 60th Birthday Anniversary, Chulabhorn Royal Academy, Bangkok, Thailand
[3]Division of Pediatric Endocrinology, Department of Pediatrics, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand
[4]Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Bangkok, Thailand
[5]Omics Sciences and Bioinformatics Center, Faculty of Science, Chulalongkorn University, Bangkok, Thailand
[6]Center of Excellence for Medical Genomics, Medical Genomics Cluster, Department of Pediatrics, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand
[7]Excellence Center for Genomics and Precision Medicine, King Chulalongkorn Memorial Hospital, the Thai Red Cross Society, Bangkok, Thailand; and
[8]Division of Human Genetics, Department of Botany, Faculty of Science, Chulalongkorn University, Bangkok, Thailand.

*These authors contributed equally to this work.
Correspondence: Patra Yeetong, PhD, Division of Human Genetics, Department of Botany, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand. E-mail: patra.y@chula.ac.th.

Abstract

Context: Congenital adrenal hyperplasia is most commonly caused by 21-hydroxylase deficiency (21-OHD), an autosomal recessive disorder resulting from biallelic pathogenic variants (PVs) in *CYP21A2*. With a highly homologous pseudogene and various types of single nucleotide and complex structural variants, identification of PVs in *CYP21A2* has been challenging.

Objective: To leverage long-read next-generation sequencing combined with locus-specific polymerase chain reaction (PCR) to detect PVs in *CYP21A2* and to determine its diagnostic yield in patients with 21-OHD.

Methods: Forty-eight Thai patients with 21-OHD comprising 38 sporadic cases and 5 pairs of siblings were enrolled. Two previously described locus-specific PCR methods were performed. Amplicons were subject to long-read sequencing.

Results: Ninety-six PVs in *CYP21A2* in the 48 patients were successfully identified. The combined techniques were able to detect 26 structural chimeric variants (27%; 26/96) in 22 patients with 18 having monoallelic and 4 having biallelic chimeras. The remaining PVs were pseudogene-derived mutations (63%; 60/96), entire gene deletions (2%; 2/96), missense variants (3%; 3/96), a splice-site variant (2%; 2/96), frameshift variants (2%; 2/96), and a nonsense variant (1%; 1/96). Notably, a splice-site variant, IVS7 + 1G > T, which was identified in a pair of siblings, has not previously been reported.

Conclusions: Our approach exploiting locus-specific PCR and long-read DNA sequencing has a 100% diagnostic yield for our cohort of 48 patients with 21-OHD.

Key Words: congenital adrenal hyperplasia, 21-hydroxylase deficiency, locus-specific polymerase chain reaction, long-read amplicon sequencing

Congenital adrenal hyperplasia (CAH) is a potentially fatal disease, most commonly caused by from 21-hydroxylase deficiency (21-OHD) leading to impaired cortisol synthesis (1). 21-OHD is an autosomal recessive disorder caused by pathogenic variants (PVs) in the *CYP21A2* gene, located on chromosome 6p21.33. *CYP21A2* has a highly homologous pseudogene-*CYP21A1P*, sharing 98% and 96% sequence identity in their exons and introns, respectively. The *CYP21A1P* is about 30 kb from the *CYP21A2* gene, adjacent to the HLA Class III region (2). Moreover, both genes are arranged in tandem loci with the *C4A* and *C4B* genes encoding the fourth component of the complement system. The RCCX module consists of *RP-C4-CYP21-TNX*. The most common RCCX haplotype is comprised of 2 segments with the genes *RP1-C4A-CYP21A1P-TNXA-RP2-C4B-CYP21A2-TNXB* arranged telomere-to-centromere (3). During meiosis, an active gene and a pseudogene might cross over unequally, resulting in the formation of nonfunctional chimeric genes (4). About 70% to 75% of disease-causing *CYP21A2* mutations are caused by the transfer of deleterious mutations from *CYP21A1P*, a process known as gene conversion (2).

Previously reported PVs in *CYP21A2* comprise several types including missense, nonsense, complex rearrangements, small/large deletions, conversions, splicing, and insertions (http://www.hgmd.cf.ac.uk/). With the complex gene arrangement and several possible types of PVs, genetic

Received: 21 September 2021. **Editorial Decision:** 22 March 2022. **Corrected and Typeset:** XX XXXX XXXX

testing designed to identify the *CYP21A2* PVs in patients with 21-OHD has been challenging. Attempts include allele-specific oligonucleotide hybridization (5-7), single-stranded conformation polymorphisms (8), allele-specific polymerase chain reaction (PCR) (9), ligation detection reaction (10), amplification-created restriction sites (5), denaturing high pressure liquid chromatography (11), minisequencing and multiplex minisequencing (12), southern blot (13), multiplex ligation-dependent probe amplification (14), real-time PCR (15), DNA Sanger sequencing (16), locus-specific PCR (17), and, recently, a short-read next-generation sequencing (NGS) assay (18-20). However, the short-read NGS assay for CAH is impeded by pseudogene homologue and commonly existing copy number variations across the RCCX locus. In fact, the copy number variations are a more common cause of mapping chaos in short-read NGS (21).

Long-read sequencing (LRS), a recently developed nucleotide sequencing technology, can sequence 10 kilo base-pairs (bps) to several mega bps in length compared to Sanger or short-read sequencing, which can sequence in the range of 100 bps. LRS therefore has a potential to read through and solve the complex regions of the human genome (22, 23) and detect difficult PVs such as structural variations, tandem repeat expansions, and pseudogenes (24).

In this study, we developed combined variant detection testing: locus-specific PCR and long-read amplicon sequencing and applied it to 48 Thai patients with 21-OHD. We successfully identified the 96 PVs, suggesting that the testing is a powerful tool to identify PVs in the complex *CYP21A2* gene.

Materials and Methods

Patients

Forty-eight patients (38 sporadic and 5 pairs of siblings) with clinical and hormonal findings suggesting 21-OHD were recruited from the pediatric endocrine clinic of the King Chulalongkorn Memorial Hospital. All were classic 21-OHD with an evidence of hyperandrogenism and cortisol insufficiency. They were subdivided into salt-wasting (SW) form, defined by hypoaldosteronism or a salt-losing crisis, and simple virilizing (SV) form if without evident hypoaldosteronism. The biochemical diagnosis of 21-OHD is based on an elevation of baseline or stimulated steroid intermediates 17-hydroxyprogesterone (17-OHP) level > 1000 ng/dL (1, 25) determined by the radioimmunoassay commercial kit (CISBIO, cat no. 17-OHP-CT, RRID:AB_2904505). This study was approved by the Institutional Review Board of Faculty of Medicine, Chulalongkorn University (COA no. 954/2016).

Locus-specific PCR

Genomic DNA was extracted from the peripheral blood using a Puregene Blood Kit (Qiagen, Hilden, Germany).

Locus-specific PCR Method 1

Primers CYP779f and Tena32F (26) were used to amplify a 8515 bp fragment using the PrimeSTAR GXL Taq DNA polymerase and the manufacturer's recommendations (Takara Bio Inc). The forward CYP779f primer (5'-AGGTGGGCTGTTTTCCTTTCA-3') could bind to the 5′ regions of both *CYP21A1P* and *CYP21A2*, while the Tena32F primer (5'-CTGTGCCTGGCTATAGCAAGC-3') binds specifically only to the intron 32 of *TNXB* (Fig. 1A). All the PCR products were detected by electrophoresis on a 1.0% agarose gel.

Locus-specific PCR Method 2

Four primers (ME0008, ME0066, ME0059, and ME0067) were used to PCR amplify the 4 possible amplicons (genuine gene, pseudogene, chimera, and rearrangement) [Supplementary Figure 1 (27)], as previously described (17). The *CYP21A1P/CYP21A2* chimera is a chimeric fusion of the 5' of the *CYP21A1P* pseudogene with the 3' of *CYP21A2*, which was identified using a *CYP21A1P* forward primer and a *CYP21A2* reverse primer combination. The *CYP21A2/CYP21A1P* rearrangement, on the other hand, is a combination of the 5′ region of *CYP21A2* and the 3′ region of the *CYP21A1P* pseudogene that was detected using the *CYP21A2* forward primer and the *CYP21A1P* reverse primer (17, 28). The PCR products of amplicon 1 was subjected to *Eco*RI (New England Biolabs, Ipswich, MA, USA) digestion for 1 hour at 37°C. All the PCR products were detected by electrophoresis on a 1.0% agarose gel.

Long-read Amplicon Sequencing of the Products From Locus-specific PCR

The 8.5-kb fragments from locus-specific PCR (LS-PCR) Method 1 using the primers CYP779f and Tena32F of all 48 patients were subjected to long-read amplicon sequencing. For the LS-PCR Method 2, only the nonchimeric and genuine *CYP21A2* (the 3.4-kb fragments of amplicon 1) of the 42 patients were subjected to long-read amplicon sequencing. Library preparation, sequencing, and data analysis of both methods were performed independently.

Amplicon SMRTbell templates and barcoding were prepared using PacBio® Barcoded Adapters for Multiplex SMRT® Sequencing with SMRTbell DNA Damage Repair samples. SMRTbell libraries were pooled and sequenced on 1M SMRT cell v3 LR with long-read single molecule real-time (SMRT) sequencing technology using the Sequel platform (Pacific Biosciences, Menlo Park, CA, USA). The single molecule real-time complex was generated by Sequel® Binding Kit 3.0 and then sequencing with Sequencing Plate 3.0 and Sequel® DNA Internal Control 3.0. To examine the overall data quality, LongQC (29) was used. After quality inspection, long amplicon analysis was performed to generate high-quality, phased consensus sequences from pooled amplicons with pblaa package (https://github.com/pacificbiosciences/pblaa/). The long amplicon analysis reads were mapped against human reference genome build 37 (GRCh37) by CoNvex Gap-cost alignMents for LRS (30). We extracted reads from the LRS aligned BAM that spanned chromosome 6, position 32005984-32009368 (GRCh37) to visually inspect and confirm variants using the Golden Helix GenomeBrowse® tool (Golden Helix, Montana, USA).

Results

Clinical Characterization of Patients With 21-OHD

Forty-eight patients (34 females, 14 males) with classic 21-OHD were included in the analysis. Thirty-six patients

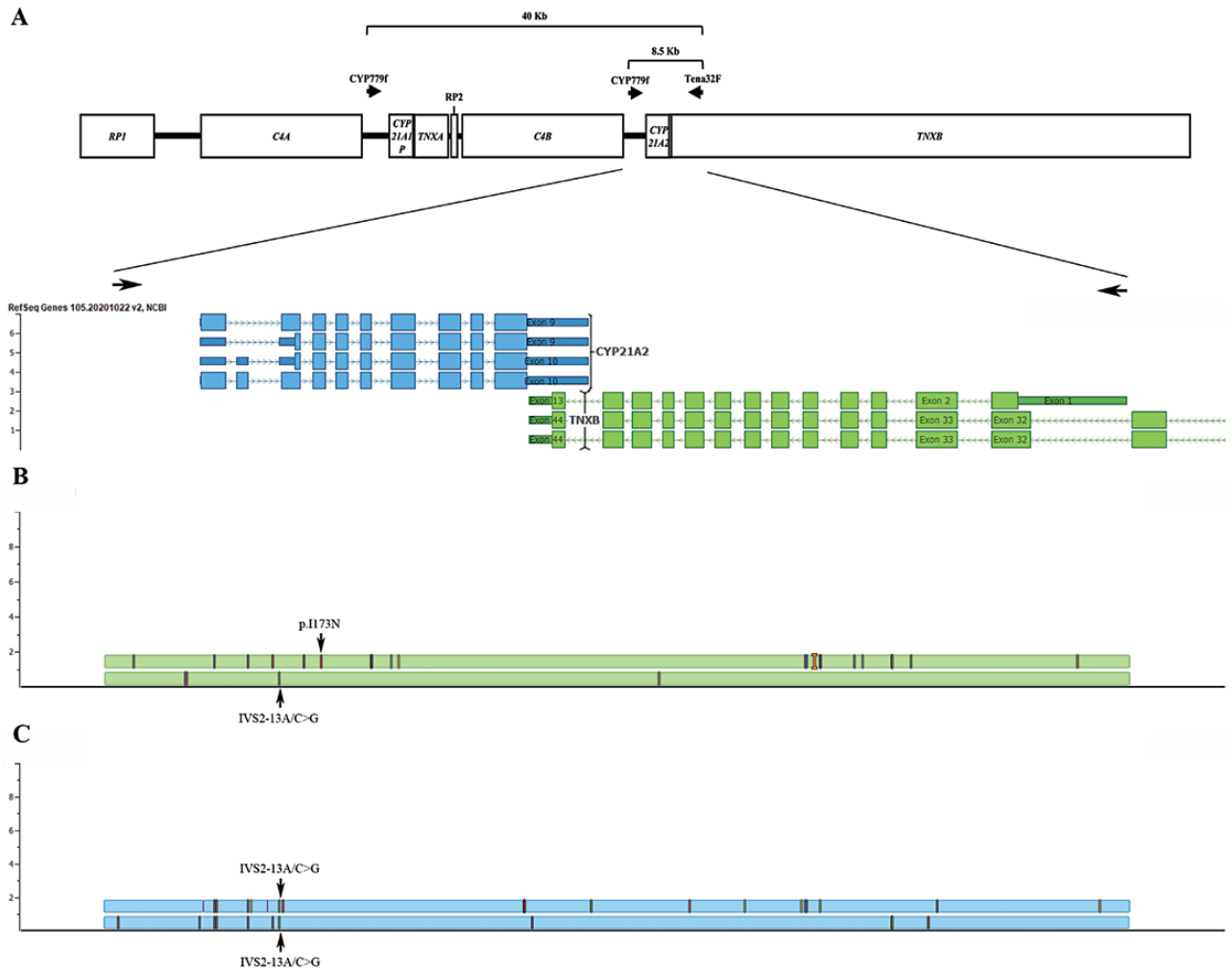

Figure 1. Findings of single nucleotide variants by long-read sequencing. (A) Schematic representation of the genes in the *CYP21A2-TNXB* region. The forward primer CYP779f can bind to the 5′ regions of both the *CYP21A1P* and *CYP21A2*. The reverse primer Tena32F specifically binds only to the intron 32 of *TNXB*. (B) The results of long amplicon analysis reads revealed 2 different 8.5-kb reads. Patient 1 was compound heterozygous for IVS2-13A/C > G and p.I173N, denoted by the black arrows. (C) Patient 3 was homozygous for IVS2-13A/C > G, yet having 2 different haplotypes, demonstrated by different variants between the 2 alleles in other parts of the genetic region.

(75%) had SW phenotypes; 12 (25%) had SV form. Since there is no 17-OHP newborn screening established in Thailand, most patients with classic 21-OHD (SW form) in our study presented with salt-losing crisis, hyperpigmentation, and ambiguous genitalia. The median age of diagnosis in the SW-CAH group was 1 month (range 1 day-7 months). By contrast, SV-CAH patients presented later with sexual precocity in boys and ambiguous genitalia in girls. The median age of diagnosis was 4 years (range 4-4.9 years) in SV-CAH boys and 2.2 years (range 14 days-9.1 years) in SV-CAH girls. Median serum levels of stimulated 17-OHP were 27 760 ng/dL (range 1 630-269 000 ng/dL).

Locus-specific PCR

For LS-PCR Method 1, DNA of 48 patients was amplified using CYP779f and Tena32F to produce the 8518-bp fragment [Supplementary Figure 2 (27)]. For LS-PCR Method 2, the DNA of all patients was amplified by 4 locus-specific PCR (amplicons 1, 2, 3, and 4). There were 42 patients with at least 1 unidentified mutant allele and at least 1 allele of the functional gene (*CYP21A2*) as determined by the presence of amplicon 1.

Long-read Amplicon Sequencing of the Entire *CYP21A2* Gene

The total output of the LS-PCR Method 1 was 34 098 838 163 bps and 7 036 270 reads. The total output of the LS-PCR Method 2 was 18 792 208 804 bps and 4 094 324 reads. After demultiplexing, the mean read depth was 7383 and 9696 in LS-PCR Methods 1 and 2, respectively. The overall data quality was investigated using LongQC. The majority of reads were close to the expected length of 8.5 kb for LS-PCR Method 1 and 3.4 kb for the LS-PCR Method 2. In both methods, the normal read's quality value is greater than 40 [Supplementary Figure 8 (27)]. The predicted accuracy in the generated consensus sequence was 0.97 to 1.0 and 0.91 to 1.0 in LS-PCR Methods 1 and 2, respectively. The total coverage in each cluster was between 23 and 494 in LS-PCR Method 1 and between 20 and 500 in LS-PCR Method 2 [Supplementary Tables 1 and 2 (27)].

Pseudogene-derived mutation, entire gene deletions, missense variants, splice site variant, frameshift variant, and nonsense variant were detected (Table 1). The most frequent mutation was pseudogene-derived *CYP21A2* mutation in the intron 2 of *CYP21A2* (IVS2-13A/C > G) (38%; 36/96 alleles)

Table 1. Clinical and molecular features of the 48 Thai patients with 21-OHD

Patient	Sex	Age of onset	Phenotype	Types of amplicons by the locus-specific PCR Method 2				Genotype		Group*
				1	2	3	4	Allele 1	Allele 2	
1	F	21 days	SV	+	+	–	–	IVS2-13A/C > G	p.I173N	B
2[a]	F	5 years	SV	+	+	–	–	p.I173N	p.I173N	B
3	F	11 days	SW	+	+	–	–	IVS2-13A/C > G	IVS2-13A/C > G	A
4	F	2 years, 2 months	SV	+	+	–	–	IVS2-13A/C > G	p.I173N	B
5	F	1 month	SW	+	+	–	–	IVS2-13A/C > G	IVS2-13A/C > G	A
6	F	1 day	SW	+	+	–	–	p.I173N; p.I237N; p.V238E; p.M240K; p.V282L; p.L308Ffs; p.Q319X; p.R357W	p.R357W	Null
7	F	1 month, 21 days	SW	+	+	–	–	IVS2-13A/C > G	IVS2-13A/C > G	A
8	F	5 years	SV	+	+	–	–	IVS2-13A/C > G	p.I173N	B
9[b]	F	14 days	SW	+	+	–	–	IVS7 + 1G > T	TNX-CH-1	D
10[b]	M	20 days	SW	+	+	–	–	IVS7 + 1G > T	CYP-CH-6	D
11	F	7 days	SW	+	+	–	–	p.Q319X	p.R484P	C
12	M	12 days	SW	+	+	–	–	p.Q319X	CYP-CH-3	Null
13[c]	F	4 days	SW	+	+	–	–	IVS2-13A/C > G	IVS2-13A/C > G	A
14[c]	F	19 days	SW	+	+	–	–	IVS2-13A/C > G	IVS2-13A/C > G	A
15	M	4 years, 10 months	SV	+	+	–	–	IVS2-13A/C > G	p.R357W	A
16	F	7 days	SW	+	+	–	–	IVS2-13A/C > G	IVS2-13A/C > G	A
17	F	27 days	SW	+	+	–	–	IVS2-13A/C > G	IVS2-13A/C > G	A
18[d]	F	1 month	SV	+	+	–	–	IVS2-13A/C > G	p.I173N	B
19[d]	M	4 years	SV	+	+	–	–	IVS2-13A/C > G	p.I173N	B
20	F	14 days	SV	+	+	–	–	IVS2-13A/C > G	CYP-CH-1	A
21	F	7 days	SW	+	+	–	–	p.R484Pfs*58	TNX-CH-1	Null
22	F	1 month	SW	+	+	–	–	p.I173N	p.Q319X	B
23	M	4 years	SV	+	+	–	–	IVS2-13A/C > G	p.I173N	B
24	F	7 years, 6 months	SV	+	+	–	–	p.I173N	p.I173N	B
25	M	1 month	SW	+	+	–	–	IVS2-13A/C > G	p.Q319X	A
26	M	21 days	SW	+	+	–	–	IVS2-13A/C > G	p.R357W	A
27	M	1 month	SW	+	+	–	–	IVS2-13A/C > G	p.G293D	A
28	F	1 day	SW	+	+	–	–	p.R317X	TNX-CH-1	Null
29[e]	M	4 months	SW	+	+	–	–	IVS2-13A/C > G	Entire gene deletion	A
30[e]	F	8 days	SW	+	+	+	–	IVS2-13A/C > G	CYP-CH-1	A
31[e]	M	3 months	SW	+	+	+	–	IVS2-13A/C > G	CYP-CH-1	A
32	F	9 years, 2 months	SV	+	+	+	–	p.I173N	CYP-CH-1	B
33	F	1 month	SW	+	+	+	–	p.R357W	CYP-CH-1	Null
34	F	1 month	SW	+	+	+	–	IVS2-13A/C > G	CYP-CH-1	A
35	F	1 month	SW	+	+	+	–	IVS2-13A/C > G	CYP-CH-1	A
36	F	7 months	SW	+	+	+	–	p.Q319X; p.R357W	CYP-CH-5	Null
37	M	1 month	SW	+	+	+	–	IVS2-13A/C > G	CYP-CH-1	A
38	F	1 month	SW	+	+	+	–	IVS2-13A/C > G	CYP-CH-1	A
39	F	2 years	SV	+	+	+	–	IVS2-13A/C > G	CYP-CH-4	C
40	M	21 days	SW	+	+	+	–	IVS2-13A/C > G	CYP-CH-1	A
41[f]	F	1 month	SW	+	+	–	+	IVS2-13A/C > G	p.I173N; p.I237N; p.V238E; p.M240K; p.V282L; p.L308Ffs; p.R357W	A

Table 1. Continued

Patient	Sex	Age of onset	Phenotype	Types of amplicons by the locus-specific PCR Method 2				Genotype		Group*
				1	2	3	4	Allele 1	Allele 2	
42[f]	F	1 month	SW	+	+	–	–	IVS2-13A/C > G	p.I173N; p.I237N; p.V238E; p.M240K; p.V282L; p.L308Ffs; p.R357W	A
43	F	1 month	SW	-	+	+	–	CYP-CH-1	CYP-CH-8	Null
44	M	6 days	SW	-	+	+	–	CYP-CH-1	TNX-CH-1	Null
45	F	4 months	SW	-	+	+	–	CYP-CH-3	CYP-CH-7	Null
46[a]	F	1 day	SW	-	+	+	–	CYP-CH-8	CYP-CH-8	Null
47	F	1 month	SW	-	+	+	+	p.L167P	CYP-CH-2	A
48[a]	M	1 month	SW	-	+	–	+	p.R484Pfs*40	Entire gene deletion	Null

Asterisk (*) denotes the groups of residual enzyme activities: null-0; A-almost 0; B-severe reduction; C-partial reduction; D-new mutation.
Abbreviations: SV, simple virilizing; SW, salt-wasting.
[a]Denotes the 2 molecular possibilities. Due to the mutation detection methods used in this study, patients 2, 29, 46, and 48 could be either homozygous for the allele 1 or hemizygous (compound heterozygous for the allele 1 and an entire gene deletion). Since the patients 2 and 46 are products of consanguineous parents, a homozygous allele 1 is assigned, while patients 29 and 48 denied consanguinity, so hemizygosity is assigned.
[b], [c], [d], [e], and [f]: the same letters indicate siblings.

(Fig. 1). Three clusters of mutations were identified. Two clusters in exons 4, 6, 7, and 8, p.I173N; p.I237N, p.V238E; p.M240K; p.V282L; p.L308Ffs; p.Q319X; p.R357W and p.I173N; p.I237N, p.V238E; p.M240K; p.V282L; p.L308Ffs; p.R357W, were identified in 3 (3%; 3/96 alleles) alleles. The cluster in exon 8 (p.Q319X; p.R357W) was identified in 1 allele (1%; 1/96 alleles). Moreover, *CYP21A1P/CYP21A2* chimera types 1, 2, 3, 4, 5, 6, 7, and 8 were found in 22 alleles (23%, 22/96), and *TNXA/TNXB* chimera type 1 was identified in 4 alleles (4%, 4/96) (Fig. 2). Notably, a variant, IVS7 + 1G > T, found in a pair of siblings (patients 9 and 10) has not been previously described. Both were compound heterozygous for the novel IVS7 + 1G > T and an entire gene deletion from *TNX* and *CYP* chimeric genes, respectively.

LS-PCR Method 2 showed that 42 patients had positive amplicon 1, indicating the presence of at least 1 allele of the functional *CYP21A2* gene [Supplementary Figure 3 (27)]. These amplicons 1 were subjected to LRS. The same pseudogene-derived mutations, entire gene deletions, missense variants, splice site variant, frameshift variants, and nonsense variant as the Method 1 were identified [Table 1; also see Supplementary Table 3 (27)].

Discussion

CAH due to 21-OHD is one of the most common autosomal recessive disorders and is a life-threatening condition if undiagnosed, especially in its classic form (31). Hyperpigmentation and ambiguous genitalia are common presentations at diagnosis. In this cohort, female-to-male ratio is 2:1. In Thailand, there has not been nationwide newborn screening program for CAH yet. Thus, a number of patients, especially boys, had delayed diagnosis and treatment until toddlers or early childhood (32). Due to its nonspecific symptoms, the diagnosis may be missed and lead to mortality, especially in boys. Steroid precursor measurements (17-OHP) are the standard for diagnosing 21-OHD. Genetic testing is usually not used as a first-line diagnostic test due to the complexity of the *CYP21A2* locus.

The *CYP21A2* and its pseudogene, *CYP21A1P*, have a high degree of sequence similarity. Unequal crossing-over during meiosis could arise causing misalignment, which can result in gene deletions, duplications, and deletions affecting *CYP21A2* and other adjacent genes (33). Moreover, transfer of deleterious mutations presents in the pseudogene to *CYP21A2*, called microconversion, could occur. We used locus-specific PCR and then long-read amplicon sequencing to detect structural variations, microconversions, and other point mutations. LRS technology enables directly phase variants several kb apart (24). It provides information whether 2 PVs are in *trans* or *cis* configuration without the need of parents testing (Figs. 1 and 2). This is important for autosomal recessive disorders.

The unequal crossover during meiosis between an active gene and a pseudogene results in a 30-kb deletion and the formation of nonfunctional chimeric genes (*CYP21A1P/CYP21A2* or *TNXA/TNXB*). We were able to detect *CYP21A1P/CYP21A2* and *TNXA/TNXB* chimeras and identify their specific types (Fig. 2). In previous studies, approximately 14% to 15% of patients with 21-OHD have CAH-X syndrome (34-36). CAH-X syndrome is the clinical combination between 21-OHD and Ehlers-Danlos syndrome. It is related to the chimeric *TNXA/TNXB* gene. These gene located downstream of the *CYP21A2* encodes for the tenascin X (*TNX*). Four patients (8%; 4/48) in our cohort were found to have the *TNXA/TNXB* chimeric gene type 1 (Table 2). Interestingly, we discovered the *TNX* chimeric gene type 1 and the *CYP* chimeric gene type 1 in patient 44.

Notably, accurate LRS allows the identification of a different PV in a pair of siblings. Patients 9 and 10 were compound heterozygous for a single nucleotide variant, the IVS7 + 1G > T, and a *TNXA/TNXB* chimera and a *CYP21A1P/CYP21A2*

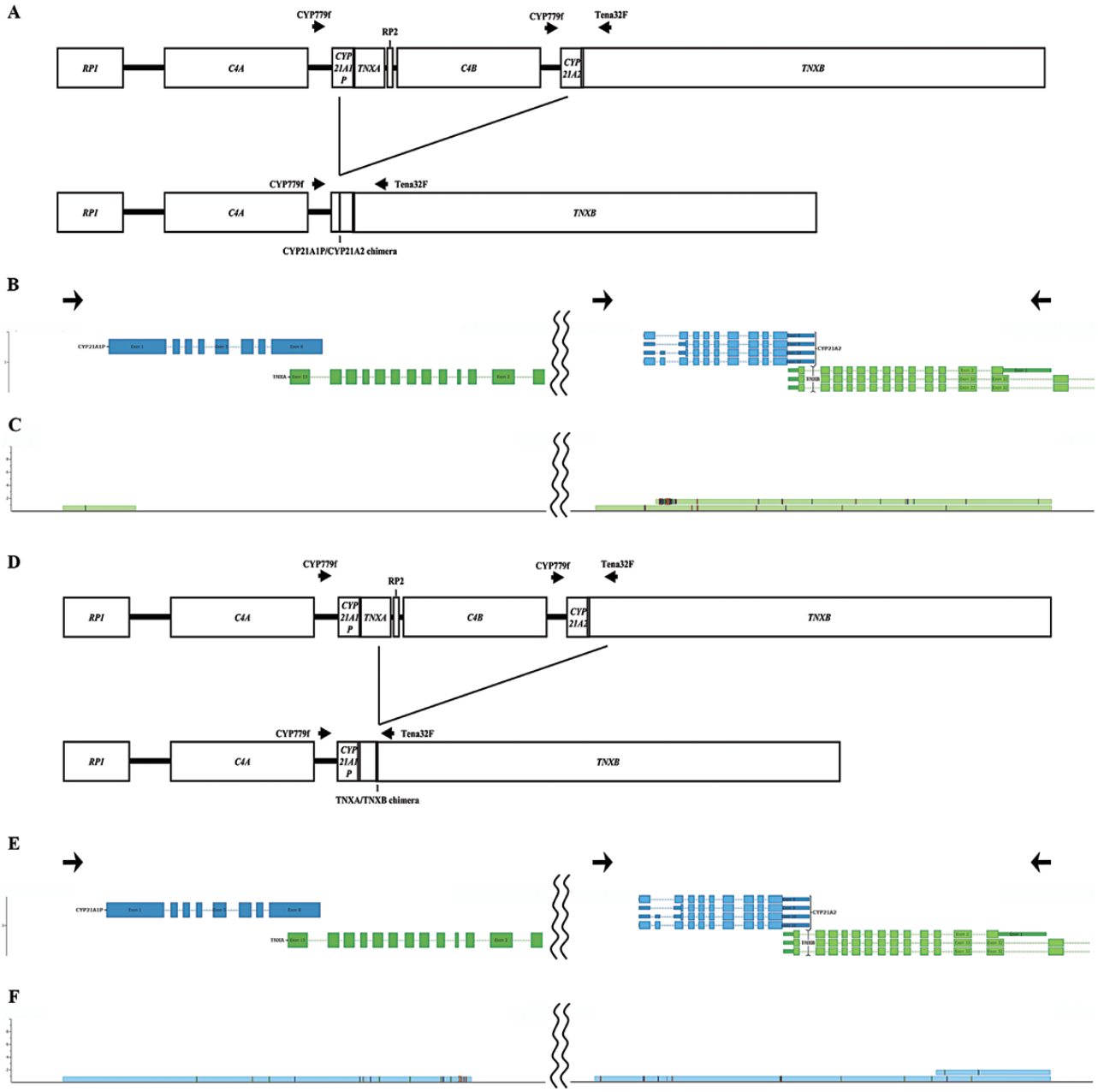

Figure 2. Findings of structural variants by long-read sequencing. (A, B) Schematic representation of *CYP21A2* chimera. (B: left panel) *CYP21A1P-TNXA* region and (B: right panel) *CYP21A2-TNXB* region. (C) Patient 32 possessed the *CYP21A1P/CYP21A2* chimeric gene type 1 (CYP-CH-1) demonstrated by the read in the left panel and the upper read of the right panel, as well as the p.I173N variant in the lower read of the right panel. (D, E) Schematic representation of *TNX* chimera. (E: left panel) *CYP21A1P-TNXA* region and (E: right panel) *CYP21A2-TNXB* region. (F) Patient 21 was compound heterozygous for a *TNX* chimera (the read in the left panel and the upper read of the right panel) and the p.R484Pfs variant (the lower read of the right panel).

chimera, respectively (Table 1). Since the parental DNA was not available, we could not demonstrate the exact mechanism. A possibility can be 1 parent carries a double deletion genotype (*CYP/TNX* chimeras), which was then passed each to the siblings.

The most common variant in our studies is the splice site IVS2-13A/C > G, which derived from the *CYP21A1P* pseudogene (Table 2). In normal individuals, the nucleotide position 13 before the end of intron 2 is A or C. Substitution to G creates an additional splice acceptor site, causing aberrant splicing of the intron 2 with retention of 19 intronic nucleotides, leading to shifts of the translational reading frame (37). In vitro expression analysis demonstrated that the enzyme activity was 1% to 5% of normal enzyme (37). The allele frequency of the mutation in each population was different ranging from 0.02 to 1 (38). In our cohort, the allele frequency was 0.38 (Table 2), which is similar to that of 0.33 of the previous study in a Thai population (38).

We have identified a novel G to T transversion at the conserved GT splice donor site of intron 7 (IVS7 + 1G > T) in a pair of siblings with the SW form of 21-OHD (http://www.hgmd.cf.ac.uk/). A variant at the same nucleotide,

Table 2. Frequencies of the pathogenic variants of 48 Thai patients with 21-OHD

Pathogenic variants	Positions	Alleles, n	Frequencies, %
IVS2-13A/C > G	Intron 2	36	38
p.I173N	Exon 4	12	13
CYP-CH-1	—	12	13
p.Q319X	Exon 8	4	4
p.R357W	Exon 8	4	4
TNX-CH-1	—	4	4
CYP-CH-8	—	3	3
IVS7 + 1G > T	Intron 7	2	2
CYP-CH-3	—	2	2
Entire gene deletion	—	2	2
p.I173N; p.I237N; p.V238E; p.M240K; p.V282L; p.L308Ffs; p.R357W	Cluster of exon 4, 6, 7, and 8	2	2
p.I173N; p.I237N; p.V238E; p.M240K; p.V282L; p.L308Ffs; p.Q319X; p.R357W	Cluster of exon 4, 6, 7, and 8	1	1
p.L167P	Exon 4	1	1
p.G293D	Exon 7	1	1
p.R317X	Exon 8	1	1
p.Q319X; p.R357W	Cluster of exon 8	1	1
p.P484Pfs*40	Exon10	1	1
p.P484Pfs*58	Exon 10	1	1
p.R484P	Exon 10	1	1
CYP-CH-2	—	1	1
CYP-CH-4	—	1	1
CYP-CH-5	—	1	1
CYP-CH-6	—	1	1
CYP-CH-7	—	1	1
Total		96	100

IVS7 + 1G > C, was previously reported in severely affected 21-OHD patients (39). The most common consequence of a splice site mutation is the exon skipping or the inclusion of the intron leading to an altered protein-coding sequence (40).

The known microconversion derived from pseudogenes p.I173N, p.Q319X, and p.R357W were reported. The p.I173N leads to the loss of the hydrophobic pocket, which reduced enzyme activity to 1% to 2% (33). The p.Q319X disrupts H-bonding and is predicted to result in a completely nonfunctional enzyme (41). The p.R357W is located in a region of the gene encoding the K helix of the enzyme, which suggests that the PV affects interactions with the cytochrome P450 reductase, and enzymatic activity is abolished (42). The previously reported missense variants, p.G293D and p.R484P, were identified. The p.G293D was identified in 1 allele of patient 27 with SW phenotype as previously reported, and in vitro study showed an activity < 1% (43). p.R484P is located within a stretch of amino acids, which is highly conserved domain. It was reported in 2 siblings with a moderate enzyme deficiency (39). The previously reported nonsense variant p.R317X in exon 8 is predicted to result in a truncated protein and the loss of the heme-binding domain, which was found in 1 allele of patient 28 with SW and was reported in 1 female who inherited the variant from her father, resulting in SW after birth (44). About 20% of PVs involve chimeric genes, which was similar to our findings (27%) (33).

Using LS-PCR Method 1 followed by LRS would be able to detect PVs in 48 out of 48 patients (100%) but unable to detect rearrangement alleles. On the other hand, using LS-PCR Method 2 followed by LRS would not be able to identify types of *CYP21A1P/CYP21A2* and *TNXA/TNXB* chimeras [in patients 9, 10, 20, 21, 28, 30-40, and 43-47) [compare Table 1 and Supplementary Table 3 (27)]. Moreover, LS-PCR Method 2 was unable to amplify some functional *CYP21A2* alleles with pathogenic single nucleotide variants, which is allele 2 (the pseudogene-derived mutant allele) of patients 41 and 42, and allele 1 (p.L167P) of patient 47, due to presence of single nucleotide polymorphisms (rs6457475 and rs6457476) in the reverse primer ME0066 region [Table 3; also see Supplementary Figure 9 (27)].

In summary, we developed a variant detection technique exploiting long-read amplicon sequencing that successfully identified 96 PVs in the *CYP21A2* in 48 Thai patients with 21-OHD.

Acknowledgments

We would like to thank the patients and their family for participating in this study.

Financial Support

This work was supported by Ratchadapiseksompotch Fund, Chulalongkorn University (764002-HE01), TSRI Fund

(CU_FRB640001_01_30_10), Thailand Research Fund (DPG6180001), and Health Systems Research Institute (64-132).

Disclosures

The authors have no conflicts of interest to declare.

Data Availability

All data sets generated during the current study are not publicly available but are available from the corresponding author on reasonable request.

References

1. Merke DP, Auchus RJ. Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *N Engl J Med*. 2020;383(13):1248-1261.
2. Claahsen-van der Grinten HL, Speiser PW, Ahmed SF, *et al*. Congenital adrenal hyperplasia-current insights in pathophysiology, diagnostics, and management. *Endocr Rev*. 2022;43(1):91-159.
3. Carrozza C, Foca L, De Paolis E, Concolino P. Genes and pseudogenes: complexity of the RCCX locus and disease. *Front Endocrinol (Lausanne)*. 2021;12:709758.
4. Koppens PF, Hoogenboezem T, Degenhart HJ. CYP21 and CYP21P variability in steroid 21-hydroxylase deficiency patients and in the general population in the Netherlands. *Eur J Hum Genet*. 2000;8(11):827-836.
5. Ezquieta B, Oliver A, Gracia R, Gancedo PG. Analysis of steroid 21-hydroxylase gene mutations in the Spanish population. *Hum Genet*. 1995;96(2):198-204.
6. Higashi Y, Tanae A, Inoue H, Fujii-Kuriyama Y. Evidence for frequent gene conversion in the steroid 21-hydroxylase P-450(C21) gene: implications for steroid 21-hydroxylase deficiency. *Am J Hum Genet*. 1988;42(1):17-25.
7. Owerbach D, Sherman L, Ballard AL, Azziz R. Pro-453 to Ser mutation in CYP21 is associated with nonclassic steroid 21-hydroxylase deficiency. *Mol Endocrinol*. 1992;6(8):1211-1215.
8. Tajima T, Fujieda K, Nakayama K, Fujii-Kuriyama Y. Molecular analysis of patient and carrier genes with congenital steroid 21-hydroxylase deficiency by using polymerase chain reaction and single strand conformation polymorphism. *J Clin Invest*. 1993;92(5):2182-2190.
9. Wilson RC, Wei JQ, Cheng KC, Mercado AB, New MI. Rapid deoxyribonucleic acid analysis by allele-specific polymerase chain reaction for detection of mutations in the steroid 21-hydroxylase gene. *J Clin Endocrinol Metab*. 1995;80(5):1635-1640.
10. Day DJ, Speiser PW, White PC, Barany F. Detection of steroid 21-hydroxylase alleles using gene-specific PCR and a multiplexed ligation detection reaction. *Genomics*. 1995;29(1):152-162.
11. O'Donovan MC, Oefner PJ, Roberts SC, *et al*. Blind analysis of denaturing high-performance liquid chromatography as a tool for mutation detection. *Genomics*. 1998;52(1):44-49.
12. Krone N, Braun A, Weinert S, *et al*. Multiplex minisequencing of the 21-hydroxylase gene as a rapid strategy to confirm congenital adrenal hyperplasia. *Clin Chem*. 2002;48(6 Pt 1):818-825.
13. Morel Y, Andre J, Uring-Lambert B, *et al*. Rearrangements and point mutations of P450c21 genes are distinguished by five restriction endonuclease haplotypes identified by a new probing strategy in 57 families with congenital adrenal hyperplasia. *J Clin Invest*. 1989;83(2):527-536.
14. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res*. 2002;30(12):e57.
15. Parajes S, Quinterio C, Dominguez F, Loidi L. A simple and robust quantitative PCR assay to determine CYP21A2 gene dose in the diagnosis of 21-hydroxylase deficiency. *Clin Chem*. 2007;53(9):1577-1584.
16. Loidi L, Quinteiro C, Parajes S, *et al*. High variability in CYP21A2 mutated alleles in Spanish 21-hydroxylase deficiency patients, six novel mutations and a founder effect. *Clin Endocrinol (Oxf)*. 2006;64(3):330-336.
17. Ma D, Chen Y, Sun Y, *et al*. Molecular analysis of the CYP21A2 gene in Chinese patients with steroid 21-hydroxylase deficiency. *Clin Biochem*. 2014;47(6):455-463.
18. Wang W, Han R, Yang Z, *et al*. Targeted gene panel sequencing for molecular diagnosis of congenital adrenal hyperplasia. *J Steroid Biochem Mol Biol*. 2021;211:105899.
19. Gangodkar P, Khadilkar V, Raghupathy P, *et al*. Clinical application of a novel next generation sequencing assay for CYP21A2 gene in 310 cases of 21- hydroxylase congenital adrenal hyperplasia from India. *Endocrine*. 2021;71(1):189-198.
20. Turan I, Tastan M, Boga DD, *et al*. 21-Hydroxylase deficiency: mutational spectrum and genotype-phenotype relations analyses by next-generation sequencing and multiplex ligation-dependent probe amplification. *Eur J Med Genet*. 2020;63(4):103782.
21. Hogan GJ, Vysotskaia VS, Beauchamp KA, *et al*. Validation of an expanded carrier screen that optimizes sensitivity via full-exon sequencing and panel-wide copy number variant identification. *Clin Chem*. 2018;64(7):1063-1073.
22. Mantere T, Kersten S, Hoischen A. Long-read sequencing emerging in medical genetics. *Front Genet*. 2019;10:426.
23. Stephens Z, Milosevic D, Kipp B, Grebe S, Iyer RK, Kocher JA. PB-Motif-a method for identifying gene/pseudogene rearrangements with long reads: an application to CYP21A2 genotyping. *Front Genet*. 2021;12:716586.
24. Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nat Rev Genet*. 2020;21(10):597-614.
25. Speiser PW, Arlt W, Auchus RJ, *et al*. Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2018;103(11):4043-4088.
26. Lee HH, Chang SF, Lee YJ, *et al*. Deletion of the C4-CYP21 repeat module leading to the formation of a chimeric CYP21P/CYP21 gene in a 9.3-kb fragment as a cause of steroid 21-hydroxylase deficiency. *Clin Chem*. 2003;49(2):319-322.
27. Tantirukdham N, Sahakitrungruang T, Chaisiwamongkol R, *et al*. Data from: supplemental materials for: long-read amplicon sequencing of the CYP21A2 in 48 Thai patients with steroid 21-hydroxylase deficiency. Deposited March 21, 2022. *Figshare*. https://doi.org/10.6084/m9.figshare.17873465
28. Keen-Kim D, Redman JB, Alanes RU, *et al*. Validation and clinical application of a locus-specific polymerase chain reaction- and minisequencing-based assay for congenital adrenal hyperplasia (21-hydroxylase deficiency). *J Mol Diagn*. 2005;7(2):236-246.
29. Fukasawa Y, Ermini L, Wang H, Carty K, Cheung MS. LongQC: a quality control tool for third generation sequencing long read data. *G3 (Bethesda)*. 2020;10(4):1193-1196.
30. Sedlazeck FJ, Rescheneder P, Smolka M, *et al*. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods*. 2018;15(6):461-468.
31. Claahsen-van der Grinten HL, Speiser PW, Ahmed SF, *et al*. Congenital adrenal hyperplasia—current insights in pathophysiology, diagnostics and management. *Endocr Rev*. 2022;**43**(1):91-159.
32. Gidlof S, Falhammar H, Thilen A, *et al*. One hundred years of congenital adrenal hyperplasia in Sweden: a retrospective, population-based cohort study. *Lancet Diabetes Endocrinol*. 2013;1(1):35-42.
33. Pignatelli D, Carvalho BL, Palmeiro A, Barros A, Guerreiro SG, Macut D. The complexities in genotyping of congenital adrenal hyperplasia: 21-hydroxylase deficiency. *Front Endocrinol (Lausanne)*. 2019;10:432.

34. Gao Y, Lu L, Yu B, *et al*. The prevalence of the chimeric TNXA/TNXB gene and clinical symptoms of Ehlers-Danlos syndrome with 21-hydroxylase deficiency. *J Clin Endocrinol Metab*. 2020;105(7):dgaa199.
35. Lao Q, Brookner B, Merke DP. High-throughput screening for CYP21A1P-TNXA/TNXB chimeric genes responsible for Ehlers-Danlos syndrome in patients with congenital adrenal hyperplasia. *J Mol Diagn*. 2019;21(5):924-931.
36. Marino R, Perez Garrido N, Ramirez P, Belgorosky A. Response to letter to the editor from Lao and Merke: "Ehlers-Danlos syndrome: molecular and clinical characterization of TNXA/TNXB chimeras in congenital adrenal hyperplasia." *J Clin Endocrinol Metab*. 2021;106(7):e2837-e2838.
37. Higashi Y, Hiromasa T, Tanae A, *et al*. Effects of individual mutations in the P-450(C21) pseudogene on the P-450(C21) activity and their distribution in the patient genomes of congenital steroid 21-hydroxylase deficiency. *J Biochem*. 1991;109(4):638-644.
38. Wilson RC, Nimkarn S, Dumic M, *et al*. Ethnic-specific distribution of mutations in 716 patients with congenital adrenal hyperplasia owing to 21-hydroxylase deficiency. *Mol Genet Metab*. 2007;90(4):414-421.
39. Wedell A, Luthman H. Steroid 21-hydroxylase deficiency: two additional mutations in salt-wasting disease and rapid screening of disease-causing mutations. *Hum Mol Genet*. 1993;2(5):499-504.
40. Anna A, Monika G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet*. 2018;59(3):253-268.
41. Globerman H, Amor M, Parker KL, New MI, White PC. Nonsense mutation causing steroid 21-hydroxylase deficiency. *J Clin Invest*. 1988;82(1):139-144.
42. Lajic S, Levo A, Nikoshkov A, Lundberg Y, Partanen J, Wedell A. A cluster of missense mutations at Arg356 of human steroid 21-hydroxylase may impair redox partner interaction. *Hum Genet*. 1997;99(6):704-709.
43. Tardy V, Menassa R, Sulmont V, *et al*. Phenotype-genotype correlations of 13 rare CYP21A2 mutations detected in 46 patients affected with 21-hydroxylase deficiency and in one carrier. *J Clin Endocrinol Metab*. 2010;95(3):1288-1300.
44. Lee HH, Chao HT, Lee YJ, *et al*. Identification of four novel mutations in the CYP21 gene in congenital adrenal hyperplasia in the Chinese. *Hum Genet*. 1998;103(3):304-310.