GeNeViSTA

Challenges of Molecular Analysis of Congenital Adrenal Hyperplasia Caused Due to Steroid 21 Hydroxylase Deficiency

Sudhisha Dubey[1], Renu Saxena[1], Vinu Narayan[2], Ratna Dua Puri[1], Ishwar C Verma[1]

[1]*Institute of Medical Genetics and Genomics, Sir Ganga Ram Hospital, Rajinder Nagar, New Delhi, India*

[2]*Rainbow Children's Hospital, Marathahalli, Bengaluru, Karnataka, India*

Correspondence to: Dr Sudhisha Dubey *Email:* sudhishadubey@gmail.com

Abstract

Congenital adrenal hyperplasia (CAH) due to 21-hydroxylase deficiency is an autosomal recessive disorder which results from inherited defects in the steroid 21-hydroxylase enzyme encoded by the *CYP21A2* gene. Molecular analysis of *CYP21A2* is important for confirming the diagnosis, carrier screening, providing accurate genetic counseling, and calculating risk of recurrence in each pregnancy. An interesting feature of the *CYP21A2* gene is its location in the variable genomic regions called RCCX and presence of its highly homologous *CYP21A1P* pseudogene that makes molecular analysis quite challenging as compared to other monogenic disorders. Here we discuss the complexity of the *CYP21A2* gene and the importance of comprehensive molecular analysis of *CYP21A2* for accurate interpretation of the results citing molecular analysis of two interesting CAH cases.

Keywords: Congenital adrenal hyperplasia, *CYP21A2* gene, *CYP21A1P*, pseudogene, variants, MLPA, deletion, duplication

Introduction

Congenital adrenal hyperplasia (CAH) due to 21-hydroxylase deficiency (OMIM# 201910), is an autosomal recessive disorder caused by inherited deficiency of steroid 21-hydroxylase (21OH) enzyme in the steroid biosynthesis pathway in the adrenal cortex. 21OH enzyme acts on progesterone and 17-hydroxyprogesterone (17OHP) and converts these to deoxycortisosterone and 11-beta-hydroxylase respectively, which are further converted into aldosterone and cortisol by other enzymes in the steroidogenic pathway. Deficiency of 21OH enzyme results in shunting of 17OHP and progesterone into the adrenal pathway resulting in excessive production of androgens and deficiency of aldosterone and cortisol (Figure 1). Excessive androgens lead to prenatal virilization in females and rapid somatic growth in both sexes (White & Speiser, 2000). Deficient cortisol level disrupts the negative feedback to the anterior pituitary that results in constant secretion of adrenocorticotropic hormone (ACTH) that overstimulates the adrenal cortex to secret more of cortisol. Due to 21OH deficiency in the adrenal pathway, the cortisol is not secreted and adrenals become hyperplastic due to overstimulation of ACTH in fetal life. That is how this condition obtained its name as "congenital adrenal hyperplasia".

CAH is divided into classic and non-classic (NC) CAH. Classic CAH is again divided into salt-wasting (SW) and simple virilizing (SV) forms. SW-CAH is a severe form characterised by deficiency of both cortisol and aldosterone and found in about 75% of patients. Aldosterone deficiency predisposes SW-CAH patients to develop hyponatremic dehydration which is fatal if not treated with glucocortcoids in time. SV-CAH is a milder form found in about 25% of CAH patients. Aldosterone levels are adequate to maintain sodium balance in the SV form and hence there is normally no salt wasting. The NC form is asymptomatic at birth and presents with various degrees of late-onset hyperandrogenism (White & Speiser, 2000). Prenatal virilisation may or may not be present in the mild NC form but is always present in the SW or SV classic forms.

The overall incidence of CAH in the general population worldwide is between 1 in 10,000 to 1 in 20,000 live births for the classic form of CAH

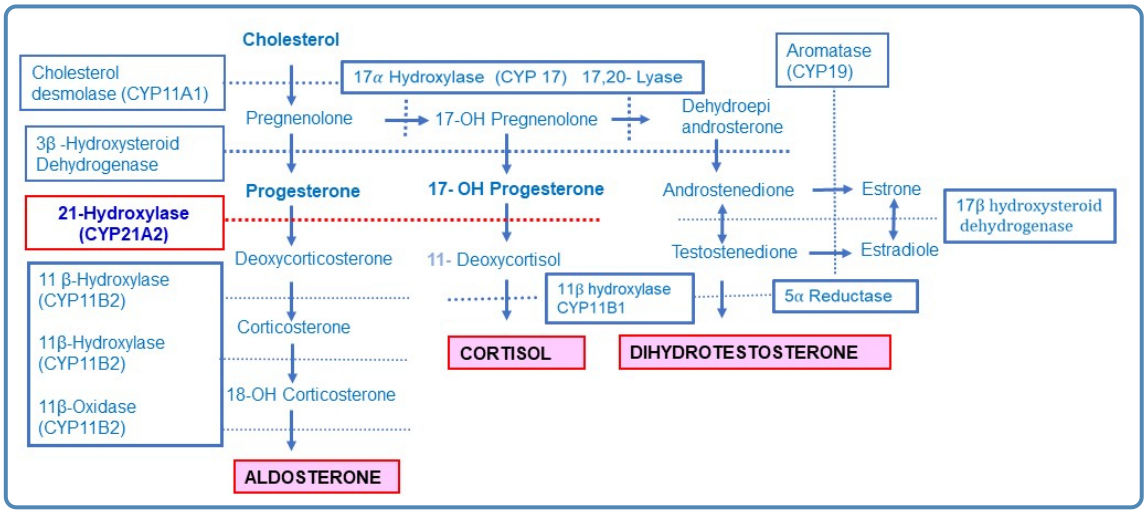

Figure 1 Steroid pathways for biosynthesis of progesterone, aldosterone, cortisol, androgens (testosterone and dihydrotestosterone), and estrogens (estradiol) are arranged from left to right. The enzymatic activities catalyzing each bioconversion are written in boxes. For those activities mediated by specific cytochromes P450, the systematic name of the enzyme ("CYP" followed by a number) is listed in parentheses. CYP11B2 and CYP17 have multiple activities. The planar structures of cholesterol, aldosterone, cortisol, dihydrotestosterone, and estradiol are placed near the corresponding labels (adapted from White & Speiser, 2000).

(Therrell et al., 2001). However, the prevalence of classic CAH in India is 1 in 5762 according to a recent survey (ICMR task force, 2018). Non-classic CAH is one of the most common autosomal recessive disorders in humans and affects approximately 1 in 1,000 individuals (Speiser et al., 1985).

Steroid 21OH enzyme, is encoded by the *CYP21A2* gene located on chromosome 6 (6p21.3) in the HLA class III of the major histocompatibility (MHC) region (Yang et al., 1999). About 30 kb upstream a non-functional pseudogene *CYP21A1P* is located that shares about 98% sequence homology to *CYP21A2*. About 95% of the pathogenic variants are pseudogene derived and are transferred from *CYP21A1P* to *CYP21A2* by gene conversion events (Higashi at al., 1986). The remaining 5% are new/rare and unique for single families or considered as population specific (White & Speiser, 2000; Stikkelbroeck et al., 2003). A compilation of 233 pathogenic variants and their clinical classification have been done recently (Concolino & Costella, 2018).

CYP21A2 gene is a part of the genetic unit comprising of *RP2-C4B-CYP21A2-TNXB* genes known as the RCCX module. Each chromosome bears two RCCX modules; one with the functional *CYP21A2* gene and other with the non-functional *CYP21A1P* as shown in Figure 2. Majority of the individuals have a bimodular haplotype i.e., two modules present on each chromosome. However, three modules have also been reported to be present on one chromosome which is known as the trimodular haplotype. In the trimodular haplotype either two *CYP21A1P* and one *CYP21A2* or one *CYP21A1P* and two *CYP21A2* are present on one chromosome (Figure 2). The later has two copies of functional gene on a chromosome resulting in duplication of the *CYP21A2* gene that complicates the molecular analysis of the *CYP21A2* gene.

In about 20-30% of cases, the large 30kb deletion extends from somewhere between exon 3 of *CYP21A1P* through C4B to the corresponding point in *CYP21A2* yielding a single copy with 5´ end of *CYP21A1P* and 3´ end of *CYP21A2*, also known as the chimeric gene. Nine different chimeras have been reported depending on the extent of deletion involved (Chen et al., 2012). Extent of the deletion also helps in determining the genotype -phenotype correlation (Narasimhan et al., 2019).

Materials and methods

Written informed consent was obtained from the parents of both patients. About 100 ng of each genomic DNA was subjected to selective

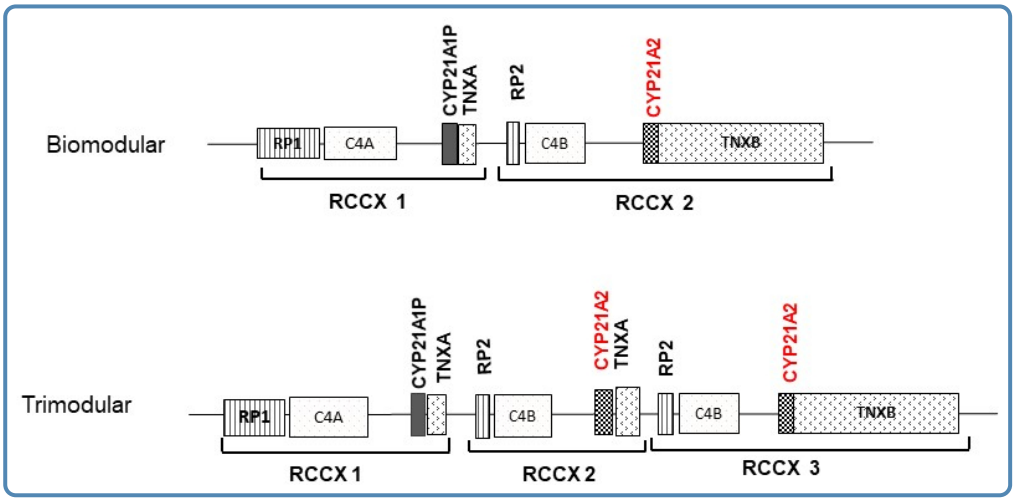

Figure 2 Schematic diagram of the organization of the *RCCX* modules. The most common is the bimodular haplotype with two *RCCX* modules, one with pseudogene *CYP21A1P* and other functional *CYP21A2* gene. Trimodular haplotype with three *RCCX* modules can result in duplication of the *CYP21A2* gene. C4 (C4A and C4B) gene encodes the fourth component of the serum complement. RP2, a truncated copy of RP1, encodes the threonine kinase enzyme and *TNXB* encodes tenascin-X an extracellular matrix protein. *TNXA* is a non-functional homologue of the *TNXB* gene (adapted from Sweeten et al., 2008).

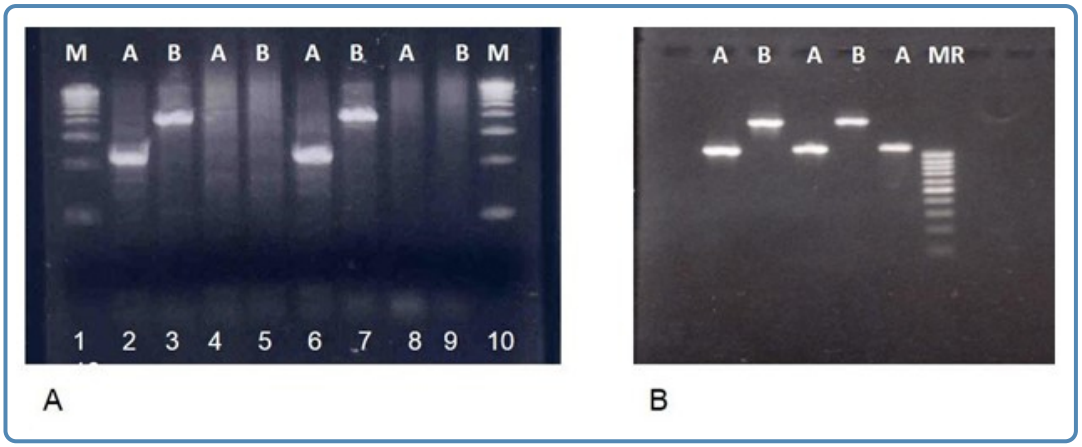

Figure 3 **A**. PCR amplification of the *CYP21A2* gene into two fragments; fragment A (1130bp) and fragment B (2127bp). M- DNA Ladder; Lanes 2 & 6-Fragment A; Lanes 3 & 7 – Fragment B; Lane 4-5, 8-9 – Absence of bands or amplification indicating gene deletion. **B**. Purified PCR products of fragment A and B with MassRuler (MR). (Dubey et al., 2017)

amplification of *CYP21A2* into two large fragments with two sets of primers highly specific to the active i.e., *CYP21A2* gene (Figure 3A). Absence of bands indicate the deletion of 8 bp of exon 3 or whole of the active gene which is confirmed by MLPA. These fragments were purified using the Qiagen kit (QIAamp PCR Clean-up, Qiagen GmbH, Hilden, Germany) and quantified with MassRuler (Fermentas Life Sciences, Thermo Fisher Scientific, Waltham MA, USA) (Figure 3B) (Dubey et al., 2017). Purified products were subjected to direct sequencing using ABI 3500 Genetic Analyser (PE Applied Biosystems, Thermo Fisher Scientific, Waltham MA, USA). Pathogenic variants were screened using Chromas v2.4 and SeqScape v2.1.1 (Applied Biosystems) against the NCBI

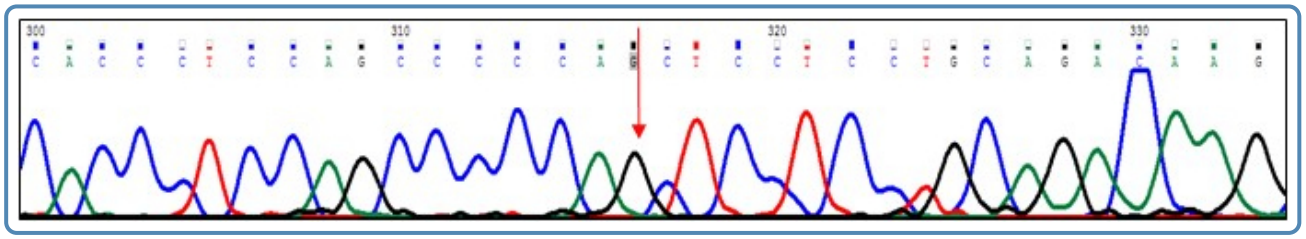

Figure 4 Partial electropherogram showing homozygous I2g (c.293-13A/C>G) mutation detected in the proband (Patient 1).

reference sequence NM_000500 and transcript ID ENST00000418967. Multiple ligation dependent probe amplification (MLPA) was done using Salsa MLPA Kit P050-C1 (MRC-Holland, Amsterdam, The Netherlands) to detect deletions and duplications.

Patient description and results

Patient 1: A five-years-old female child presented with ambiguous genitalia at birth. She had complete labial fusion and clitoral hypertrophy. Her karyotype was normal female (46, XX) and ultrasound-abdomen revealed bilateral ovaries. She had elevated levels of 17 OHP (greater than 37 ng/mL), renin (greater than 500 ng/mL/hour), potassium (7.8 mEq/L) and low level of sodium (116 mEq/L). She was reported to have seven mutations i.e., I2g (c.293-13A/C>G) (intron2), c.332_339delGAGACTAC (exon 3), c.515T>A (exon 4), c.710T>A (exon 6), c.713T>A (exon 6), c.719T>A (exon 6), and c.923_924insT (exon 7) by NGS. All mutations were in heterozygous form except splice site mutation I2g (c.293-13A/C>G) in intron 2 of the *CYP21A2* gene. Snapshots of Integrative genome viewer (IGV) software and MLPA ratio chart were also provided that clearly illustrated presence of these mutations.

The proband was referred to us for validation and segregation of pathogenic variants in her parents, her paternal aunt and the aunt's husband, as her aunt was pregnant and the family wanted prenatal diagnosis (PND) to be done. Sequencing of the proband was carried out to validate the seven reported pathogenic variants. However, only I2g pathogenic variant was found in homozygous state and all other mutations were clearly absent (Figure 4). To know whether this mutation was in homozygous or hemizygous form, MLPA was carried out for detection of deletion. Half dosage was seen in the probes covering exon 3,4,6 and 7 indicating heterozygous deletion from exon 3 to 7. I2g (intron 2 splice) mutation was found in homozygous state by the two probes included in the MLPA kit P050-C1 for detection of I2g mutation. (Figure 5).

Her parents were analysed for segregation of mutations by Sanger sequencing and MLPA. Mother was found to carry the I2g mutation as expected but father was negative for the same. He was then checked for deletion by MLPA that showed normal dosage for all probes indicating that he was negative for the deletion which was unexpected.

Paternal aunt (sister of proband's father) was checked for deletion and duplication by MLPA. She was found to harbor a heterozygous duplication shown by 3 copies of *CYP21A2* (Figure 6). After analysing results of paternal aunt, MLPA results of the father were reinterpreted and it was inferred that father harboured both a duplication and a deletion together, due to which he was showing normal dosage. And his sister had inherited the duplicated allele but not the deletion, and hence was not a carrier of CAH. Her husband too was checked and he was found to be negative for deletion and duplication.

Hence it was confirmed that the proband was compound heterozygous for whole gene deletion and I2g mutation. The deletion was inherited from the father and the I2g mutation from the mother. Proband's aunt and uncle were counseled about the insignificant risk of having a child affected with CAH.

Patient 2: A five-years-old female child clinically confirmed to have CAH was referred to our genetic clinic for molecular analysis. Her mother was 18 weeks pregnant and the family wanted PND to be done.

Deletions being more common in the *CYP21A2* gene, MLPA was first done that showed half dosage of exon 4,6, and 7 indicating heterozygous deletion from exon 4-7 (Figure

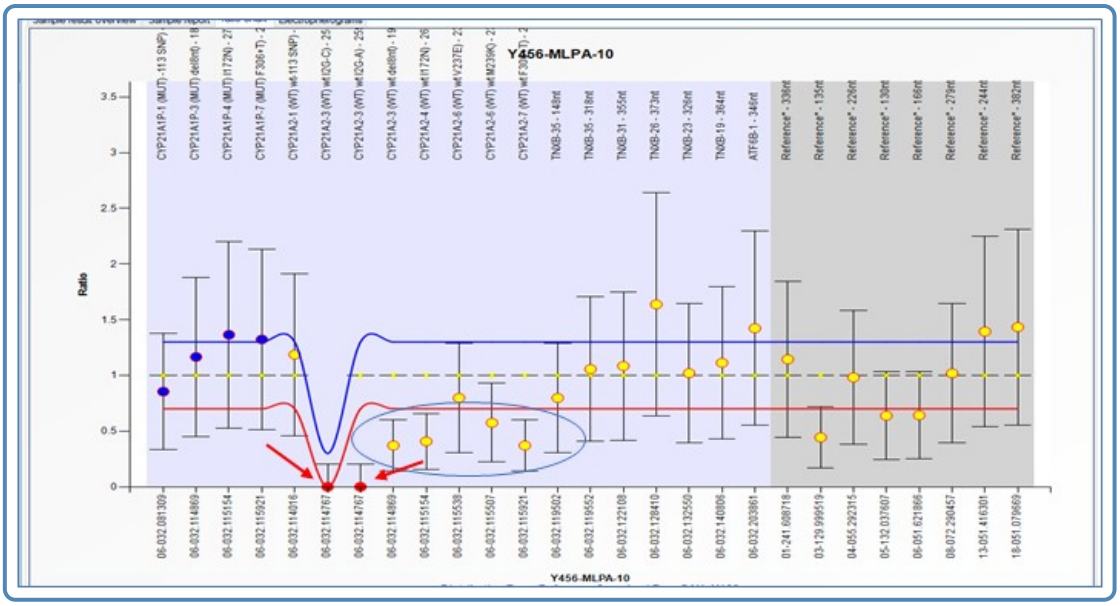

Figure 5 MLPA analysis using Coffalyser software showing heterozygous deletion of exon 3- 7 of *CYP21A2* gene in Patient 1. Deletions of exons 3 and 7 are marked by red circle. SALSA MLPA kit P050-C1 was used to detect deletion in our patients. Normal alleles A and C at I2g showing zero value (shown by arrows) indicate absence of both A and C alleles and presence of homozygous allele G. Normalized peak height ratio between 0.7 and 1.3 was considered as normal in patient DNA w.r.t. control DNA.

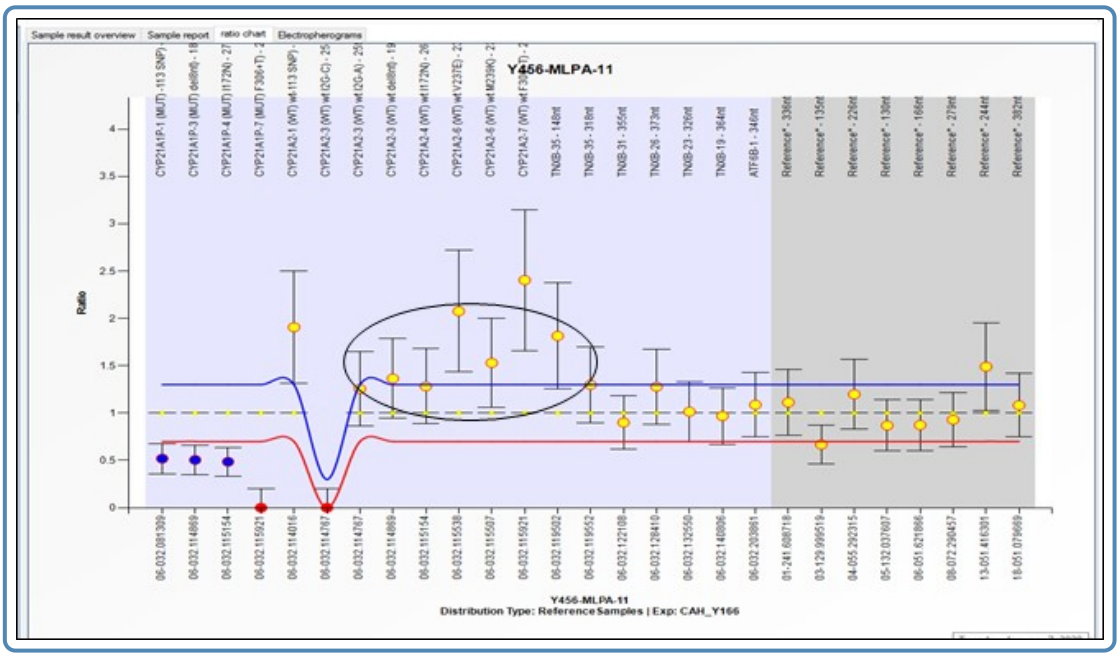

Figure 6 MLPA analysis using Coffalyser software showing heterozygous duplication indicated by the red circle. All probes fall above the normal ratio (1.5) indicating three copies of *CYP21A2* gene in the paternal aunt of patient 1. SALSA MLPA kit P050-C1 was used to detect deletion in our patients. Normalized peak height ratio between 0.7 and 1.3 was considered as normal in patient DNA w.r.t. control DNA.

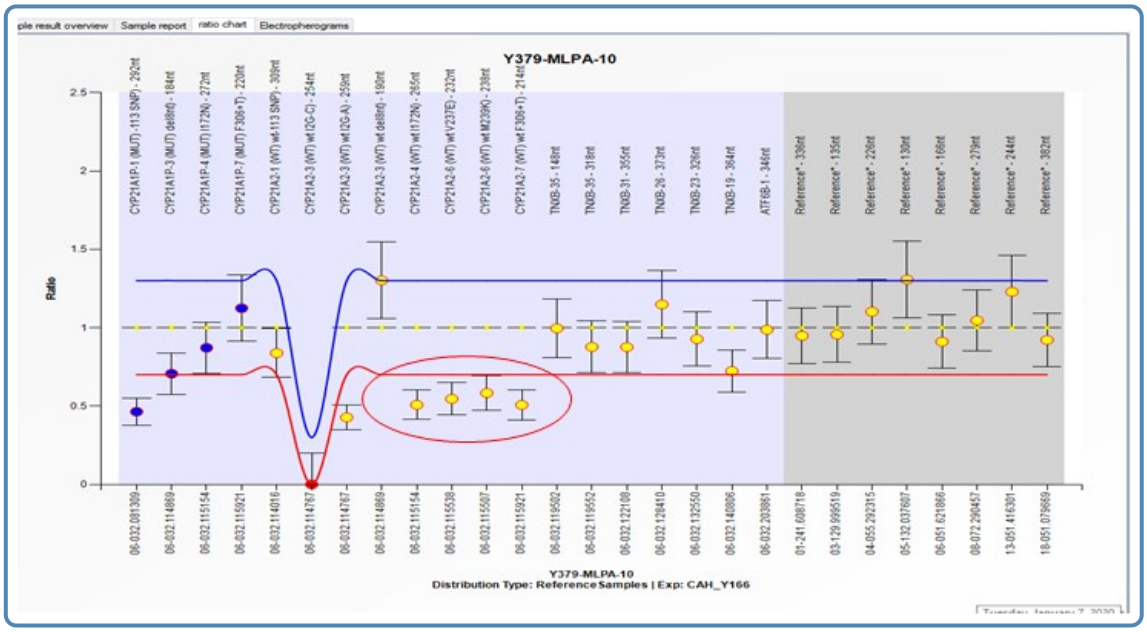

Figure 7 MLPA analysis using Coffalyser software showing half ratios of exon 4-7 indicating heterozygous deletion from exon 4-7 of *CYP21A2* gene in Patient 2. SALSA MLPA kit P050-C1 was used to detect deletion in our patients. Normalized peak height ratio between 0.7 and 1.3 was considered as normal in patient DNA w.r.t. control DNA.

7). To look for second mutation, Sanger sequencing was done and the proband was found to harbour c.515T>A (p.Ile172Asn) in exon 4, E6 cluster [c.710T>A (p.Ile236Asn); c.713T>A (p.Val237Glu); c.719T>A (p.Met239Lys)] in exon 6, c.923_924insT (p.Leu306+T) in exon 7, and c.955C>T (p.Gln319Ter) in exon 8, all in heterozygous form (Figure 8).

Her parents were then checked for segregation analysis to confirm whether these mutations were present in *cis* or *trans*. Mother was found to have E6 cluster [p.Ile236Asn, p.Val237Glu, p.Met239Lys], p.Leu306+T and p.Glu319Ter mutations, and father was heterozygous for the p.Ile172Asn mutation. Hence it was confirmed that the child was compound heterozygous for the point mutations.

Discussion

Molecular genetic diagnosis of CAH is more complicated than for many other monogenic disorders due to the location of the *CYP21A2* gene in the highly variable genomic region with more than one RCCX repeat unit on the same chromosome. Presence of a non-functional pseudogene further complicates the amplification of the functional gene. The 11 most common mutations known to cause CAH are present in the pseudogene too. Due to this reason, it is extremely important that the functional gene should only be amplified in the background of pseudogene. It is quite difficult as there is not much difference in the sequence between the two genes. The most significant difference is the 8 base pairs GAGACTAC present in exon 3 of *CYP21A2* and these 8 base pairs are deleted in exon 3 of *CYP21A1P*. This '8bp site' has been exploited extensively to design primers for selective amplification of the active gene. To be twice as sure, two primers - forward as well as reverse, were designed at the wild type sequence of the "8bp site" to amplify the *CYP21A2* gene into two large fragments. This ensures that amplification occurs only when both primers bind on the wild type sequence at the '8bp site'. Absence of amplification indicates the absence of the active gene or presence of the homozygous 8 bp deletion or presence of only the pseudogene (Figure 3A). The extent of deletion can then be analysed by MLPA.

It is important to know that due to presence of the pseudogene, the capture-based NGS approach is not considered appropriate as it may interfere with the analysis and give erroneous results. Recently, a customized work flow involving

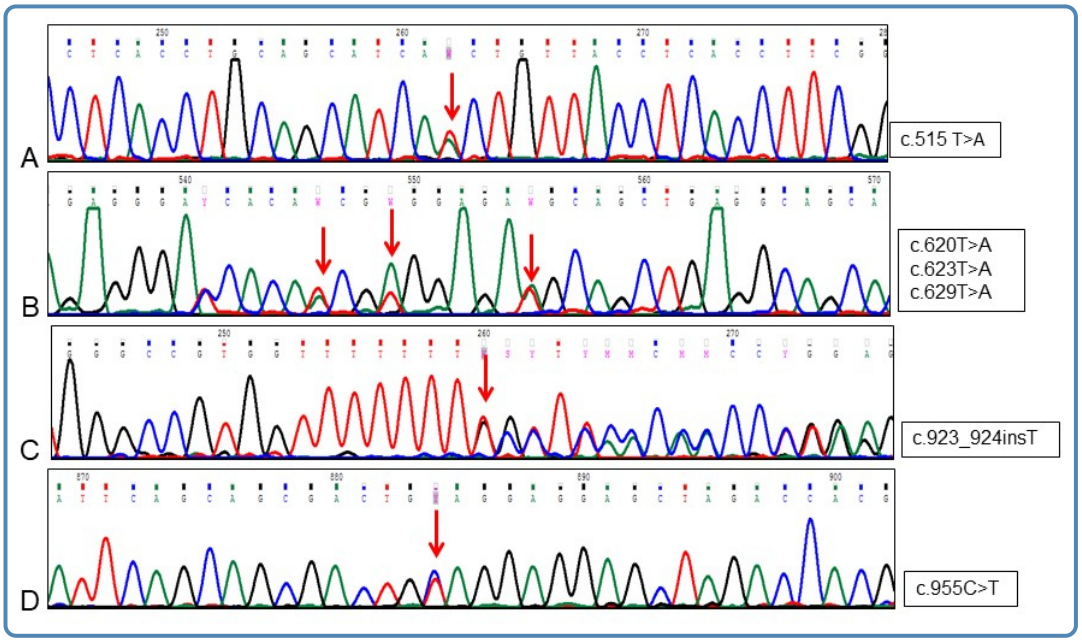

Figure 8 Partial electropherograms showing mutations detected in Patient 2. A. Mutation c.515T>A in exon 4; B. E6 cluster mutation (c.710T>A, 713T>A, 719T>A) in exon 6; C. c.923_924 insT in exon 7; and D. c.955C>T in exon 8 of *CYP21A2* gene. All mutations are shown by arrows.

selective amplification of *CYP21A2* followed by NGS has been used to correctly detect variants in CAH patients (Gangodkar et al., 2020).

In Patient 1, all pathogenic variants except I2g were reported in heterozygous form by NGS. These pathogenic variants appeared in IGV as heterozygous state as half reads were generated from the active gene and half reads from the pseudogene that harboured the corresponding mutant allele. I2g variant was seen in homozygous form as there was no wild type allele present in the proband. MLPA Kit P050-C1 probes are complimentary to the sequences encompassing the pathogenic variants present in different exons, thus their ratios indicate deletions as well as zygosity of the variants present in the sample. In this patient, half ratios of the probes were wrongly interpreted as heterozygous variants. Since MLPA results were concordant with NGS results, all variants were reported without validating by Sanger sequencing. However, these ratios were actually indicating deletions in exons 1-7. Sanger validation in this patient could have avoided the erroneous interpretation.

The scenario for Patient 2 was completely opposite to that of Patient 1. In Patient 2, MLPA was first performed and heterozygous deletion of exon 3 to 7 was detected. Only after performing Sanger sequencing, the proband was found to harbour 4 pathogenic variants, [p.Ile172Asn, E6 cluster, p.Leu306+T, p.Glu319Ter], all in heterozygous state. Since there is no probe available for exon 8 in the MLPA Kit P050-C1 used, p.Glu319Ter a common pathogenic variant present in exon 8 was not picked up by MLPA. Therefore, one should keep in mind while analysing the MLPA results that half ratio (0.5) or zero ratio observed in any exon indicates heterozygous or homozygous deletion of the corresponding exon respectively. However, these ratios could also indicate the presence of heterozygous/ homozygous variant in that exon as seen in Patient 2. Thus, MLPA results should always be complemented with Sanger sequencing. On the contrary, whenever homozygous variants are detected by Sanger sequencing, MLPA should be done to verify whether the pathogenic variant is homozygous or hemizygous.

Therefore, for molecular analysis of the *CYP21A2* gene, more than one method should be used for comprehensive analysis. For example, while performing PND for the I2g variant, microsatellite linkage analysis should also be performed in addition to direct DNA sequencing and MLPA, as this variant is known to have a high rate of allele drop out (Tsai & Lee, 2012)

Secondly, mutant alleles must be segregated in the parents to verify their presence on different alleles for correct interpretation of the molecular genetics results, as observed in Patient 2. In Patient 1, we could have missed carrier status of the father if we had not analysed the proband's aunt. Duplications and deletions of the *CYP21A2* gene are now being detected relatively frequently due to the use of MLPA, a valid alternative to Southern blotting. However, the interpretation of MLPA results requires extensive knowledge of *CYP21A2* gene rearrangements (Concolino et al., 2009). Duplications have been reported to be quite frequent in Caucasians (Parajes et al., 2008), however, no data is available from Indian subjects. Duplications have great impact on the carrier status of an individual therefore they represent a significant pitfall in the molecular diagnosis of steroid 21-hydroxylase deficiency (Koppens et al., 2002). Hence, it is imperative to screen duplications in all couples referred for preconceptional carrier screening.

References

1. Chen W, et al. Junction site analysis of chimeric *CYP21A1P*/*CYP21A2* genes in 21-hydroxylase deficiency. Clin Chem. 2012; 58: 421–430.
2. Concolino P, et al. Multiplex ligation-dependent probe amplification (MLPA) assay for the detection of *CYP21A2* gene deletions/duplications in congenital adrenal hyperplasia: first technical report. Clin Chim Acta. 2009; 402: 164–170.
3. Concolino P, Costella A. Congenital Adrenal Hyperplasia (CAH) due to 21-Hydroxylase Deficiency: A Comprehensive Focus on 233 Pathogenic Variants of *CYP21A2* Gene. Mol Diagn Ther. 2018; 22: 261–280.
4. Dubey S, et al. Prenatal diagnosis of steroid 21-hydroxylase-deficient congenital adrenal hyperplasia: Experience from a tertiary care centre in India. Indian J Med Res. 2017; 145: 194–202.
5. Gangodkar 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: 189–198.
6. Higashi Y, et al. Complete nucleotide sequence of two steroid 21-hydroxylase genes tandemly arranged in human chromosome: a pseudogene and a genuine gene. Proc Natl Acad Sci U.S.A. 1986; 83: 2841–2845.
7. ICMR Task Force on Inherited Metabolic Disorders. Newborn screening for congenital hypothyroidism and congenital adrenal hyperplasia. Indian J Pediatr. 2018; 85: 935–40.
8. Koppens PF, et al. Duplication of the *CYP21A2* gene complicates mutation analysis of steroid 21-hydroxylase deficiency: characteristics of three unusual haplotypes. Hum Genet. 2002; 111: 405–410.
9. Narasimhan ML, Khattab A. Genetics of congenital adrenal hyperplasia and genotype-phenotype correlation. Fertil Steril. 2019; 111: 24–29.
10. Parajes S, et al. High frequency of copy number variations and sequence variants at *CYP21A2* locus: Implication for the genetic diagnosis of 21-hydroxylase deficiency. PLoS One. 2008; 3: e2138.
11. Speiser PW, et al. High frequency of nonclassical steroid 21-hydroxylase deficiency. Am J Hum Genet. 1985; 37: 650–667.
12. Stikkelbroeck NM, et al. CYP21 gene mutation analysis in 198 patients with 21-hydroxylase deficiency in The Netherlands: six novel mutations and a specific cluster of four mutations. J Clin Endocrinol Metab. 2003; 8: 3852–3859.
13. Sweeten TL, et al. C4B null alleles are not associated with genetic polymorphisms in the adjacent gene *CYP21A2* in autism. BMC Med Genet. 2008; 9: 1. doi: 10.1186/1471-2350-9-1
14. Therrell B. Newborn screening for congenital adrenal hyperplasia. Endocrinol Metab Clin North Am. 2001; 30: 15–30.
15. Tsai LP, Lee HH. Analysis of *CYP21A1P* and the duplicated *CYP21A2* genes. Gene 2012; 506: 261–262.
16. White PC, Speiser PW. Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. Endocrine Rev. 2000; 2: 245–91.
17. Yang Z, et al. Modular variations of the human major histocompatibility complex class III genes for serine/threonine kinase RP, complement component C4, steroid 21-hydroxylase CYP21, and tenascin TNX (the RCCX module). A mechanism for gene deletions and disease associations. J Biol Chem. 1999; 274: 12147–12156.