Analytical Biochemistry xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

Analysis of the *CYP21A1P* pseudogene: Indication of mutational diversity and *CYP21A2*-like and duplicated *CYP21A2* genes

Li-Ping Tsai [a,b], Ching-Feng Cheng [c,d], Shu-Hua Chuang [a], Hsien-Hsiung Lee [a,e,*]

[a] *Department of Pediatrics, Buddhist Tzu Chi General Hospital, Taipei Branch, Sindian, Taipei County 231, Taiwan*
[b] *School of Medicine, Tzu Chi University, Hualien 970, Taiwan*
[c] *Department of Pediatrics, Tzu Chi University, Hualien 970, Taiwan*
[d] *Department of Medical Research, Tzu Chi General Hospital, Hualien 970, Taiwan*
[e] *School of Chinese Medicine, College of Chinese Medicine, China Medical University, Taichung 404, Taiwan*

ARTICLE INFO

Article history:
Received 16 November 2010
Received in revised form 3 February 2011
Accepted 4 February 2011
Available online xxxx

Keywords:
CYP21A1P
XA
CYP21A2-like gene
Duplicated *CYP21A2* genes
PCR-based amplification
Mutational detection

ABSTRACT

The *CYP21A1P* gene downstream of the *XA* gene, carrying 15 deteriorated mutations, is a nonfunctional pseudogene that shares 98% nucleotide sequence homology with *CYP21A2* located on chromosome 6p21.3. However, these mutations in the *CYP21A1P* gene are not totally involved in each individual. From our analysis of 100 healthy ethnic Chinese (i.e., Taiwanese) (n = 200 chromosomes) using the polymerase chain reaction (PCR) products combined with an amplification-created restriction site (ACRS) method and DNA sequencing, we found that approximately 10% of *CYP21A1P* alleles (n = 195 chromosomes) presented the *CYP21A2* sequence; frequencies of P30, V281, Q318, and R356 in that locus were approximately 24%, 21%, 11%, and 34%, respectively, and approximately 90% of the *CYP21A1P* alleles had 15 mutated loci. In addition, approximately 2.5% (n = 5 chromosomes) showed four haplotypes of the 3.7-kb *Taq*I-produced fragment of the *CYP21A2*-like gene and one duplicated *CYP21A2* gene. We conclude that the pseudogene of the *CYP21A1P* mutation presents diverse variants. Moreover, the existence of the *CYP21A2*-like gene is more abundant than that of the duplicated *CYP21A2* gene downstream of the *XA* gene and could not be distinguished from the *CYP21A2–TNXB* gene; thus, it may be misdiagnosed by previously established methods for congenital adrenal hyperplasia caused by a 21-hydroxylase deficiency.

Congenital adrenal hyperplasia (CAH)[1] is an inherited disorder resulting mainly from a defect in the steroid 21-hydroxylase (*CYP21A2*) gene that causes approximately 90–95% of all CAH cases. It is one of the most common inborn errors of metabolism in humans. The wide range of CAH phenotypes is associated with multiple mutations known to affect 21-hydroxylase enzymatic activity. CAH is classified into three forms: classical salt wasting, classical simple virilizing, and nonclassical [1,2]. The incidence of the disease caused by these two classical forms is reported to be 1:10,000–1:18,000, depending on race [3,4]. The nonclassical form is milder and commonly occurs at a rate of 1:1700 in the general population [3,5].

The gene coding for P450c21 is designated *CYP21A2.* A duplicate copy designated *CYP21A1P* exists and shares 98% nucleotide sequence homology with *CYP21A2* in exons and 96% in noncoding sequences [6,7]. These two genes are located within the HLA class III human histocompatibility complex locus on chromosome 6p21.3, adjacent to and alternating with the *C4A* and *C4B* genes, which encode the fourth components of the serum complement [8]. The specific constitution of the *CYP21A2* locus is probably caused by an ancestral duplication of 30 kb encompassing the *CYP21A2* and *C4* genes. In this region, *C4* (*C4A* and *C4B*) [9], *CYP21A2* (and *CYP21A1P*), *RP* (*RP1* and *RP2*) [9], and tenascin (*XA* and *TNXB*) [10,11] comprise the most frequent bimodular RCCX (or the C4-CYP21 repeat) of the *RP1–C4–CYP21A1P–XA–RP2–C4–CYP21A2–TNXB* gene sequence in 69% of alleles in Caucasians [12]. The sequence length of the gene cluster is approximately 120 kb [13]. However, *CYP21A1P* downstream of *XA* is a pseudogene and is inactive due to the existence of 15 deteriorated mutations that include four nucleotides of the promoter region at nt −103 (G), −110 (C), −113 (A), and −126 (T) as well as a 30L, I2 splice (nt 655 or IVS2-12A/C > G), nt 707–714del, 172N, cluster E6, 281L, F306ΛL307insT, 318X, and 356W. In addition, there are two different gene sizes of 3.7 kb representing *CYP21A2* and 3.2 kb representing *CYP21A1P*, as determined by *Taq*I endonuclease digestion from the genome [8].

The RCCX module has three possible forms: monomodular, bimodular, and trimodular. Chromosomes with four RCCX modules

* Corresponding author at: School of Chinese Medicine, College of Chinese Medicine, China Medical University, Taichung 404, Taiwan. Fax: +886 3 9389073.
E-mail addresses: hhlee@mail.cmu.edu.tw, leehsienhsiung@gmail.com (H.-H. Lee).

[1] *Abbreviations used:* CAH, congenital adrenal hyperplasia; PCR, polymerase chain reaction; ACRS, amplification-created restriction site; dNTP, deoxynucleoside triphosphate; mRNA, messenger RNA.

0003-2697/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2011.02.016

are very rare [14]. In the monomodule, the *CYP21A1P* gene does not exist in approximately 17% of Caucasians, as reported by Blanchong and coworkers [14]. However, haplotypes of the RCCX module with more than one *CYP21A2* gene in the trimodular form were observed [14–19]. Obviously, the gene located downstream of the *XA* gene can possibly include the *CYP21A2* gene as well as the *CYP21A1P* gene.

The current study was undertaken to use specifically paired primers to amplify a PCR product containing *XA* and the adjacent upstream *CYP21A1P* gene to identify 15 mutational loci commonly appearing in *CYP21A1P* and to examine whether all of these loci occurred in the *CYP21A1P* gene in our healthy individuals. In addition, we determined whether the duplicated *CYP21A2* gene exists downstream of *XA* and established a protocol to distinguish the *CYP21A2* gene existing in the *XA* and *XB* genes, which may produce a misdiagnosis of CAH caused by a 21-hydroxylase deficiency.

Materials and methods

Subjects

For a carrier analysis of CAH caused by the 21-hydroxylase deficiency in our population, 1000 blood samples, used mainly for biochemical testing at the time of a health examination, were obtained from Veterans General Hospital – Taipei in 2000 [20]. Among these samples, we used 100 samples to analyze the *CYP21A1P* gene. The institute review board of Veterans General Hospital – Taipei approved the protocol, and the study followed its guidelines strictly.

Amplification and restriction endonuclease analysis of the primary 6.1-kb PCR product containing the CYP21A1P and XA genes

All of the *CYP21A1P* gene, except the monomodule in the RCCX module, is located downstream of the *XA* gene (Fig. 1A). To amplify *CYP21A1P*, a 6.1-kb polymerase chain reaction (PCR) product amplified with the specifically paired primers, CYP-779f/XA-36F (Table 1), was derived from 100 healthy individuals. The 6.1-kb PCR product contained the entire *CYP21A1P* and partial *XA* genes (Fig. 1A). Primer CYP-779f (5′-aggtgggctgttttcctttca-3′, nt −779 to −759) [6] is located at the 5′ end of the *CYP21A1P* and *CYP21A2* genes, and XA-36F (5′-ggacccagaaactccaggtgg-3′, nt 4252–4272, GenBank accession No. S38953) is located in the *XA* gene of exon 36/intron 36 (Fig. 1A). The following PCR conditions were used: initial denaturation at 94 °C for 4 min, followed by 12 cycles at 94 °C for 30 s, 62 °C for 40 s, and 68 °C for 5 min, and another 16 cycles at 94 °C for 30 s and 68 °C for 5 min. The 6.1-kb PCR product with 10 U in a 10-μl volume of *Taq*I cleavage enzymes was incubated at 65 °C for 2 h. The completely digested PCR products were analyzed by electrophoresis on a 1.0% agarose gel.

ACRS method for identifying 10 mutational loci and DNA sequencing of five loci of the CYP21A1P gene using the secondary PCR product

Identification of *CYP21A1P* gene mutations was followed by amplification-created restriction site (ACRS) primer detection (see Supplemental Table 1 in supplementary material) [21] using the 6.1-kb PCR product as a template. ACRS primer pairs and the restriction enzyme reaction were used to detect mutations (Table 2), including 30L (A), IVS2-12G (or I2 splice) (B), 172N (C), 236N (D), 281L (G), 318X (J1), and 356W (J2) (Fig. 2A), as described previously [21]. To detect the IVS2-12G (I2 splice) in combination with nt 707–714del, the 3B/C4 paired primers were used as described previously [22]. To detect the promoter region (P in Table 2) of nt −209 (C), −198 (T), −188Λ−189 (T), −126 (T), −113 (A), −110 (C), −103 (G), and −4 (T) [6], 291-bp PCR fragments derived from the paired primers, CYP-270f/C120 (Table 1), were generated. To detect the F306ΛL307inseT mutation (H in Table 2), 519-bp PCR fragments derived by the paired primers, Ex6/In7R (Table 1), were generated. These two secondary PCR products were subjected to DNA sequencing. The 519-bp PCR fragment contains the *CYP21A1P* sequence from nt 1306 to 1825, which also includes five SNP site detections of nt 1375 (C), 1420 (G), 1421 (T), 1586 (G), and 1789 (C) (Table 2) [6]. The sequence and location of the sequencing primers are listed in Table 1. The reaction mixture of the secondary PCR amplification contained 0.5 μl of the primary PCR product, 5 μl of 10× PCR buffer (100 mmol/L Tris–HCl [pH 8.8], 500 mmol/L KCl, 25 mmol/L $MgCl_2$, and 0.1% Tween 20), 0.08 U of Thermoprime Plus DNA polymerase (Advanced Biotechnologies, UK), 7.5 pmol of each primer, and 80 μmol/L of each deoxynucleoside triphosphate (dNTP) in a final volume of 50 μl. The PCR consisted of 37 cycles at 94 °C for 30 s, 58 °C for 40 s, and 72 °C for 50 s.

DNA sequencing

The secondary PCR products (Table 1) for DNA sequencing were used in a Taq Dye Deoxy Terminator Cycle Sequencing Kit (Applied Biosystems, USA).

Results

Use of the 6.1-kb PCR products for CYP21A1P amplification and TaqI restriction endonuclease analysis of 100 normal individuals

A 6.1-kb primary PCR product (Fig. 1B, lane 1) (data from only one person) was generated by the paired primers, CYP779f/XA-36F, from 100 healthy individuals. From the *Taq*I cleavage analysis of the 6.1-kb PCR product, two *Taq*I sites (TCGA) in this region (nt −209 and 2995) (Fig. 1A) [6] produced three fragments of 3207, 2315, and 591 bp (the last one of which ran out of the gel) (Fig. 1B, lane 2) (data from only one sample) that appeared in 96 healthy individuals (192 chromosomes). However, three healthy individuals (samples 350, 984, and 988) produced fragments of 3738, 3207, 2315, 591, and 60 bp (the last two of which ran out of the gel) (data from only sample 984) (Fig. 1B, lane 3), and one healthy individual (sample 646) produced 3738, 2315, and 60 bp (the last one of which ran out of the gel) (Fig. 1B, lane 4). This indicated that the *CYP21A1P* gene downstream of the *XA* gene in most healthy individuals (n = 195 chromosomes) presented a putative size of the 3.2-kb *Taq*I-produced fragments (Fig. 1B, lane 2) that showed a 97.5% frequency. Otherwise, one allele of the *CYP21A2* gene of samples 350, 984, and 988 had the 3.7-kb *Taq*I-produced fragment, and the other allele had 3.2 kb of the *CYP21A1P* gene. Sample 646 had the 3.7-kb *Taq*I-produced fragment of the *CYP21A2* gene in two alleles. This implies that 2.5% (five chromosomes) of healthy individuals had *CYP21A2* genes (Table 3).

ACRS analysis and DNA sequencing of the secondary PCR product for 15 loci detections in 96 healthy individuals carrying the 3.2-kb TaqI-produced fragments

To detect the 10 loci of A, B, C, D (cluster E6), G, J1, and J2 (Fig. 2A) from 96 healthy individuals who showed the 3.2-kb *Taq*I-produced fragments, the 6.1-kb primary PCR product (Fig. 1B, lane 1) generated by the paired primers, CYP-779f/XA-36F, as the template was subjected to secondary ACRS primer amplification (data from only one sample) using the paired primers, CIN/C2 (Table 2), and this generated a 195-bp fragment of locus (A) (Fig. 2B, lane 1); C3B/C4 generated a 124-bp fragment of locus (B) (Fig. 2B, lane 5); C5/C6 generated a 148-bp fragment of locus (C) (Fig. 2B, lane 10); C7D1/C8 generated a 140-bp frag-

Please cite this article in press as: L.-P. Tsai et al., Analysis of the *CYP21A1P* pseudogene: Indication of mutational diversity and *CYP21A2*-like and duplicated *CYP21A2* genesCYP21A1P pseudogene –>, Anal. Biochem. (2011), doi:10.1016/j.ab.2011.02.016

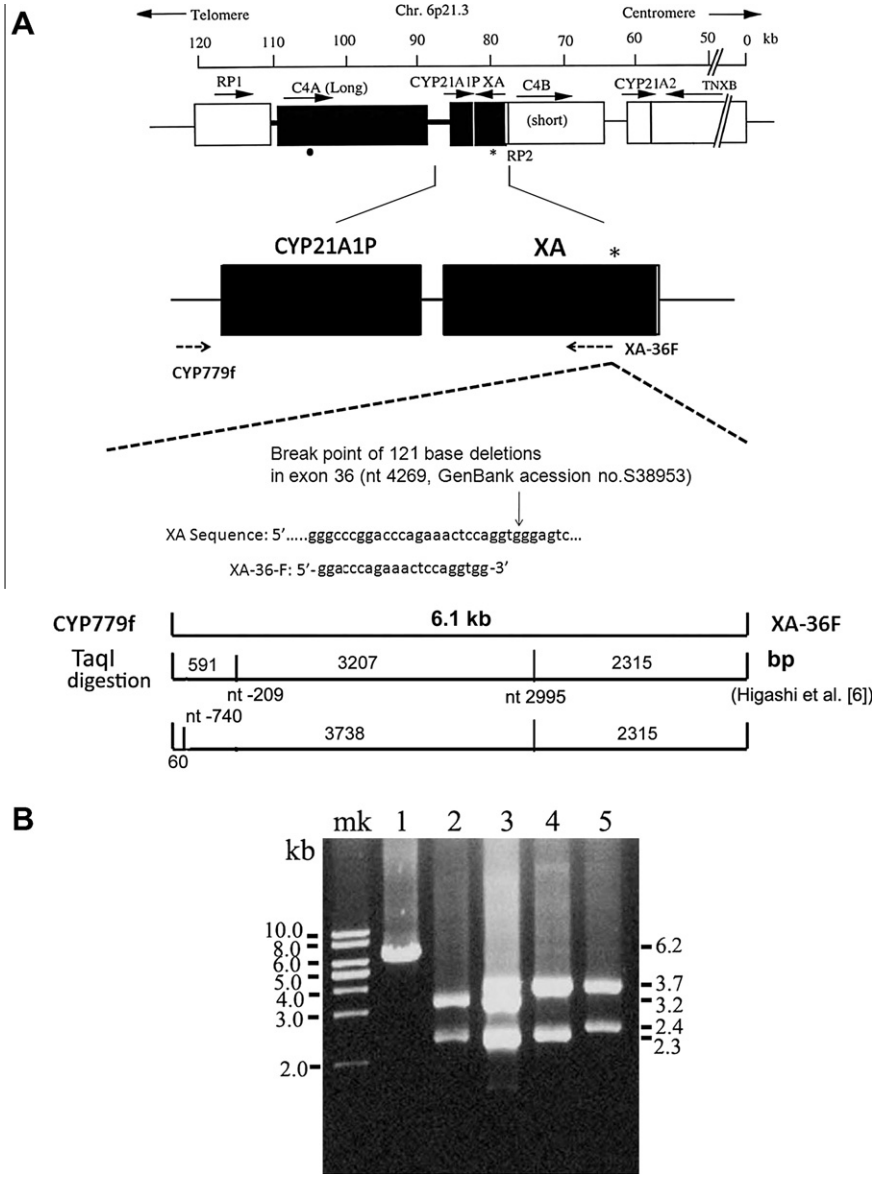

Fig.1. (A) Bimodular form (or the C4-CYP21 repeat module) (*RP1–C4A–CYP21A1P–XA–RP2–C4B–CYP21A2–TNXB*) of the RCCX region of chromosome 6p21.3 and the strategy for PCR amplification of a 6.1-kb fragment encompassing the *XA* gene to the *CYP21A1P* gene. The white box indicates the structures of the *RP1*, *CYP21A2*, *C4B*, and *TNXB* genes, whereas the black box represents the *C4A*, *CYP21A1P*, *XA*, and *RP2* genes. Sizes of the genes from the ATG start codon to the TGA stop codon, including *RP1*, *C4A*, *CYP21A1P*, *XA*, *RP2*, *C4B*, *CYP21A2*, and *TNXB*, in the RCCX module of the figure are based on sequences of GenBank accession numbers AL049547 and AF019413. The presence of *C4A* (the long gene of 20.4 kb) or *C4B* (the short gene of 14.1 kb) depends on the presence of HERV-K (C4), the endogenous 6.7-kb retroviral sequence, in intron 9 indicated by a solid circle (●). Solid arrows indicate the orientation of transcription. A 121-bp deletion in exon 36 of the *XA* gene is marked with an asterisk (*). Top: Scale in kilobases (kb), with the *TNXB* gene starting at 0. Horizontal dashed arrows represent the direction and location of primers XA-36F and CYP779f. Bottom: Restriction digestion analysis of the *Taq*I restriction fragment on the 6.1-kb PCR product amplified with the paired primers, CYP779F/XA-36F. (B) *Taq*I cleavage analysis of the 6.1-kb PCR product amplified from *XA* to the 5′ end of the *CYP21A1P* gene. The analysis was carried out on a 1.0% agarose gel. Lane 1: 6.1-kb PCR product amplified with the CYP779/XA-36F primer pair from a wild-type individual; lane 2: 6.1-kb PCR product digested with *Taq*I from one of 195 healthy individuals that produced three fragments of 3207, 2315, and 591 bp (the last one of which ran out of the gel); lane 3: one of three healthy individuals (sample 984) that produced five fragments of 3738, 3207, 2315, 591, and 60 bp (the last two of which ran out of the gel); lane 4: sample 646 that produced three fragments of 3740, 2315, and 60 bp (the last one of which ran out of the gel); lane 5: 6.2-kb fragment amplified with the paired primers, CYP779f/Tena36F2 [22], that produced three fragments of 3740, 2491, and 60 bp (the last one of which ran out of the gel); lane mk: 1-kb DNA ladder (domestic).

ment of locus (D) (Fig. 2B, lane 14); C9/C10-1 generated a 213-bp fragment of locus (G) (Fig. 2B, lane 18); and C11/C12 generated a 197-bp fragment of loci (J1) and (J2) (Fig. 2B, lane 22). In addition, for DNA sequencing, we also prepared (data not shown) paired primers of CYP-270f/C120 to generate a 291-bp fragment of the promoter region (P in Table 2) and EX6/In7R to generate a 519-bp fragment of locus (H) (Table 2). After appropriate endonuclease cleavages (Table 2) of the six different ACRS PCR products (Fig. 2B, lanes 1, 5, 10, 14, 18, and 22), only two alleles (data from only sample S1) of 24 healthy individuals showed the *CYP21A1P* sequence in 10 loci (Table 2) that presented a 164-bp fragment of *CYP21A1P* locus (A) (Fig. 2B, S1), a 93-bp fragment of *CYP21A1P* locus (B) (Fig. 2B, S1), a 118-bp fragment of *CYP21A1P* locus (C) (Fig. 2B, S1), a 140-bp fragment of *CYP21A1P* locus (D) (Fig. 2B, S1), a 213-bp fragment of *CYP21A1P* locus (G) (Fig. 2B, S1), a 197-bp fragment of *CYP21A1P* locus (J1) (Fig. 2B, S1), and a 167-bp fragment of *CY-*

Please cite this article in press as: L.-P. Tsai et al., Analysis of the *CYP21A1P* pseudogene: Indication of mutational diversity and *CYP21A2*-like and duplicated *CYP21A2* genesCYP21A1P pseudogene –>, Anal. Biochem. (2011), doi:10.1016/j.ab.2011.02.016

Table 1
Primers for amplification of the *CYP21A1P* gene and secondary PCR product for DNA sequencing.

Designation	Primer (5′ → 3′)	Location (nt)	Specificity
Primers for amplification of the 6.1-kb PCR product			
CYP-779f	CCAGAAAGCTGACTCTGGATG	−779 to −759[a]	CYP21A1P/ CYP21A2
XA-36F	GGACCCAGAAACTCCAGGTGG	4252–4272[b]	XA
Primers for amplification of the 291-bp PCR product for DNA sequencing			
CYP-270f	CCAGAAAGCTGACTCTGGATG	−271 to −251[a]	CYP21P/CYP21
C120	AGCAGGCCCAGGAGCAGCAT	20–1[a]	CYP21P/CYP21
Primers for amplification of the 519-bp PCR product for DNA sequencing			
EX6	TCATGCTTCCTGCCGCAGTTC	1306–1326[a]	CYP21P/CYP21
In7R	GCCAGGTTGCTGGGAAGGAGC	1825–1805[a]	CYP21P/CYP21
Primers for amplification of the 469-bp PCR product for DNA sequencing			
1A2	CTGCTGGCTGGCGCCCGCCT	31–50[a]	CYP21P/CYP21
In2	AGGTGGGAGGATCATTTGAG	500–481[a]	CYP21P/CYP21

[a] Based on Higashi and coworkers [6].
[b] Based on GenBank accession number S38953.

Table 2
Primer pairs for the ACRS analysis of 10 mutation loci and DNA sequencing of the promoter region and F306ΛL307insT locus in the *CYP21A1P* gene.

ACRS analysis

Designation	Mutation locus[a]	Paired primer	PCR (bp)	Restriction site		Fragment (bp)	
				Natural	Created	CYP21A2	CYP21A1P
A	30L	ClN/C2	195	–	*Pst*I	195	164 + 31
B	I2splice and 707–714del[b]	C3B/C4	124	–	*Sac*I	132	93 + 31
C	172N	C5/C6	148	–	*Mse*I	148	118 + 30
D	236N[c]	C7D1/C8	140	–	*Mbo*I	114 + 26	140
G	281L	C9/C10-1	213	*Apa*LI	–	116 + 101	213
J1	318X	C11/C12	197	*Pst*I	–	146 + 51	197
J2	356W	C11/C12	197	–	*Msc*I	197	167 + 30

DNA sequencing analysis

Designation	Locus (nt/aa)[a]	Paired primers	PCR (bp)	Sequence[a]	
				CYP21A2	CYP21A1P
H	F306ΛL307	EX6/In7R	519	–	T
	nt 1375			T	C
	1420			A	G
	1421			C	T
	1586			C	G
	1789			G	C
P	nt −209	CYP-270f/C120	291	T	C
	−198			C	T
	−188Λ189			–	T
	−126			C	T
	−113			G	A
	−110			T	C
	−103			A	G
	−4			C	T
H62		1A2/In2	469	CAC	CAC/CTC

[a] Based on Higashi and coworkers [6].
[b] I2 splice denotes IVS2-12A/C > G or nt 655.
[c] E6 cluster indicates I236, V237, and M239.

P21A1P locus (J2) (Fig. 2B, S1). In addition, two secondary PCR products of the 291- and 519-bp fragments for DNA sequencing (data not shown) also contained the *CYP21A1P* sequence in two alleles in the promoter region of four SNP sites of nt −126 (T), −113 (A), −110 (C), and −103 (G) (Table 2) and the F306ΛL307inseT locus (data not shown). This indicates that 24% (24/100) of *CYP21A1P* downstream of the *XA* gene of healthy individuals showed 15 mutated loci in two alleles. The allelic frequencies of healthy individuals (n = 195 chromosomes) with normal and mutated loci in the *CYP21A1P* gene are summarized in Table 4.

ACRS analysis and DNA sequencing of the secondary PCR product for 15 loci detections of five chromosomes carrying 3.7-kb TaqI-produced fragments

Three healthy individuals (samples 350, 984, and 988) had the 6.1-kb primary PCR product (Fig. 1B, lane 3) (data from only sample 984) that produced fragments of 3.7, 3.2, 2.3, 0.5, and 0.06 kb on the *Taq*I digestion analysis. With the ACRS analysis of the A, B, C, D (cluster E6), G, J1, and J2 loci (Fig. 2B), sample 984 (S2) presented fragments of the 195-bp *CYP21A2* and 164-bp *CYP21A1P* loci (A) (Fig. 2B, S2), the 132-bp *CYP21A2* and 93-bp *CYP21A1P* loci (B)

Please cite this article in press as: L.-P. Tsai et al., Analysis of the *CYP21A1P* pseudogene: Indication of mutational diversity and *CYP21A2*-like and duplicated *CYP21A2* genesCYP21A1P pseudogene –>, Anal. Biochem. (2011), doi:10.1016/j.ab.2011.02.016

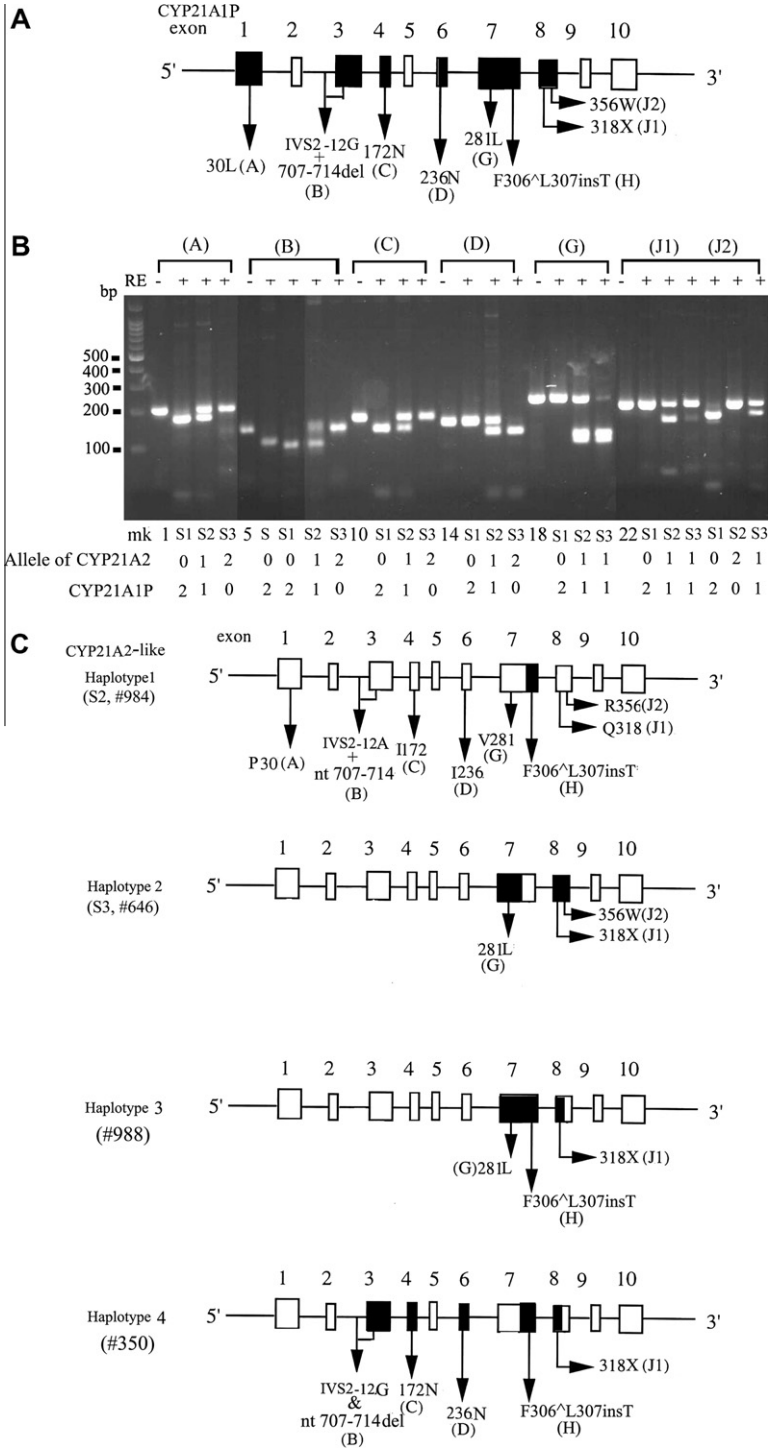

Fig.2. Restriction analysis of the amplification product of the ACRS method and map for detecting mutational loci of the *CYP21A1P* gene. (A) Detection of an ACRS amplification product of the *CYP21A1P* gene for 30L (A), for IVS2-12G (or the I2 splice) and 707–714del (B), for 172N (C), for 236N (D), for 281L (G), for 318X (J1), and for 356W (J2). The ACRS primers were described previously (Supplemental Table 1) [21,22]. Mutational loci of the *CYP21A1P* gene are indicated by a black box, and normal loci are indicated by a white box. (B) Restriction analysis of the ACRS amplification products of three healthy individuals on a 2.7% agarose gel. The loci are designated (A), (B), (C), (D), (G), (J1), and (J2) corresponding to map A. Lanes 1, 5, 10, 14, 18, and 22 are PCR products for (A), (B), (C), (D), (G), and (J1) and (J2), respectively. Lane S1: one healthy individual with the 3.2-kb *Taq*I fragment of the *CYP21A1P* gene in two alleles; lanes S2 and S3: samples 984 and 646, respectively; lane S: one CAH patient with the I2 splice mutation (without the 707–714del mutation) in two alleles that produced a 103-bp fragment by the ACRS analysis [22]. Each amplification product was either untreated ("−" lanes) or treated ("+" lanes) with an appropriate restriction enzyme (RE) (Table 2). Lane mk: 100-bp DNA ladder (domestic). (C) Four proposed haplotypes of the *CYP21A2*-like gene present in four healthy individuals with the 3.7-kb *Taq*I-produced fragment from the ACRS and DNA sequencing analyses.

(Fig. 2B, S2), the 148-bp *CYP21A2* and 118-bp *CYP21A1P* loci (C) (Fig. 2B, S2), the 114-bp *CYP21A2* and 140-bp *CYP21A1P* loci (D) (Fig. 2B, S2), the 213-bp *CYP21A1P* and 116- plus 101-bp *CYP21A2* loci (G) (Fig. 2B, S2), the 197-bp *CYP21A2* and 146-bp *CYP21A1P* loci (J1) (Fig. 2B, S2), and the 197-bp *CYP21A2* locus (J2) (Fig. 2B, S2). From DNA sequencing of the 291-bp fragment to identify the promoter region of four SNP sites, sample 984 (S2) presented the status of nt −126 (C/T), −113 (G/A), −110 (T/C), and −103 (A/G) in

Please cite this article in press as: L.-P. Tsai et al., Analysis of the *CYP21A1P* pseudogene: Indication of mutational diversity and *CYP21A2*-like and duplicated *CYP21A2* genesCYP21A1P pseudogene –>, Anal. Biochem. (2011), doi:10.1016/j.ab.2011.02.016

Table 3
Frequency of the 3.2- and 3.7-kb *Taq*I-produced fragments of the *CYP21A1P* gene downstream of the *XA* gene in healthy ethnic Chinese (i.e., Taiwanese).

Chromosomes		
	n	%
CYP21A1P with the 3.2-kb *Taq*I-produced fragment	195	97.5
CYP21A1P with the 3.7-kb *Taq*I-produced fragment	5[a]	2.5
Total	200	100

[a] Four haplotypes of the *CYP21A2*-like gene and one haplotype of the *CYP21A2* gene were included.

Table 4
Allelic frequencies of the *CYP21A1P* gene with the *CYP21A2* sequence in healthy ethnic Chinese (i.e., Taiwanese) (n = 195 chromosomes).

Locus[a]	P	A	B	C	D	G	H	J1	J2
Two alleles	0	7	0	0	0	7	0	2	14
One allele	0	33	0	0	0	27	0	17	38
Total individuals	0	40	0	0	0	34	0	19	52
Total chromosomes	0	47	0	0	0	41	0	21	66
Locus frequency (%)	0	24.1	0	0	0	21	0	10.8	33.8
Allelic frequency ~10% (175/9 × 195)									

Note: The population frequency of the normal locus in the *CYP21A1P* gene was approximately 10%. The population frequency of the mutated locus in the *CYP21A1P* gene was approximately 90%.

[a] P: nt −126 (C), −113 (G), −110 (T), and −103 (A); A: P30; B: I2 splice (IVS2-12A/T, nt 655); C: I172; D: E6 cluster (I236, V237, and M239); G: V281; H: F306/L307; J1: Q318; J2: R356 (based on Higashi et al. [6]).

two alleles (see Supplemental Table 2 in supplementary material) and 519-bp fragments showed the presence of F306ΛL307inseT in two alleles (see Supplemental Table 2). This indicated that sample 984 (S2) might carry one *CYP21A2*-like gene with a 3.7-kb *Taq*I-produced fragment (Fig. 1B, lane 3) that had the F306ΛL307inseT mutation (H) (Fig. 2C, haplotype 1), and the other allele with a 3.2-kb *Taq*I-produced fragment of the *CYP21A1P* gene (Fig. 1B, lane 3) had the 197-bp *CYP21A2* locus (J2) (Fig. 2B, S2). In addition, the ACRS analysis of sample 646 (S3) with two alleles presenting the 3.7-kb *Taq*I-produced fragment (Fig. 1B, lane 4) showed fragments of the 195-bp *CYP21A2* locus (A) (Fig. 2B, S3), the 132-bp *CYP21A2* locus (B) (Fig. 2B, S3), the 148-bp *CYP21A2* locus (C) (Fig. 2B, S3), the 114-bp *CYP21A2* locus (D) (Fig. 2B, S3), the 213-, 116-, and 101-bp *CYP21A1P* and *CYP21A2* loci (G) (Fig. 2B, S3), the 197-bp *CYP21A1P* and 146-bp *CYP21A2* loci (J1) (Fig. 2B, S3), and the 197-bp *CYP21A2* and 164-bp *CYP21A1P* of loci (J2) (Fig. 2B, S3). From DNA sequencing of the 291-bp fragment (see Supplemental Table 2) of the promoter region (P in Table 2), sample 646 (S3) presented the *CYP21A2* sequence of nt −126 (C), −113 (G), −110 (T), and −103 (A) in two alleles (see Supplemental Table 2), and the sequenced 519-bp fragments did not carry the F306ΛL307inseT mutation (see Supplemental Table 2) in two alleles. This indicated that sample 646 (S3) might carry one *CYP21A2*-like gene with 3.7-kb *Taq*I-produced fragments (Fig. 1B, lane 4) that had mutations of 281L, 318X, and 356R (Fig. 2C, haplotype 2) and another allele with a normal *CYP21A2* gene (see Supplemental Table 2). Two samples, 988 and 350, were also identified by ACRS and DNA sequencing, and the data are listed in Supplemental Table 2. Haplotypes of the *CYP21A2*-like genes with the 3.7-kb *Taq*I-produced fragment are shown in Fig. 2C.

Four healthy individuals showed normal allele in the CYP21A2 gene downstream of the TNXB gene

With an examination of four healthy individuals (samples 984, 646, 988, and 250) for a carrier analysis in 2000 [20] using differential PCR amplification with the paired primers, BF1/21BR [21], their haplotypes of *CYP21A2* showed two normal alleles. Primers BF1 and 21BR are specific for the 5′ and 3′ ends of the *CYP21A2* sequence, respectively. Therefore, this implies that the *CYP21A2*-like gene that existed downstream of the *XA* gene in these four healthy individuals cannot be generated by PCR amplification using the paired primers, BF1/21BR, as in a previously established method.

The presence of normal locus P30 (CCG) in the CYP21A1P gene is attributed to the production of L62 (CTC), and the L30 (CTG) mutation goes with H62 (CAC) in the gene sequence

Our analysis revealed a 24.1% frequency (Table 4) of the normal P30 (CCG) locus (A) (Fig. 3A) in the *CYP21A1P* gene with a 3.2-kb *Taq*I-produced fragment in 40 healthy individuals (Table 4) in whom locus H62 (CAC) was altered to L62 (CTC) (Fig. 3B) and mutated L30 (CTG) (Fig. 3D) retained locus H62 (CAC) in the gene sequence (Fig. 3E). However, a normal P30 was present in four haplotypes of the *CYP21A2*-like genes (samples 984, 646, 988, and 350; see Supplemental Table 2), and H62 was not changed to L62 (see Supplemental Table 2). Because L62 does not exist in the *CYP21A1P* or *CYP21A2* gene, normal P30 presenting in *CYP21A1P* with a 3.2-kb *Taq*I-produced fragment may play a role in creating the L62 sequence through a spontaneous recombination.

Discussion

Pseudogenes are divided into two types: processed and nonprocessed. Processed pseudogenes lack introns and are presumably generated from mature messenger RNA (mRNA) that carries out retrotransposition and is then integrated into the genome. The integration mechanism is by random chance, and the location is frequently not near the parental gene [23]. In this class, some pseudogenes were identified as having "live" transcriptional activity in humans [24–26]. Nonprocessed pseudogenes arise from duplication of the parental gene. The duplication region is close to the parental gene, within 500 kb (22.6% of frequencies) [23]. Nonprocessed pseudogenes contain intron sequences and genetic lesions such as in-frame termination codons produced by a single base change and base insertions and deletions that cause a change in the reading frame and, thus, premature termination. Therefore, the sequence of the functional gene being converted to that of the pseudogene may cause genetic mutations.

CAH is a term that describes several inheritable disturbances in steroid hormone metabolism. There are six enzymes, *CYP11A*, *CYP17* (17,20-lyase), *CYP21A2*, *CYP11B1*, *CYP11B2*, and 17β-hydroxylsteroid dehydrogenase, that are required for the synthesis of steroid hormones. Among them, only *CYP21A2* contains the *CYP21A1P* pseudogene; they are separated by 30 kb in chromosome 6p21.3 and show great similarity. Therefore, this seems to be the most likely reason for misalignment and gene conversions to occur at meiosis between these two genes. As a result, the most frequent steroid hormone abnormality occurs as a *CYP21A2* deficiency that causes approximately 90–95% of all CAH cases. Genetic defects of the *CYP21A2* gene in CAH may commonly lead to one of two categories of (i) small-scale conversions of the *CYP21A1P* sequence (commonly one of 15 mutations) and (ii) chimeras of the RCCX module, including the chimeric *CYP21A1P/CYP21A2* and *TNXA/TNXB* genes [27–29]. The *CYP21A1P* pseudogene belongs to the nonprocessed type [23], for which *CYP21A1P* transcription itself was not detected under any culture conditions or with various regulatory factors [30] solely because of two deteriorated mutations of the 707–714del and I2 splice. However, the promoter can drive transcription to an unrelated gene termed *YA* [31], although the function of *YA* is not yet clear.

Please cite this article in press as: L.-P. Tsai et al., Analysis of the *CYP21A1P* pseudogene: Indication of mutational diversity and *CYP21A2*-like and duplicated *CYP21A2* genesCYP21A1P pseudogene –>, Anal. Biochem. (2011), doi:10.1016/j.ab.2011.02.016

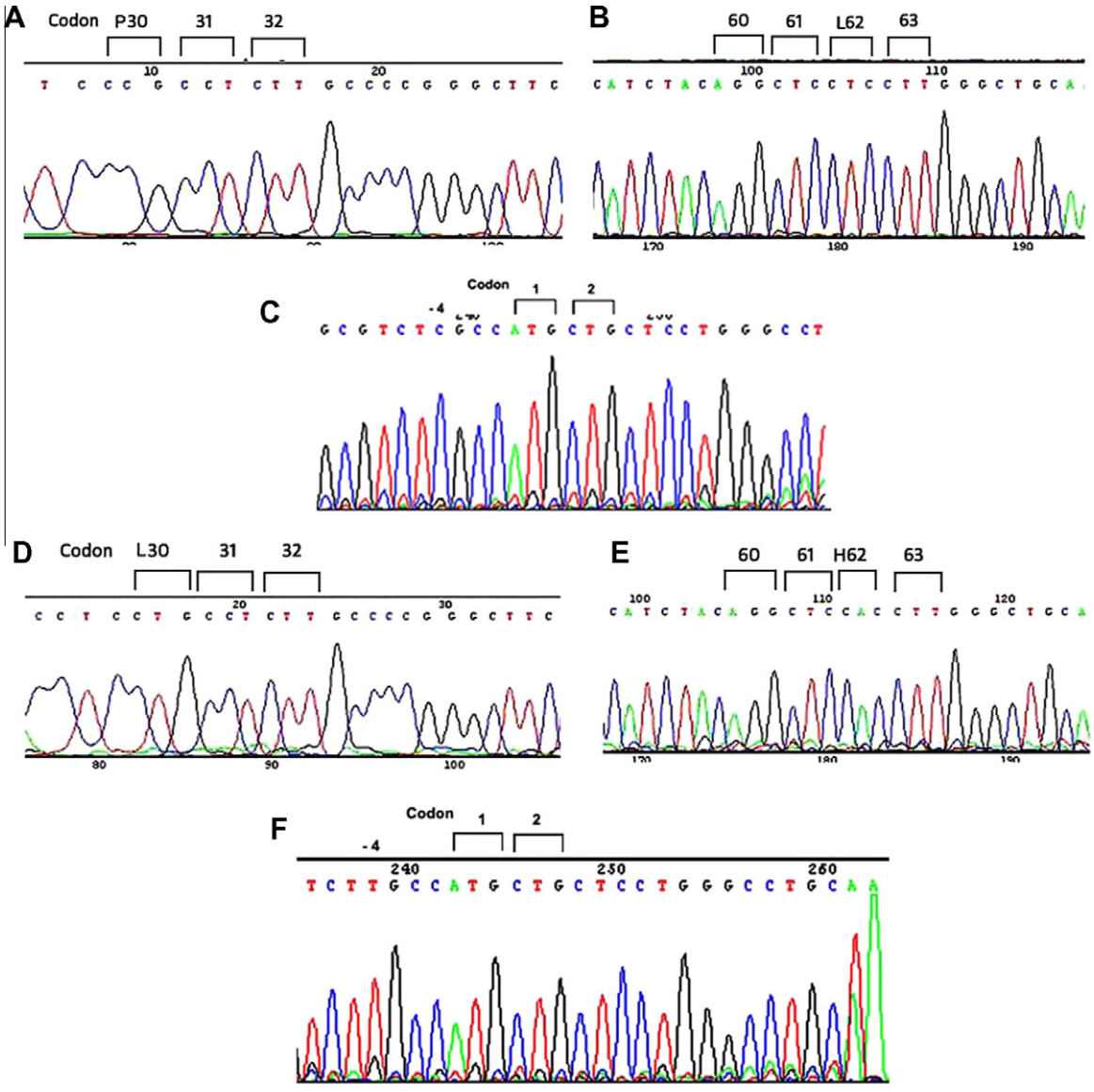

Fig.3. DNA sequencing analysis of codons 30 and 62, and the promoter region of the nt -4 sequence of the CYP21A1P gene with the 3.2-kb TaqI-produced fragment from the secondary PCR products described in the context. A, B, and C for individuals with locus of P30, L60, and C at nt -4. D, E, and F for individuals with locus of L30, H62, and T at nt -4.

In this study, we used the specific paired primers, CYP779/XA-36F, to amplify the *CYP21A1P* gene located downstream of the *XA* gene and examined the status of the gene size and mutational locus distributions. In an analysis of the alleles ($n = 47$ chromosomes) (Table 4) of the 3.2-kb *Taq*I-produced fragment of the *CYP21A1P* gene with the P30 (CCG) locus (A) of the *CYP21A2* sequence, the promoter sequence of nt −4 contained C of the *CYP21A2* sequence (Fig. 3C), and the allele with L30 (CTG) (Fig. 3F) contained sequence T. However, promoter regions showed the *CYP21A1P* sequence in these two different alleles at nt −126 (T), −113 (A), −110 (C), and −103 (G) (Table 2) (data not shown). This implies that the 5′ end region of the *CYP21A1P* gene with P30 of the *CYP21A2* sequence includes only nt −4 to −103, indicating that the 5′ end region extending beyond nt −103 is the *CYP21A1P* sequence. Furthermore, analysis of the 3′ end region in either the *CYP21A1P* gene (195 chromosomes) (Table 3) or four haplotypes of the *CYP21A2*-like genes and one duplicated *CYP21A2* (five chromosomes) (Table 3) showed the *XA* sequence at nt 3074 (C), 3096 (G), 3146 (T), 3149 (C), 3170 (C), and 3180 (T) (data not shown). This indicates that *CYP21A2*-like genes and duplicated *CYP21A2* genes downstream of the *XA* gene show a chimeric structure with the 5′ end of the *CYP21A2* structure and the *CYP21A1P* sequence in the 3′ end. Therefore, identification of *CYP21A2*-like genes and duplicated *CYP21A2* genes using either a two-step *CYP21A2* amplification [32–34] or a one-step differential PCR amplification of the *CYP21A2* gene [21,35] may produce PCR dropoff. In addition, the use of an allele-specific PCR to detect 10 different mutations [16–18,36] might not be able to distinguish whether *CYP21A2* exists downstream of the *XB* or *XA* gene. Predictably, the MLPA assay [37] might "catch" both the *CYP21A2*-like (Fig. 2C, haplotypes 1, 2, and 3) and *CYP21A2* genes downstream of the *TNXB* gene. A Southern blot analysis by *Taq*I digestion might also show fragments of the 3.7 kb that present the *CYP21A2* gene downstream of the *TNXB* gene, and 2.3 kb may be present as from the *CYP21A1P* fraction in the RCCX region. Therefore, these two methods might lead to a misinterpretation of *CYP21A2* genotyping. Because the orientation of transcription of the *CYP21A2* gene begins from a telomere to a centromere in chromosome 6p21.3, these four haplotypes of *CYP21A2*-like genes (samples 984, 646, 988, and 350;

Please cite this article in press as: L.-P. Tsai et al., Analysis of the *CYP21A1P* pseudogene: Indication of mutational diversity and *CYP21A2*-like and duplicated *CYP21A2* genesCYP21A1P pseudogene –>, Anal. Biochem. (2011), doi:10.1016/j.ab.2011.02.016

see Supplemental Table 2) downstream of the *XA* gene with one more mutational locus (Fig. 2C) might not produce enzyme activity and might not be considered as carriers due to normal functioning of *CYP21A2* downstream of the *XB* gene.

Four normal loci of P30, V281, Q318, and R356 (Table 4) of the *CYP21A2* sequence apparently exist in the *CYP21A1P* pseudogene in our populations and may represent an imprint that gene conversions between these two genes previously occurred, or they might have originated from a by-product of the "incomplete" release of selective pressures on the functional gene in evolution. However, the other four normal loci of the I2 splice, I172, cluster E6, and F306/L307 were not found in the *CYP21A1P* gene with the 3.2-kb *Taq*I-produced fragment. This might have been caused by "complete" selective pressure of the functional gene. Because gene conversions may occur at meiosis between these two genes, we propose that the 10% allelic frequencies (Table 4) of these individuals with the *CYP21A2* sequence in the *CYP21A1P* allele might produce protection against "invasion" of the next gene conversion caused by misalignment.

In conclusion, we have established a protocol to identify *CYP21A2*-like and duplicated *CYP21A2* genes downstream of the *XA* gene that can distinguish the *CYP21A2* gene downstream of the *TNXB* gene, and it indicates that the existence of the *CYP21A2*-like gene is more abundant than that of the duplicated *CYP21A2* gene. We believe that a better understanding of the underlying genetic mechanisms will contribute to more precise diagnoses. We suggest that preparing two PCR products with a full *CYP21A2* gene containing the downstream sequence of the *XB* gene [38] and a full *CYP21A1P* gene containing the downstream sequence of the *XA* gene can faultlessly and accurately detect the molecular *CYP21A2* gene of the RCCX module in CAH due to 21-hydroxylase deficiency.

Acknowledgment

This work was supported by a grant (TCRD-TPE-97-C1-4) from Buddhist Tzu Chi General Hospital of Taiwan.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ab.2011.02.016.

References

[1] M.I. New, R.C. Wilson, Steroid disorders in children: congenital adrenal hyperplasia and apparent mineral corticoid excess, Proc. Natl. Acad. Sci. USA 96 (1999) 12790–12797.
[2] M.I. New, Extensive clinical experience nonclassical 21-hydroxylase deficiency, J. Clin. Endocrinol. Metab. 91 (2006) 4205–4214.
[3] P.C. White, P.W. Speiser, Congenital adrenal hyperplasia due to 21-hydroxylase deficiency, Endocr. Rev. 21 (2000) 245–291.
[4] B.L. Therrel, Newborn screening for congenital adrenal hyperplasia, Endocrinol. Metab. Clin. North Am. 30 (2001) 15–30.
[5] J. Fitness, N. Dixit, D. Webster, T. Torresani, R. Pergolizzi, P.W. Speiser, D.J. Day, Genotyping of CYP21, linked chromosome 6p marker, and a sex-specific gene in neonatal screening for congenital adrenal hyperplasia, J. Clin. Endocrinol. Metab. 84 (1999) 960–966.
[6] Y. Higashi, H. Yoshioka, M. Yamane, O. Gotoh, Y. Fujii-Kuriyama, Complete nucleotide sequence of two steroid 21-hydroxylase genes tandemly arranged in human chromosome: a pseudogene and a genuine gene, Proc. Natl. Acad. Sci. USA 83 (1986) 2841–2845.
[7] P.C. White, M.I. New, B. Doupont, Structure of human steroid 21-hydroxylase genes, Proc. Natl. Acad. Sci. USA 83 (1986) 5111–5115.
[8] P.C. White, M.I. New, B. Dupont, HLA-linked congenital adrenal hyperplasia results from a defective gene encoding a cytochrome P-450 specific for steroid 21-hydroxylation, Proc. Natl. Acad. Sci. USA 81 (1984) 7505–7509.
[9] L. Shen, L.C. Wu, S. Sanlioglu, R. Chen, A.R. Mendoza, A.W. Dangel, M.C. Carroll, W.B. Zipf, C.Y. Yu, Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and the C4B genes in the HLA class III region: molecular cloning, exon-intron structure, composition retroposon, and breakpoint of gene duplication, J. Biol. Chem. 269 (1994) 8466–8476.
[10] S.E. Gitelman, J. Bristow, W.L. Miller, Mechanism and consequences of the duplication of the human C4/P450c21/gene X locus, Mol. Cell Biol. 12 (1992) 2124–2134.
[11] J. Bristow, M.K. Tee, S.E. Gitelman, S.H. Mellon, W.L. Miller, Tenascin-X: A novel extracellular matrix protein encoded by the human XB gene overlapping P450c21B, J. Cell Biol. 122 (1993) 265–278.
[12] C.A. Blanchong, B. Zhou, K.L. Rupert, E.K. Chung, K.N. Jones, J.F. Sotos, W.B. Zipf, R.M. Rennebohm, C.Y. Yu, Deficiencies of human complement component C4A and C4B and heterozygosity in length variants of RP-C4-CYP21-TNX (RCCX) modules in Caucasians: the load of RCCX genetic diversity on major histocompatibility complex-associated disease, J. Exp. Med. 191 (2000) 2183–2196.
[13] H.H. Lee, Diversity of the CYP21P-like gene in CYP21 deficiency, DNA Cell Biol. 24 (2005) 1–9.
[14] C.A. Blanchong, E.K. Chung, K.L. Rupert, Y. Yang, Z. Yang, B. Zhou, J.M. Moulds, C.Y. Yu, Genetic, structural, and functional diversities of human complement components C4A and C4B and their mouse homologues, Slp and C4, Intl. Immunopharmacol. 1 (2001) 365–392.
[15] P.F. Koppens, T. Hoogenboezem, H.J. Degenhart, Duplication of the CYP21A2 gene complicates mutation analysis of steroid 21-hydroxylase deficiency: characteristics of three unusual haplotypes, Hum. Genet. 111 (2002) 405–410.
[16] A. Wedell, B. Stengler, H. Luthman, Characterization of mutations on the rare duplicated C4/CYP21 haplotype in steroid 21-hydroxylase deficiency, Hum. Genet. 94 (1994) 50–54.
[17] S.M. Baumgartner-Parzer, P. Nowotny, W. Waldhäusl, H. Vierhapper, A rare duplicated 21-hydroxylase haplotype and a de novo mutation: a family analysis, J. Clin. Endocrinol. Metab. 88 (2003) 2794–2796.
[18] B. Ezquieta, M. Beneyto, R. Munoz-Pacheco, R. Barrio, M. Oyarzabal, J.L. Lechuga, C. Luzuriaga, F. Hermoso, S. Quinteiro, S. Martinez, Gene duplications in 21-hydroxylase deficiency: the importance of accurate molecular diagnosis in carrier detection and prenatal diagnosis, Prenat. Diagn. 26 (2006) 1172–1178.
[19] S. Kleinle, R. Lang, G.F. Fischer, H. Vierhapper, F. Waldhauser, M. Födinger, S.M. Baumgartner-Parzer, Duplications of the functional CYP21A2 gene are primarily restricted to Q318X alleles: evidence for a founder effect, J. Clin. Endocrinol. Metab. 94 (2009) 3954–3958.
[20] H.H. Lee, J.M. Kuo, H.T. Chao, Y.J. Lee, J.G. Chang, C.H. Tsai, B.C. Chung, Carrier analysis and prenatal diagnosis of congenital adrenal hyperplasia caused by 21-hydroxylase deficiency in Chinese, J. Clin. Endocrinol. Metab. 85 (2000) 597–600.
[21] H.H. Lee, H.T. Chao, H.T. Ng, K.B. Choo, Direct molecular diagnosis of CYP21 mutations in congenital adrenal hyperplasia, J. Med. Genet. 33 (1996) 371–375.
[22] H.H. Lee, S.F. Chang, F.J. Tsai, L.P. Tsai, C.Y. Lin, Mutation of IVS2-12 A/C->G in combination with 707-714delGAGACTAC in the CYP21 gene is caused by deletion of the C4-CYP21 repeat module with steroid 21-hydroxylase deficiency, J. Clin. Endocrinol. Metab. 88 (2003) 2726–2729.
[23] J.M. Bischof, A.P. Chiang, T.E. Scheetz, E.M. Stone, T.L. Casavant, V.C. Scheffied, T.A. Braun, Genome-wide identification of pseudogene capable of disease-causing gene conversion, Hum. Mutat. 27 (2006) 545–552.
[24] G.H. Fujii, A.M. Morimoto, A.E. Berson, J.B. Bolen, Transcriptional analysis of the PTEN/MMAC1 pseudogene, CPTEN, Oncogene 18 (1999) 1765–1769.
[25] D. Zheng, Z. Zhang, P.M. Harrison, J. Karro, N. Carriero, M. Gerstein, Integrated pseudogene annotation for human chromosome 22: evidence for transcription, J. Mol. Biol. 349 (2005) 27–45.
[26] P.M. Harrison, D. Zheng, Z. Zhang, N. Carriero, M. Gerstein, Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability, Nucleic Acids Res. 33 (2005) 2374–2383.
[27] H.H. Lee, The chimeric CYP21P/CYP21 and 21-hydroxylase deficiency, J. Hum. Genet. 49 (2004) 65–72.
[28] H.H. Lee, Chimeric CYP21P/CYP21 and TNXA/TNXB genes in the RCCX module, Mol. Genet. Metab. 84 (2005) 4–8.
[29] P. Concolino, E. Mello, A. Minucci, E. Giardina, C. Zuppi, V. Toscano, E. Capoluongo, A new CYP21A1P/CYP21A2 chimeric gene identified in an Italian woman suffering from classical congenital adrenal hyperplasia form, BMC Med. Genet. 10 (2009) 72.
[30] A. Endoh, L. Yang, P.J. Hornsby, CYP21 pseudogene transcripts are much less abundant than those from the active gene in normal human adrenocortical cells under various conditions in culture, Mol. Cell Endocrinol. 137 (1998) 13–19.
[31] J. Bristow, S.E. Gitelman, M.K. Tee, B. Staels, W.L. Miller, Abundant adrenal-specific transcription of the human P450c21A "pseudogene", J. Biol. Chem. 268 (1993) 12919–12924.
[32] L. Loidi, C. Quinteiro, S. Parajes, J. Barreirot, D.G. Leston, J.M. Cabezas-Agricola, A.M. Sueiro, D. Araujo-Vilar, L. Catro-Feijoo, J. Costas, M. Pombo, F. Dominguez, High variability in CYP21A2 mutated alleles in Spanish 21-hydroxylase deficiency patients, six novel mutations, and a founder effect, Clin. Endocrinol. 64 (2006) 330–336.
[33] D. Owerbach, Y.M. Crawford, M.B. Draznin, Direct analysis of CYP21B genes in 21-hydroxylase deficiency using polymerase chain reaction amplification, Mol. Endocrinol. 4 (1990) 125–131.
[34] D.J. Day, P.W. Speiser, P.C. White, F. Barany, Detection of steroid 21-hydroxylase alleles using gene-specific PCR and a multiplexed ligation detection reaction, Genomics 29 (1995) 152–162.

[35] D. Keen-Kim, J.B. Redman, R.U. Alanes, M.M. Eachus, R.C. Wilson, M.I. New, J.M. Nakamoto, R.G. Fenwick, 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. 7 (2005) 236–246.
[36] R.C. Wilson, J.Q. Wei, K.C. Chen, A.B. Mercado, M.I. New, Rapid deoxyribo nucleic acid analysis by allele-specific polymerase chain reaction for detection of mutations in the steroid 21-hydroxylase gene, J. Clin. Endocrinol. Metab. 80 (1995) 1635–1640.
[37] P. Concolino, E. Mello, V. Toscano, F. Ameglio, C. Zuppi, E. Capoluongo, Multiple ligation-dependent amplification (MLPA) assay for the detection of CYP21A2 gene deletions/duplications in congenital adrenal hyperplasia: first technical report, Clin. Chim. Acta 402 (2009) 164–170.
[38] H.H. Lee, Y.J. Lee, C.Y. Lin, PCR-based detection of the CYP21 deletion and TNXA/TNXB hybrid in the RCCX module, Genomics 83 (2004) 944–950.

Please cite this article in press as: L.-P. Tsai et al., Analysis of the *CYP21A1P* pseudogene: Indication of mutational diversity and *CYP21A2*-like and duplicated *CYP21A2* genesCYP21A1P pseudogene –>, Anal. Biochem. (2011), doi:10.1016/j.ab.2011.02.016