Risk Management and Healthcare Policy

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Optimized Homologous Sequence Alignment for the Identification of CYP21A2 Variants in 21-Hydroxylase Deficiency Using Next-Generation Sequencing Technology

Yibo Chen[1], Qi Yu[2], Lisha Ge[3,4], Lixin Weng[5], Xiaoli Pan[3,4], Xiaoxia Zhou[6], Nani Zhou[6], Yanjie Wang[6], Jia Jia[6], Haibo Li[3,4,7]

[1]Department of Clinical Laboratory, Women and Children's Hospital of Ningbo University, Ningbo, 315012, People's Republic of China; [2]Neonatal Screening Center, Women and Children's Hospital of Ningbo University, Ningbo, 315012, People's Republic of China; [3]The Central Laboratory of Birth Defects Prevention and Control, Women and Children's Hospital of Ningbo University, Ningbo, 315012, People's Republic of China; [4]Ningbo Key Laboratory of Genomic Medicine and Birth Defects Prevention, Women and Children's Hospital of Ningbo University, Ningbo, 315012, People's Republic of China; [5]Department of Clinical Laboratory, Ningbo Yinzhou District Maternity and Child Healthcare Hospital, Ningbo, 315000, People's Republic of China; [6]Shanghai Fujungenetics Biotechnology Co., Ltd, Shanghai, 201900, People's Republic of China; [7]Ningbo Key Laboratory for the Prevention and Treatment of Embryogenic Diseases, Women and Children's Hospital of Ningbo University, Ningbo, 315012, People's Republic of China

Correspondence: Haibo Li, The Central Laboratory of Birth Defects Prevention and Control, Ningbo Key Laboratory of Genomic Medicine and Birth Defects Prevention, Ningbo Key Laboratory for the Prevention and Treatment of Embryogenic Diseases, Women and Children's Hospital of Ningbo University, Ningbo, 315012, People's Republic of China, Tel +86 574 83882401, Email lihaibo-775@163.com; Jia Jia, Shanghai Fujungenetics Biotechnology Co., Ltd, Shanghai, 201900, People's Republic of China, Tel +86 18658176000, Email jiajia@fujungenetics.com.cn

Risk Management and Healthcare Policy downloaded from https://www.dovepress.com/
For personal use only.

Objective: This study aimed to develop a novel homologous sequence analysis technique using high-throughput sequencing data to enhance CYP21A2 mutation detection. The approach leverages next-generation sequencing to overcome existing limitations and improve 21-hydroxylase deficiency diagnostic accuracy.

Methods: From April 21, 2022, to February 21, 2023, a total of 100 unrelated participants were enrolled at the Women and Children's Hospital of Ningbo University, selected based on clinical manifestations and genetic testing results. The study used next-generation sequencing combined with a homologous sequence alignment (HSA) algorithm, which calculated the sequencing read ratios from homologous regions to identify pathogenic or likely pathogenic variants in the CYP21A2 gene. All detected variants were further validated using long-range PCR or multiplex ligation-dependent probe amplification. The accuracy of the HSA algorithm was systematically assessed.

Results: Among the 100 participants, 84 were identified as carriers of CYP21A2 mutations, while 16 were diagnosed with 21-hydroxylase deficiency. A total of 107 pathogenic mutations were detected using the homologous sequence alignment algorithm, comprising of 99 single nucleotide variants or insertions/deletions, 6 copy number variants, and 8 fusion mutations. Additionally, eight cases of CYP21A2-CYP21A1P gene conversions were identified based on HSA scores and confirmed through long-range PCR or multiplex ligation-dependent probe amplification. The algorithm demonstrated a positive predictive value of 96.26% for identifying mutations in CYP21A2. The most frequently observed mutations included c.955C > T, c.844G > T, c.293–13C > G, c.518T > A, and exon-level deletions.

Conclusion: In genetic testing, particularly when addressing misalignment challenges associated with highly homologous genes such as CYP21A2, application of the HSA algorithm enables accurate mutation detection using commonly employed short-read sequencing methods. Through the characterization of homologous sequence features and optimization of the HSA algorithm, accurate mutation detection can be achieved in more homologous gene families (eg, HBA1/HBA2, SMN1/SMN2, GBA/GBAP1).

Keywords: CYP21A2, 21-hydroxylase deficiency, fusion mutations, highly homologous sequences, HSA algorithm, next generation sequencing

Received: 18 February 2025
Accepted: 26 July 2025
Published: 16 September 2025

© 2025 Chen et al. This work is published and licensed by Dove Medical Press Limited. The full terms of this license are available at https://www.dovepress.com/terms.php and incorporate the Creative Commons Attribution – Non Commercial (unported, v4.0) License (http://creativecommons.org/licenses/by-nc/4.0/). By accessing the work you hereby accept the Terms. Non-commercial uses of the work are permitted without any further permission from Dove Medical Press Limited, provided the work is properly attributed. For permission for commercial use of this work, please see paragraphs 4.2 and 5 of our Terms (https://www.dovepress.com/terms.php).

Introduction

Congenital adrenal hyperplasia (CAH) is an autosomal recessive disorder resulting from a deficiency in steroid synthase.[1] This disease has a global incidence of being 1 in 10,000 and 1 in 2,000,[2] while in Zhejiang Province, China, it is reported to be 0.38 per 10,000.[3] Approximately 95% of CAH cases are attributed to 21-hydroxylase deficiency (21-OHD) caused by mutations in the CYP21A2 gene.[4]

The CYP21A2 gene encodes cytochrome P450c21 (CYP21), which plays a key role in the biosynthesis of glucocorticoids. During steroidogenesis, 21-hydroxylase catalyzes the conversion of 17-hydroxyprogesterone (17-OHP) to 11-deoxycortisol, which is further converted to cortisol and aldosterone through subsequent reactions. A loss of 21-hydroxylase activity disrupts this pathway, leading to an accumulation of 17-OHP and overproduction of androgens. This disruption manifests clinically as feminization and male pseudo-precocious puberty.[5]

The clinical phenotype of CAH, particularly 21-OHD, is closely related to the residual activity of the 21-hydroxylase enzyme. Severe deficiency, with enzyme activity below 1%, can result in significant salt wasting and early adrenal crisis. When enzyme activity is between 1% and 2%, aldosterone production may be sufficient to prevent salt wasting under normal conditions, although stress-induced adrenal crises remain a risk and may contribute to neonatal mortality.[6] Hyperandrogenism associated with this condition can lead to abnormal sexual development and gonadal axis dysfunction during childhood, while adults may experience long-term complications such as infertility, tumors, obesity, hypertension, osteoporosis, and reduced quality of life. Notably, when enzyme activity ranges between 20% and 50%, cortisol synthesis remains largely unaffected.[7]

Clear treatment principles are well-established for patients diagnosed with 21-OHD. Early etiological diagnosis is key, as it enables effective clinical interventions and timely genetic counseling for affected individuals.[8,9] Current expert consensus in China recommends the use of gas chromatography-mass spectrometry (GC-MS) or liquid chromatography-tandem mass spectrometry (LC-MS/MS) for analyzing urine and blood samples.[10,11] However, these traditional diagnostic methods are associated with limitations, including high false-positive rates, inability to detect non-classic forms of CAH, and challenges in accurately genotyping CAH, which is crucial for timely and appropriate management.[12–14]

Next-generation sequencing (NGS) technology has significantly advanced the diagnostic landscape for CAH. NGS not only facilitates the differentiation of various forms of CAH based on distinct diagnostic criteria and also enhances the efficiency of the diagnostic process. By integrating with or streamlining conventional mass spectrometry methods, NGS reduces diagnostic time and associated treatment costs.[15]

The CYP21A2 gene is located within a cluster of neighboring genes known as the RCCX (RP-C4-CYP21-TNX) module.[16,17] Tandem duplication events frequently occur in this 30kb genomic region, leading to variability in the number of RCCX modules among individuals. The most common configurations include two consecutive modules: RP1-C4A-CYP21A1P-TNXA and RP2-C4B-CYP21A2-TNXB.[16] Notably, CYP21A1P is a pseudogene that shares high sequence homology with CYP21A2 (about 98%) but is nonfunctional due to the presence of 15 mutations.[17] Accurate diagnosis of 21-OHD therefore necessitates precise characterization of mutations in the CYP21A2 gene.

Despite advancements in NGS, existing technologies face limitations in differentiating highly homologous sequences and detecting copy number variations (CNVs), particularly within gene families and pseudogenes. These challenges complicate the diagnostic process.[2,18] Although increasing sequencing depth and improving data quality can mitigate some of these issues, such measures significantly raise costs and increase the likelihood of experimental errors due to excessive sequencing depth.[19,20] Conventional sequence alignment algorithms such as BLAST struggle to resolve mismatches in highly homologous regions like the CYP21A2 gene and its pseudogene, leading to misalignment and complicating accurate variant detection. To address this, Illumina Dragen has developed a specialized tool—the CYP21A2 caller—that employs haplotype-specific analysis of the RCCX region.[21] This method has demonstrated high performance in whole-genome sequencing (WGS) data, achieving a positive predictive value (PPV) of 98.5% (201/204), as it fully captures the sequence context necessary for analyzing gene conversion events between CYP21A2 and its pseudogene. However, its utility remains limited in clinical practice due to its commercial nature, high cost, and lack of validation on short-read exome data, which often lacks full RCCX coverage. Additionally, while third-generation sequencing technologies paired with tools like FreeBayes or pbmm2 offer promise for analyzing complex genomic regions, they are currently constrained by high costs and technical complexity,

making them impractical for widespread clinical deployment. In routine diagnostics, the genetic identification of congenital adrenal hyperplasia (CAH) still relies heavily on phenotype-driven methods and single-gene tests such as MLPA or LR-PCR, which may miss atypical cases and cannot support large-scale, multiplexed gene screening. Against this backdrop, we developed the HSA algorithm—a targeted analytical method designed to work effectively with only CYP21A2 exonic regions using short-read sequencing data. This cost-efficient and scalable approach not only overcomes the limitations of traditional tools in homologous regions but also enhances variant detection accuracy. By enabling reliable analysis of homologous gene pairs from exonic short-read data, the HSA algorithm holds strong potential for large-scale screening and simultaneous multi-gene diagnostics in clinical settings.

Consequently, the development of robust analytical methods to improve exome-focused NGS discrimination of homologous sequences offers a more cost-effective and reliable solution for diagnosing 21-OHD.

The objective of this study is to develop a homologous sequence comparison algorithm to better identify the mutations in the CYP21A2 gene. The broader objective is to improve the diagnostic accuracy of CAH and establish efficient screening frameworks, facilitating the expanded application of tertiary prevention gene detection technologies for birth defect prevention in regional populations.

Materials and Methods

Study Participants

From April 21, 2022, and February 21, 2023, a total of 100 unrelated participants with CYP21A2 mutations were enrolled at the Women and Children's Hospital of Ningbo University, based on clinical manifestations and genetic testing. Among these participants, 16 were diagnosed with 21-hydroxylase deficiency, while the remaining 84 were identified as carriers of pathogenic mutations in CYP21A2 (Table S1).

All patient samples were collected following the principles outlined in the Helsinki Declaration, with written informed consent obtained from each participant. Clinical data for all participants were reviewed and approved by the ethics committee of the Women and Children's Hospital of Ningbo University (approval number: 2021-ky-036).

Experimental Methods

High Throughput Sequencing (NGS) and Data Analysis

Genomic DNA was purified from frozen blood samples using the QIAamp® DNA Blood Mini Kit (Qiagen, Germantown, MD, USA), following the protocol provided by the manufacturer. A total of 300–500 ng of genomic DNA was fragmented and size-selected to 400–600 bp. Sequencing adapters were ligated, and the DNA fragments were amplified by PCR to prepare the standard library. Whole exome sequencing (WES) libraries were prepared using the Twist Human Core Exome Multiplex Hybridization Kit. For samples undergoing custom WES (cWES) testing, library amplification was performed using HiFi HotStart ReadyMix (KAPA), and exome hybrid capture was conducted following the standard protocol provided by Integrated DNA Technologies (IDT, Coralville, IA, USA).

The libraries were sequenced on an Illumina NovaSeq platform (Illumina, Inc., San Diego, CA, USA). The raw image files were processed using bcl2fastq2 conversion software v2.20 (Illumina, Inc., San Diego, CA, USA). Sequencing reads were aligned to the human reference genome (hg19/GRCh37) using the Burrows-Wheeler Alignment (BWA) tool, and PCR duplicates were removed using Picard software v1.57 (http://picard.sourceforge.net/).

Identification and Interpretation of Variations

The Genomic Analysis Toolkit (GATK4) (https://software.broadinstitute.org/gatk/) was used for mutation detection. Single nucleotide variants (SNVs) and insertions or deletions (Indels) were annotated and interpreted using ANNOVAR software (http://www.openbioinformatics.org/annovar/). CNVs were analyzed using an internally developed tool, CNVexon, which calculates copy number variations based on regional coverage and depth.[22,23] This tool was used to detect heterozygous deletions or amplifications at the exon level.

All identified variants were manually reviewed according to the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) guidelines for mutation interpretation. Variants were classified into five categories: "pathogenic", "likely pathogenic", "uncertain significance", "likely benign", and "benign." Variants

categorized as "pathogenic" or "likely pathogenic", including SNVs and CNVs, were compared with exprmental results of long range polymerase chain reaction (LR-PCR), Sanger sequencing, and multiplex ligation-dependent probe amplification (MLPA) techniques.[24] A comprehensive list of all identified variants is available at https://www.ncbi.nlm.nih.gov/SNP/snp_viewTable.cgi?handle=NBFRCYP.

Protocol for LR-PCR

For each reaction, 25 μL of mixture was prepared in PCR tubes or plates. The mixture consisted of 12.5 μL EmeraldAmp® MAX HS PCR Master Mix (2X Premix), 0.5 μL LD-PCR primer F (10 μM), 0.5 μL LD-PCR primer R (10 μM), 2.5 μL genomic DNA, and 9.5 μL ddH_2O. The tubes or plates were placed in a thermal cycler, and LR-PCR was conducted under the following conditions: initial denaturation at 98 °C for 3 minutes; 22 cycles of denaturation at 98 °C for 10 seconds, annealing at 68 °C (decreasing by 0.5 °C/cycle) for 30 seconds, and extension at 72 °C for 6 minutes; 25 additional cycles of denaturation at 98 °C for 10 seconds, annealing at 57 °C for 30 seconds, and extension at 72 °C for 6 minutes. A final extension step at 72 °C for 10 minutes was conducted. Samples were then held at 10 °C for storage.

After LR-PCR, 5 μL of the product was loaded onto a 1.5% agarose gel in 1× TAE buffer, and electrophoresis was conducted at 120 V for 30 minutes. The presence of a band indicated successful amplification, allowing progression to the next step.

For subsequent PCR reactions, the total reaction volume was adjusted to 25 μL, composed of 12.5 μL 2× PCR Buffer for KOD FX, 5 μL dNTPs, 0.5 μL KOD FX DNA Polymerase (1.0 U/μL), 0.75 μL PCR primer F (10 μM), 0.75 μL PCR primer R (10 μM), 4.25 μL ddH_2O, and 2 μL of a 1:20 dilution of the LR-PCR product. The reaction was carried out in a thermal cycler under the following conditions: initial denaturation at 95 °C for 5 minutes; 35 cycles of denaturation at 94 °C for 30 seconds, annealing at 60 °C for 30 seconds, and extension at 72 °C for 30 seconds. A final extension step at 72 °C for 5 minutes was conducted. Samples were then held at 10 °C for storage.

Following the reaction, 5 μL of the PCR product was loaded onto a 1.5% agarose gel in 1× TAE buffer, and electrophoresis was performed at 120 V for 30 minutes to confirm the presence of a band. If a band was observed, the PCR product was sent to a third-party provider for gel cutting, purification, and subsequent Sanger sequencing.

Protocol for MLPA

Following the instructions provided by the manufacturer (MRC Holland, Amsterdam, The Netherlands), analysis of CYP21A2 was conducted using the SALSA MLPA probe mixture P050-C1. For each reaction, 100 ng of genomic DNA was used. Quality control and data analysis were carried out using Coffalyser.net software (MRC Holland, www.mlpa.com).

Optimized Homologous Sequence Alignment (HSA) Algorithm for Accurate Variant Detection

We developed an optimized approach based on the HSA algorithm to identify genomic dislocations and differentiate between true gene mutations and pseudogene mutations by analyzing read depth across highly homologous genomic regions. This method leverages key concepts such as gene-specific nucleotides (GSNs), which are distinct nucleotide sites within homologous regions that differentiate true genes from pseudogenes.

Misalignment occurs when sequencing data aligns incorrectly to homologous genes due to principles of optimal alignment, as depicted in Figure 1. The ratio of sequencing depth between the true gene and pseudogene is quantified as the Gene-to-Pseudogene ratio (G2P), while the natural logarithm of this ratio is referred to as the Misalignment Index (MI, ln G2P).

In an ideal alignment scenario (Figure 1a), the true gene and pseudogene exhibit equal coverage (G2P = 1, MI = 0). However, misalignment can result in distinct patterns: When a variant occurs at a GSN site of the pseudogene, the pseudogene sequence may align to the true gene, resulting in false-positive duplication in the true gene or false-positive detection at the true gene site (G2P = 3, MI = 1.1) (Figure 1b). Conversely, when a variant occurs at a GSN site of the true gene, the sequence may be misaligned to the pseudogene, leading to a false-positive deletion or a false-negative single nucleotide variant (SNV) in the true gene (G2P = 0.33, MI = −1.1) (Figure 1c).

In practical scenarios, calibration of the MI is necessary to account for sequencing errors, CNVs, or gene conversions within the CYP21A2/CYP21A1P region. For example: in a sample with a normal CYP21A2 sequence, the MI is calculated as 0 (Figure 1a). A heterozygous deletion in the target gene adjusts the MI baseline to approximately −0.69 (G2P = 1/2) (Figure 1d). Duplication of the target gene shifts the MI baseline to approximately 0.41 (G2P = 3/2) (Figure 1e).

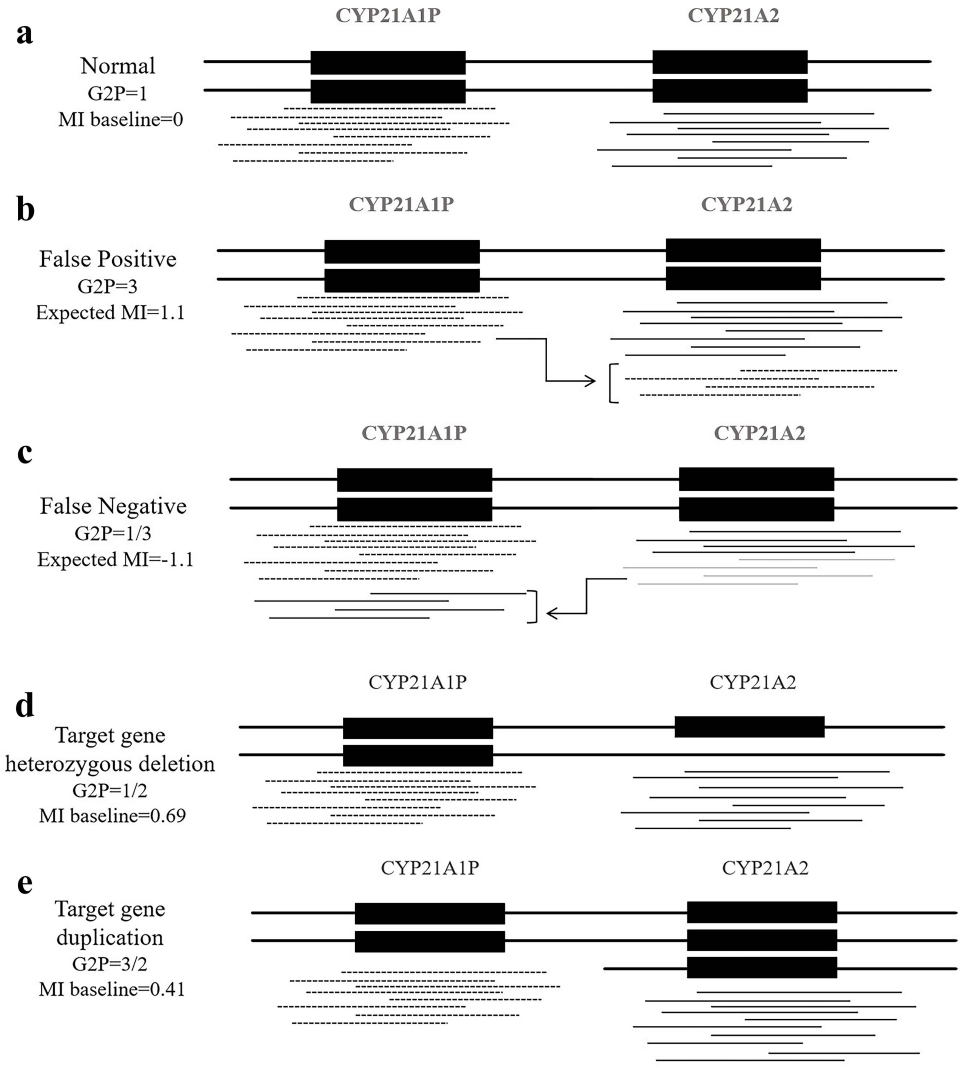


Figure 1 Impact of CNVs and GSN changes on misalignment and MI baseline adjustment. This illustration demonstrates the impact of single nucleotide differences between a gene and its pseudogene on read alignment and mutation calling due to alignment programs favoring the best-matching sequence. Key scenarios are outlined as follows: (**a**) Correct Mapping (MI = 0): In the absence of any variants, the alignment accurately maps four paired-end reads to the target gene (Orange) and its pseudogene (blue), resulting in a Mapping Integrity (MI) value of 0. (**b**) Variant in the Pseudogene GSN (False-Positive Duplication, MI = 1.1): When a heterozygous variant occurs in the GSN of the pseudogene, two paired-end reads corresponding to one allele are misaligned to the target gene. This misalignment generates a false-positive duplication, with MI = 1.1. Any variants located in-cis with the GSN variant are incorrectly mapped to the target gene, leading to lower allele fraction false-positive calls. (**c**) Variant in the Target Gene GSN (False-Positive Deletion, MI = −1.1): When a heterozygous variant occurs in the GSN of the target gene, reads are misaligned to the pseudogene. This results in a false-positive deletion with MI = −1.1. Variants located in-cis with the GSN variant are not called in the target gene, leading to false negatives. (**d**) Heterozygous Deletion of the Target Gene (MI Baseline = −0.69): A heterozygous deletion in the target gene alters the MI baseline, with a Gene-to-Pseudogene ratio (G2P) of 1/2, corresponding to MI = −0.69. This shift impacts the interpretation of read alignment and variant calling. (**e**) Duplication of the Target Gene (MI Baseline = 0.41): Duplication of the target gene changes the G2P to 3/2, resulting in an MI baseline of 0.41 (ln 1.5). This adjustment reflects the increased read depth for the target gene and influences subsequent variant detection. These scenarios highlight the key role of precise alignment and MI baseline adjustment in addressing misalignment challenges, particularly for genes with highly homologous pseudogenes, to ensure accurate mutation detection.

This approach enhances the detection accuracy of mutations and the differentiation between true and pseudogene-derived variants in highly homologous genomic regions.

The MI was calculated across the exons of the CYP21A2/CYP21A1P genes in a group of 511 individuals from a healthy population with no reported mutations in this region (Figure 2a). For exons 1–7, the calculated sequencing depth ratios for CYP21A2 and CYP21A1P were approximately 1, indicating that gene conversion events between these two genes are rare within these exons in the healthy population. However, sequencing depth ratios for exons 8–10 were significantly greater than 1, suggesting frequent conversion from CYP21A1P to CYP21A2 in these regions. As a result, MI values for exons 8–10 require specific adjustments. This observation accounts for the higher frequency of pathogenic

https://doi.org/10.2147/RMHP.S514355

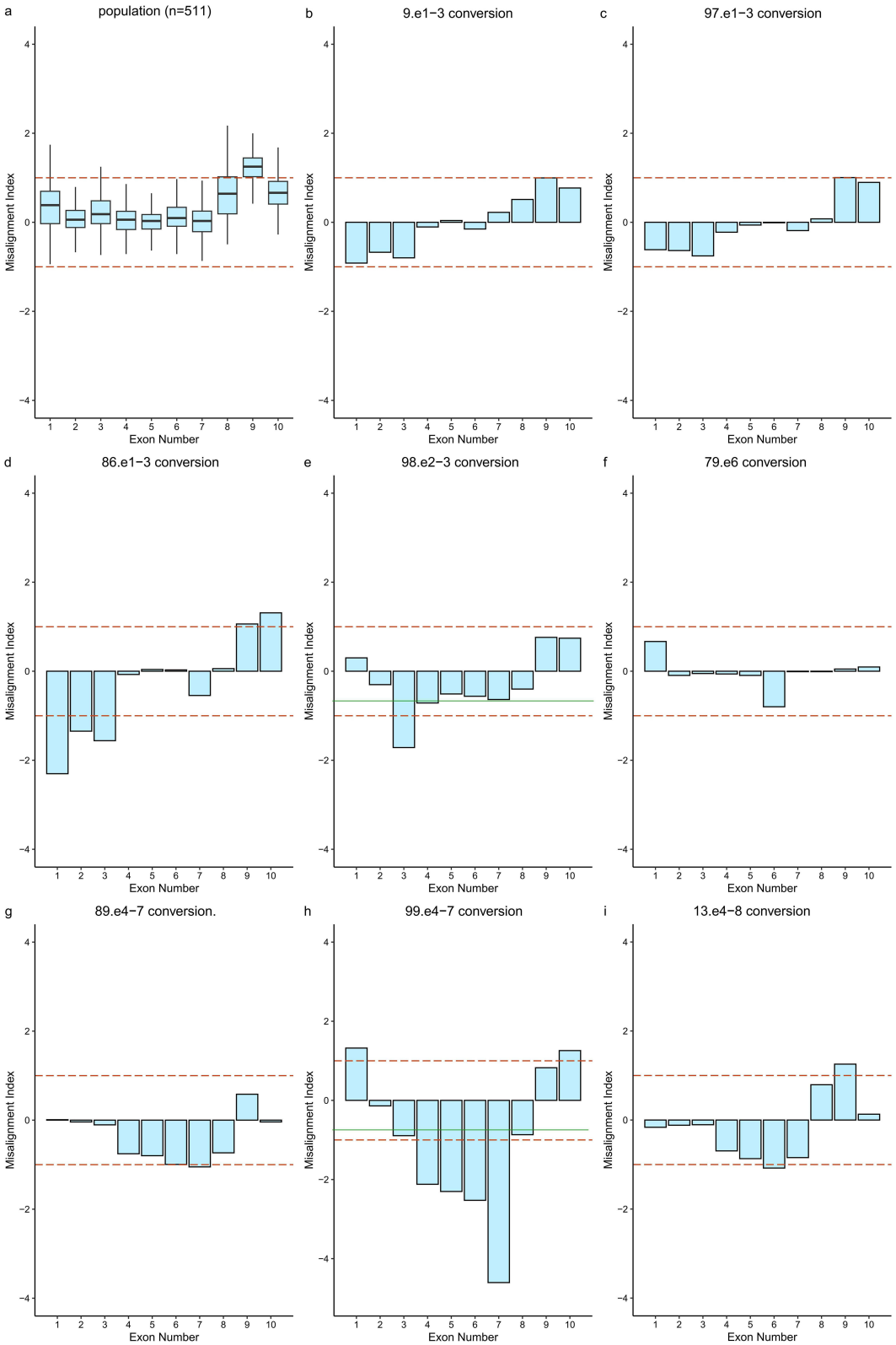


Figure 2 CYP21A2 MI analysis to identify gene conversion. The expected MI values when alleles are mapped to homologous areas are demonstrated in this analysis by the red lines close to +1 and −1. The threshold is established using the MI readings and the predicted MI baseline as criterion. (**a**) the 511 random samples' median MI. Exons 8–10 commonly map CYP21A1P readings to CYP21A2. The standard error of MI is displayed by the error bars. (**b–d**) CYP21A2 exons 1–3 were gene converted to CYP21A1P in three samples (9, 97, and 86). (**e**) Gene conversion of CYP21A2 exons 2–3 to CYP21A1P in the No.98 sample. (**f**) The changes in CNV that impact MI The No.79 sample, CYP21A2 exon 6 gene conversion to CYP21A1P. (**g** and **h**) CYP21A2 exons 4–7 are converted to CYP21A1P in the No.89 and 99 samples. (**i**) CYP21A2 exons 4–8 gene conversion to CYP21A1P, No.13 sample.

allele mutations reported in exons 8–10 of the CYP21A2 gene in population databases. For example, the mutation NM_000500.7:c.955C>T (p.Gln319*), located in exon 8, exhibits a higher frequency.

To ensure accurate variant detection, a true positive baseline for MI was defined. For exons 2–7, the baseline MI was set at 0, with a true positive MI range established between −0.5 and 0.5. Variants within this range are reliably identified from NGS data without the need for additional validation. A lower boundary for expected false negatives was set at −1.1, while an upper boundary for expected false positives was set at 1.1. Variants with MI values exceeding 0.5 indicate potential pseudogene-derived false positives, necessitating validation using MLPA or PCR. Conversely, MI values below −0.5 may indicate variants within GSNs or potential false negatives. In such cases, MLPA or PCR validation is recommended if the variants are mapped to the target gene and are clinically significant.

For exon 1, exons 8–10, or in the presence of suspected CNVs or gene conversion events, recalibration of MI baselines and cutoffs is required to maintain detection accuracy and efficiency. Custom Python scripts were developed to facilitate MI calculation and variant detection analysis.

Results

Among the 100 participants, a total of 107 PLPVs were identified using the HSA method. All variants were validated through Sanger sequencing, MLPA, and LR-PCR. Of these, 103 variants identified by the HSA method were consistent with the validation results, yielding a PPV of 96.26%. The group included 84 carriers and 16 patients (12 males and 4 females).

As summarized in Table 1, small variants accounted for 86.41% (n = 93) of all identified PLPVs, CNVs (exons deletion) for 5.83% (n = 6), and fusion mutations—representing sequence conversions between CYP21A2 and CYP21A1P—for 7.77% (n = 8).

Performance of Small Variants Identification

Table S2 provides a summary of the small variants identified in this study. Among these, nonsense mutations were the most prevalent, accounting for 66.67% of all variants, followed by missense mutations at 30.11%, frameshift mutations at 2.15%, and splicing mutations at 1.08%. Notably, 62 nonsense mutations were identified, all located in exon 8. Among these, 98.39% (n = 61) were c.955C > T (p.Gln319*), with the remaining mutation being c.949C > T (p.Arg317*). Of the 19 missense mutations, 68.42% occurred in exons 7 and 8, with c.518T > A (p.Ile173Asn) and c.844G > T (p.Val282Leu) each accounting for 42.11%. Additionally, 10 mutations were detected in intron 2, including c.293–13C > G (p.?) and c.292+1G > A (p.?).

In all, 83 loci were analyzed using the HSA method, excluding c.293–13C > G (p.?) and c.292+1G > A (p.?). Table 2 presents the concordance between the predicted results from HSA and the PCR/Sanger validation results for individual SNVs. The overall concordance rate between HSA predictions and validation results was 95.24%. As shown in Table S2, 51 SNVs exhibited MI values at or above the upper threshold, indicating false positives, while 5 SNVs had MI values below the lower threshold, indicating false negatives. PCR validation confirmed that these false positives and false negatives identified by HSA accurately reflected the true mutation status in the samples. Additionally, 23 SNVs predicted as positive by HSA were all confirmed as true positive variants through PCR.

Specifically, for the variant NM_000500.7: c.955C > T (p.Gln319*) in exon 8 of the CYP21A2 gene, 61 SNVs were detected. Discrepancies were observed between the MI threshold determination and PCR validation results for 4 SNVs.

Table 1 Mutation Spectrum of 100 Samples Containing Detected PLPVs

Type of Samples	Number of Small Variants (SNV/Indels)				Number of Exons Deletion	Number of Fusion Mutations
	Nonsense	Missense	Frameshift	Splicing		
Carrier	61	18	0	0	2	3
Proband	1	10	2	1	4	5
In total	62	28	2	1	6	8

Table 2 Characteristics and Predictive Capabilities of Minor Variations Identified in the Study

Specific Mutation	Mutation Type	Mutation Site	Number of Carriered Variants	Number of Patient's Variants	Total Number Variants	MI	MI	MI	Number of PCR Validated Mutation	Concordance
						FP	FN	TP		
c.955C>T (p.Gln319*)	Nonsense	Exon8	60	1	61	54	0	7	57	93.44%
c.518T>A (p.Ile173Asn)	Missense	Exon4	4	4	8	0	4	4	8	100%
c.844G>T (p.Val282Leu)	Missense	Exon7	6	0	6	0	0	6	6	100%
c.913G>A (p.Val305Met)	Missense	Exon7	2	0	2	0	0	2	2	100%
c.92C>T (p.Pro31Leu)	Missense	Exon1	1	0	1	0	0	1	1	100%
c.949C>T (p.Arg317*)	Nonsense	Exon8	1	0	1	0	0	1	1	100%
c.1360C>T (p.Pro454Ser)	Missense	Exon10	1	0	1	0	0	1	1	100%
c.1069C>T (p.Arg357Trp)	Missense	Exon8	0	1	1	0	1	0	1	100%
c.1451_1452delinsC (p.Arg484Profs*58)	Frameshift	Exon10	0	1	1	0	0	1	1	100%
c.292+1G>A (p.?)	Splicing	Intron 2	0	1	1	0	0	1	1	100%
c.923dup (p.Leu308Phefs*6)	Frameshift	Exon7	0	1	1	0	0	1	1	100%

Note: The symbol (*) represent stop codon.

Of these, three were identified as false positives by HSA, while one was confirmed as a true positive variant. The concordance rate for this specific variant was 93.44%.

Performance of CNVs Identification

A total of 6 CNVs were identified, all involving exon deletions in the CYP21A2 gene. Of these, 5 CNVs exhibited deletions spanning all 10 exons of CYP21A2, while one CNV involved the deletion of exons 1–6. Table 3 provides a summary of the concordance between HSA-predicted results and MLPA validation for each CNV. The HSA analysis revealed that all 6 CNVs were characterized by the loss of multiple exons (Table S3). Subsequent MLPA validation confirmed the authenticity of these CNVs, demonstrating complete concordance with the results of the HSA analysis (Figure S1a–f).

The presence of CNVs significantly influenced the MI baseline of the CYP21A2 gene. In cases of a single-copy deletion, the MI baseline was adjusted to approximately −0.69 (ln(1/2)). For a double-copy deletion, where more than half of the CYP21A2 reads were assumed lost, the MI baseline dropped further to −1.38 (ln(0.5/2)).

For instance: Sample 46: MI values for exons 1–6 were close to −0.69, suggesting a single-copy deletion, while the MI value for exon 7 remained normal, indicating a heterozygous deletion spanning exons 1–6. MLPA analysis showed a probe ratio of 0.5 for exons 1–6 and 1 for exon 7, confirming a single-copy deletion in CYP21A2 exons 1–6 (Figure S1a). Sample 92: MI values for exons 1–7 were below −1.38, suggesting a potential homozygous deletion. MLPA revealed a probe ratio of 1 for the CYP21A1P pseudogene and 0 for the CYP21A2 gene, confirming a homozygous deletion consistent with the HSA result (Figure S1b). Sample 30: MI values for all exons were close to 0 (ln(1) = 0), and the overall CYP21A2 and CYP21A1P copy number was approximately 2. MLPA revealed probe ratios of 0.5 for both CYP21A1P and CYP21A2, indicating a single-copy deletion in each, which adjusted the MI baseline to 0 (ln(1/1)) (Figure S1c).

Prediction and Validation of Fusion Mutation in CYP21A2 and CYP21A1P

By using the HSA method, eight conversion mutations between CYP21A2 and CYP21A1P were detected, as detailed in Table S4, and calculated the sequencing depth ratio of CYP21A2/CYP21A1P for each exon (Figure 2b–i). Among these, five distinct types of CYP21A2/CYP21A1P fusion variants were identified, involving exon regions 1–3, 2–3, 6, 4–7, and 4–8 (Table 4), Subsequent MLPA or Sanger sequencing validation experiments confirmed the occurrence of the fusion events (Figure 3a–f, Figure S1e–i).

For three loci (Nos. 9, 97, and 86), HSA analysis revealed abnormal sequencing depth ratios for exons 1–3 and the presence of false-negative sites, suggesting potential conversion from CYP21A2 to CYP21A1P (Figure 2b–d). MLPA analysis confirmed this finding, showing a probe ratio for CYP21A1P exons 1 and 3 of ≥ 1.5 and a corresponding ratio for CYP21A2 of ≤ 0.5 (Figure 3a, Figure S1g and h). This result confirmed the occurrence of a true-pseudogene fusion event (Figure 3b).

For the fusion variant at exon 6, HSA analysis detected an abnormal ratio and a false-negative site (Figure 2f). LR-PCR sequencing identified heterozygous mutations at four GSN sites in exon 6 (c.[705T > A; 710T > A; 713T > A; 719T > A]), indicating that the exon 6 region of CYP21A2 had been converted into the homologous sequence of CYP21A1P (Figure 3c and d).

At the rare fusion site involving exons 4–8, HSA analysis revealed an MI value close to −1 (Figure 2i). Misalignment of reads from the 3' end of CYP21A1P (exons 8–10) to CYP21A2 frequently canceled MI changes, leading to false-positive variants. Variant annotation via NGS proved challenging. MLPA results showed probe ratios of ≥ 1.5 for CYP21A1P exons 4 and 7, while the ratio for CYP21A2 was approximately 0.5. LR-PCR sequencing further identified six GSN sites in CYP21A2

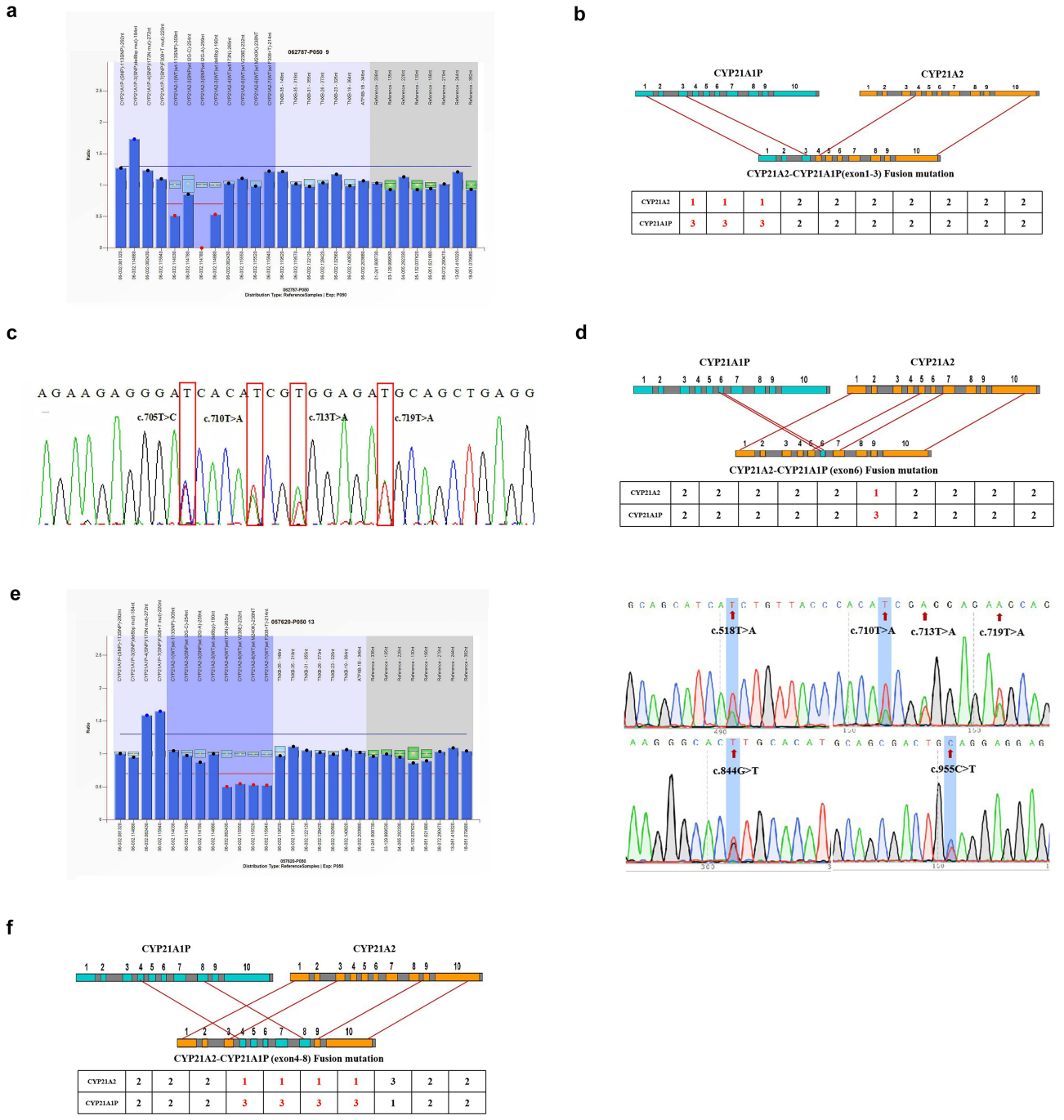
Table 3 CNV Profile and Prediction Performance in the Study

Specific Mutation	Mutation Type	Mutation Site	Number of Carriered Variants	Number of Patient's Variants	Total Number of Variants	MI Indicated Deletion Variants	MLPA Concordance
ex.1_10del (p.?)	Exon Deletion	Exon1-10	1	4	5	5	100%
ex.1_6del (p.?)	Exon Deletion	Exon1-6	1	0	1	1	100%

Table 4 Gene Conversion Profile and Prediction Performance Between CYP21A2 and CYP21A1P

Specific Mutation	Mutation Site	Number of Carriered Variants	Number of Patient's Variants	Total Number Variants	MI	MI Indicated Deletion Variants	MLPA Concordance
					FN		
c.[−113G>A; c.293–13C/A>G; c.332del8](p.?)	5'UTR-exon3	1	2	3	3	0	100%
c.[518T>A; 710T>A; 844G>T](p.?)	Exon4-7	0	2	2	2	1	100%
c.[705T>A;710T>A;713T>A;719T>A] (p.[Asp235Glu;Ile237Asn;Val238Glu;Met240Lys])	Exon6	1	0	1	1	0	100%
c.[518>A;710T>A;713T>A;719T>A;844G>T;923dup;955C>T](p.?)	Exon4-8	1	0	1	1	0	100%
c.[293–13C>G; 332_339del](p.?)	Exon2-3	0	1	1	1	1	100%

https://doi.org/10.2147/RMHP.S514355



CYP21A2	1	1	1	2	2	2	2	2	2	2
CYP21A1P	3	3	3	2	2	2	2	2	2	2

CYP21A2	2	2	2	2	2	1	2	2	2	2
CYP21A1P	2	2	2	2	2	3	2	2	2	2

CYP21A2	2	2	2	1	1	1	1	3	2	2
CYP21A1P	2	2	2	3	3	3	3	1	2	2

Figure 3 Schematic depiction of the CYP21A1P/CYP21A2 fusion mutation with MLPA and LR-PCR results for the CYP21A2 gene. (**a**) gene conversion event between these two genes at the exon 1–3 segment is indicated by the heterozygous deletion of exons 1–3 in the CYP21A2 gene and the equivalent duplication in the CYP21A1P gene. (**b**) Depiction of the copy number ratio and fusion mutation for CYP21A1P/CYP21A2 exons 1–3. (**c**) In exon 6 of the CYP21A2 gene, LR-PCR analysis identified heterozygous mutations at four GSN sites: c. [705T > C; 710T > A; 713T > A; 719T > A], indicative of a gene conversion event. (**d**) Depiction of the copy number ratio and fusion mutation for CYP21A1P/CYP21A2 exons 6. (**e**) A gene conversion event between the CYP21A2 and CYP21A1P genes at the exon 4–8 segment is shown by the heterozygous deletion of exons 4–8 in the CYP21A2 gene and the equivalent duplication in the CYP21A1P gene. Six GSN sites (c.518 > A;710T > A;713T > A;719T > A;844G > T;955C > (T) inside exons 4–8 were found to have heterozygous mutations by LR-PCR detection, which further supported the gene conversion occurrence in the exon 4–8 region. (**f**) Depiction of the copy number ratio and fusion mutation for CYP21A1P/CYP21A2 exons 4–8.

exons 4–8 (c.518T > A; 710T > A; 713T > A; 719T > A; 844G > T; 955C > T) that had been converted to CYP21A1P, with a heterozygous mutation detected at c.955C > T (p.Gln319) in exon 8 (Figure 3e). These findings indicated a conversion of the CYP21A2 exon 4–8 segment to the homologous sequence of CYP21A1P (Figure 3f).

https://doi.org/10.2147/RMHP.S514355

HSA analysis of additional sites identified one fusion variant in exons 2–3 (Figure 2e, No. 98) and two in exons 4–7 (Figure 2g and h, Nos. 89 and 99). These sites displayed abnormal sequencing depth ratios, indicating false-negative sites. LR-PCR and MLPA validation confirmed the presence of CYP21A2/CYP21A1P fusion genes at these loci, consistent with the HSA analysis results (Figure S1e,f,i and Figure S2a-b).

For fusion variants involving heterozygous deletions, the MI baseline is adjusted to approximately −0.69 (ln(1/2)), and the expected high MI value is reduced to 0.69 (ln(2/1)). If half of the CYP21A2 reads are mis mapped to CYP21A1P, the expected low MI value drops further to below −1.6 (ln(0.5/2.5)).

The HSA analysis in Figure 2e and h highlights the deletion of exons 1–10 and indicates the presence of false-negative sites. Consequently, the MI baseline for these two loci was recalibrated to −0.69. MLPA results, as revealed in Figure S1e and Figure S1f, confirmed the CNV deletions at these loci, aligning with the findings from the HSA analysis. These results provide a robust basis for accurate diagnosis.

Discussion

Gene duplication represents a fundamental evolutionary mechanism that facilitates the development of novel gene functions while preserving the original genes and their associated biological roles. This process often results in the formation of pseudogenes and highly homologous genes.[25,26] However, conventional bioinformatics approaches for sequence read analysis frequently encounter difficulties in distinguishing genomic regions with high sequence similarity, such as gene families and pseudogenes. Although NGS has significantly enhanced the resolution and accuracy of genomic analyses, challenges related to misalignment persist in diagnostic contexts.

To address these issues, various alignment algorithms have been developed over time, beginning with the Smith-Waterman algorithm and evolving to include tools such as the Burrows-Wheeler Aligner, Bowtie, Novoalign, and SOAP.[27–30] These algorithms aim to map sequencing reads to their respective genomic regions with minimal error. Despite these advances, regions with extensive sequence similarity resulting from gene duplications continue to pose challenges, including undetected mutations, erroneous variant calls, and inaccuracies in CNV assessments caused by misaligned reads. Nucleotide variations specific to homologous regions further exacerbate misalignment, undermining the precision of genetic diagnoses.

The high sequence homology between CYP21A1P and CYP21A2 genes presents a significant obstacle in the diagnosis and study of 21-hydroxylase deficiency syndrome, complicating the differentiation of mutation spectra and underlying genetic mechanisms.[15,31] To address this challenge, a novel data analysis method, referred to as the HSA algorithm, was developed. This method uses read-depth comparisons across homologous regions to identify and mitigate misalignment issues.

The findings indicate that integrating high-throughput sequencing with the HSA algorithm enables accurate mutation detection in the CYP21A2 gene. This approach effectively distinguishes pseudogene-derived matches and identifies common SNVs and CNVs. To ensure reliable results, validation using LR-PCR or MLPA is recommended. This combined strategy provides precise genetic mutation information, supporting the clinical management and treatment of 21-hydroxylase deficiency.

In this study, 100 samples were analyzed, resulting in the identification of 107 variants. Of these, 104 were validated using Sanger sequencing, MLPA, and LR-PCR following HSA analysis, yielding a PPV of 97.19% for variants detected in CYP21A2. The results demonstrated concordance between HSA analysis and validation methods for 93 variants, comprising of 83 SNVs/indels, 6 CNVs, and 8 gene conversion events between CYP21A2 and CYP21A1P.

Among the SNVs, 23 were confirmed as true positives, while 51 were identified as false positives and 5 as false negatives. Some false positives, such as the c.955C > T mutation in CYP21A2 exon 8, likely resulted from misalignment of sequencing reads to the 3' end of CYP21A1P. Additionally, 10 variants, including c.518T > A, c.844G > T, and c.913G > A, were initially misclassified as pathogenic but were ultimately attributed to mis mapping of reads to pseudogenes. Notably, pathogenic variants such as c.518T > A and c.1069C > T were undetected in standard NGS analysis without the integration of HSA, likely due to mutations in CYP21A2 causing reads to be misaligned with CYP21A1P, resulting in false negatives.

The combination of HSA analysis with orthogonal validation methods, such as LR-PCR or MLPA, proved effective in identifying mis mapping events, thereby improving the accuracy of variant detection and enhancing the reliability of clinical diagnoses. Additionally, incorrect read mapping contributed to erroneous CNV calls. The HSA algorithm estimated the copy number of CYP21A2 and its pseudogene, adjusted the MI threshold accordingly, and identified six

CNV deletions and eight unique CYP21A1P/CYP21A2 gene conversion events. Corresponding changes in MI values were observed and validated through MLPA and LR-PCR, consistent with the HSA analysis.

These results highlight the limitations of relying solely on NGS with standard variant calling protocols, which may overlook certain mutations and hinder the establishment of causal relationships.

CAH is a genetic disorder associated with symptoms such as abnormal sexual development, growth retardation, hypertension, and diabetes. Although mass spectrometry offers a rapid means of measuring 17-hydroxyprogesterone levels to help in CAH diagnosis, its limited specificity and sensitivity can result in misdiagnoses or missed diagnoses, particularly in milder cases.[31] Studies have identified over 200 pathogenic variants associated with CAH.[32,33]

This study used NGS in combination with the HSA algorithm to effectively detect 20 known and rare fusion variants of the CYP21A2 gene, significantly enhancing the accuracy and sensitivity of variant identification. This approach reduces the time and cost associated with detection while enabling the precise identification of reportable mutations that require further validation. Comprehensive genotyping is essential for accurate CAH diagnosis, evaluation of disease severity, and differentiation from other conditions with overlapping symptoms, such as polycystic ovary syndrome.

It is important to note that false-positive results have been observed for certain variants, such as c.955C > T, underscoring the need for robust validation methods. This study focuses on 21-hydroxylase deficiency caused by mutations in the CYP21A2 gene and highlights the limitations of traditional diagnostic methods, including their potential for generating misleading results. Strategies for improving genetic diagnostics are examined, including the implementation of an HSA-based workflow designed to address mis mapping issues and provide guidance for conducting orthogonal genetic testing. The core design principles of the HSA method endow it with the theoretical potential to be applied to other complex genomic regions, such as HBA1/HBA2 and SMN1/SMN2. However, for each specific gene, a detailed analysis of the homologous sequence characteristics and identification of gene-specific nucleotide (GSN) sites are required. Additionally, it is necessary to establish a baseline and more accurate analytical thresholds based on second-generation sequencing data. Only after these steps can a preliminary analysis algorithm be formed for a specific homologous sequence region. Subsequently, the results of the HSA technology must be compared with those of gold-standard experimental methods, such as MLPA, to confirm its accuracy in clinical applications.

Different highly homologous regions, such as HBA1/HBA2 and SMN1/SMN2, possess unique sequence features, degrees of homology, types of variations, and potential interfering factors. Therefore, when directly applying the HSA method to these regions, it may be necessary to recalibrate and optimize the baselines and thresholds according to the specific data of the target region. In addition to contributing to diagnostic inaccuracies, mis mapping may interfere with early-stage research by complicating the establishment of accurate gene-disease associations.

Limitations

Despite its numerous advantages, NGS-based analysis of CYP21A2 presents certain challenges. The accuracy of detecting rare pathogenic loci that have not yet been extensively studied remains uncertain. Additionally, the current analysis still heavily relies on the expertise and judgment of curators. The limitations of this study should be acknowledged. Firstly, the study only compared the consistency between our method and the gold-standard experimental methods for positive samples, thus validating the feasibility of identifying SNVs, indels, CNVs, and fusions in the CYP21A2 gene. However, this approach does not provide a comprehensive evaluation of the algorithm's performance across different populations or its compatibility with multiple NGS platforms. Future research should involve large-scale screening and analysis to demonstrate a more comprehensive analytical performance. Secondly, the study did not systematically validate sites judged by the algorithm as "no need for validation" (negative prediction), which means the number of false negatives (FN) and true negatives (TN) remains unknown, precluding the calculation of sensitivity, specificity, and NPV. Lastly, while the HSA method shows promise for application to other highly homologous genomic regions, its direct application to these areas may require recalibration and optimization of baselines and thresholds according to the specific data of the target region. Future research should prioritize the development of standardized protocols and automated analysis tools to reduce reliance on manual curation. Expanding reference databases and integrating functional assays will further enhance the clinical use of NGS in CAH diagnosis and management.

Conclusion

The accurate identification of pathogenic mutations in patients remains a fundamental objective of genetic testing, and the incorporation of misalignment checks is essential for enhancing diagnostic accuracy. For carriers without evident clinical manifestations, precise genetic testing plays a key role in reducing the likelihood of disease transmission to offspring. Misalignment during sequencing is a well-documented challenge, particularly for certain genes.

Based on the findings of this study, the proposed technology has demonstrated its efficacy in identifying misalignment issues, thereby significantly reducing misdiagnosis rates. By providing a simple and cost-effective method for detecting misalignments through the comparison of read depths across highly homologous regions, this approach offers a reliable solution. Furthermore, this method can be generalized and applied to mutation detection in other genes with similar challenges related to highly homologous sequences.

Abbreviation

CAH, Congenital adrenal hyperplasia; 21-OHD, 21-hydroxylase deficiency; CYP21, cytochrome P450c21; 17-OHP, 17-hydroxyprogesterone; GC-MS, gas chromatography-mass spectrometry; LC-MS/MS, liquid chromatography-tandem mass spectrometry; NGS, next-generation sequencing; CNV, Copy number variations; GATK4, Genomic Analysis Toolkit; SNVs, Single nucleotide variants; ACMG, American College of Medical Genetics and Genomics; AMP, Association for Molecular Pathology; LR-PCR, long range polymerase chain reaction; MLPA, multiplex ligation-dependent probe amplification; HSA, homologous sequence alignment; GSN, Gene-specific nucleotide; G2P, Gene-to-Pseudogene ratio.

Data Sharing Statement

The datasets generated and analysed during the current study are available in the dbSNP repository (https://www.ncbi.nlm.nih.gov/SNP/snp_viewTable.cgi?handle=NB_FRCYP).

Ethics Approval and Consent to Participate

The study was conducted in accordance with the Declaration of Helsinki (as was revised in 2013). The study was approved by Ethics Committee of the Women and Children's Hospital of Ningbo University (2021-ky-036). Written informed consent was obtained from all participants.

Acknowledgments

The authors appreciate Xiaoxia Liu to provide suggestion and guidance on the curation of the mutations. The authors thank Xinyuan Tong and Hui Tian, for the help in carrying out the experiments. The authors also extend their gratitude to Bailei Zhang for the support provided through the Ningbo Key Support Medical Discipline (2022-B16).

Funding

Medical and Health Project of Zhejiang (2022KY1153, Yibo Chen), Key Technology Breakthrough Program of "Ningbo Sci-Tech Innovation YONGJIANG 2035" (2024Z221, Haibo Li), Municipal Public Welfare Project (2022S035, Haibo Li), Ningbo Key Support Medical Discipline (2022-B16, Bailei Zhang).

Disclosure

The authors declare that they have no competing interests in this work.

References

1. Subspecialty Group of Endocrinologic. Hereditary and metabolic diseases, the Society of Pediatrics, Chinese Medical Association. Consensus statement on diagnosis and treatment of congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Zhonghua Er Ke Za Zhi*. 2016;54(8):569–576. doi:10.3760/cma.j.issn.0578-1310.2016.08.003
2. El-Maouche D, Arlt W, Merke DP. Congenital adrenal hyperplasia [published correction appears in Lancet. 2017;390(10108):2142. doi: 10.1016/S0140-6736(17)32818-0]. *Lancet*. 2017;390(10108):2194–2210. doi:10.1016/S0140-6736(17)31431-9
3. Li Q, Zhou Y, Xu YH, et al. Analysis on neonatal screening for inherited metabolic diseases in Zhejiang Province from 1999 to 2018. *Prev Med*. 2019;31:1081–1085.

4. Subspecialty Group of Newborn Screening; Society of Birth Defects Prevention and Control, Chinese Preventive Medicine Association; Subspecialty Group of Clinical Genetics, Society of Adolescent Medicine, Chinese Medical Doctor Association; Subspecialty Group of Endocrinologic, Hereditary and Metabolic Diseases, The Society of Pediatrics, Chinese Medical Association. Consensus statement on neonatal screening for congenital adrenal hyperplasia. *Zhonghua Er Ke Za Zhi*. 2016;54(6):404–409. doi:10.3760/cma.j.issn.0578-1310.2016.06.003
5. Haider S, Islam B, D'Atri V, et al. Structure-phenotype correlations of human CYP21A2 mutations in congenital adrenal hyperplasia. *Proc Natl Acad Sci U S A*. 2013;110(7):2605–2610. doi:10.1073/pnas.1221133110
6. Korula S, Chapla A, Ravichandran L, et al. Comprehensive overview of congenital adrenal hyperplasia and its genetic diagnosis among children and adolescents. *J Pediatric Endocrinol Diabetes*. 2022;2(3). doi:10.25259/JPED_4_2023
7. Pan P, Yang DZ. Interpretation of the new guidelines on congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *J Pract Obstet Gynecol*. 2020;36:818–821.
8. Han LS. Current status of diagnosis and treatment of congenital adrenal hyperplasia. *Chin J Pediatr*. 2016;31:410–413.
9. Odenwald B, Nennstiel-Ratzel U, Dörr HG, Schmidt H, Wildner M, Bonfig W. Children with classic congenital adrenal hyperplasia experience salt loss and hypoglycemia: evaluation of adrenal crises during the first 6 years of life. *Eur J Endocrinol*. 2016;174(2):177–186. doi:10.1530/EJE-15-0775
10. Mass Screening Committee; Japanese Society for Pediatric Endocrinology; Japanese Society for Mass Screening, et al. Guidelines for diagnosis and treatment of 21-hydroxylase deficiency (2014 revision). *Clin Pediatr Endocrinol*. 2015;24(3):77–105. doi:10.1297/cpe.24.77
11. Kamrath C, Hartmann MF, Boettcher C, Zimmer KP, Wudy SA. Diagnosis of 21-hydroxylase deficiency by urinary metabolite ratios using gas chromatography-mass spectrometry analysis: reference values for neonates and infants. *J Steroid Biochem Mol Biol*. 2016;156:10–16. doi:10.1016/j.jsbmb.2015.10.013
12. Hayashi GY, Carvalho DF, de Miranda MC, et al. Neonatal 17-hydroxyprogesterone levels adjusted according to age at sample collection and birthweight improve the efficacy of congenital adrenal hyperplasia newborn screening. *Clin Endocrinol*. 2017;86(4):480–487. doi:10.1111/cen.13292
13. Fingerhut R. False positive rate in newborn screening for congenital adrenal hyperplasia (CAH)-ether extraction reveals two distinct reasons for elevated 17alpha-hydroxyprogesterone (17-OHP) values. *Steroids*. 2009;74(8):662–665. doi:10.1016/j.steroids.2009.02.008
14. White PC. Neonatal screening for congenital adrenal hyperplasia. *Nat Rev Endocrinol*. 2009;5(9):490–498. doi:10.1038/nrendo.2009.148
15. Eggermann T, Elbracht M, Kurth I, et al. Genetic testing in inherited endocrine disorders: joint position paper of the European reference network on rare endocrine conditions (Endo-ERN). *Orphanet J Rare Dis*. 2020;15(1):144. doi:10.1186/s13023-020-01420-w
16. Chen W, Xu Z, Sullivan A, et al. Junction site analysis of chimeric CYP21A1P/CYP21A2 genes in 21-hydroxylase deficiency. *Clin Chem*. 2012;58(2):421–430. doi:10.1373/clinchem.2011.174037
17. Narasimhan ML, Khattab A. Genetics of congenital adrenal hyperplasia and genotype-phenotype correlation. *Fertil Steril*. 2019;111(1):24–29. doi:10.1016/j.fertnstert.2018.11.007
18. Mandelker D, Schmidt RJ, Ankala A, et al. Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing. *Genet Med*. 2016;18(12):1282–1289. doi:10.1038/gim.2016.58
19. Zhang H, Jain C, Aluru S. A comprehensive evaluation of long read error correction methods. *BMC Genomics*. 2020;21(Suppl 6):889. doi:10.1186/s12864-020-07227-0
20. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol*. 2020;21(1):30. doi:10.1186/s13059-020-1935-5
21. Behera S, Catreux S, Rossi M, et al. Comprehensive and accurate genome analysis at scale using DRAGEN accelerated algorithms. Preprint. *bioRxiv*. 2024:2024.01.02.573821. doi:10.1101/2024.01.02.573821
22. Strom SP, Hossain WA, Grigorian M, et al. A streamlined approach to Prader-Willi and Angelman syndrome molecular diagnostics. *Front Genet*. 2021;12:608889. doi:10.3389/fgene.2021.608889
23. Chen SC, Zhou XY, Li SY, et al. Carrier burden of over 300 diseases in Han Chinese identified by expanded carrier testing of 300 couples using assisted reproductive technology. *J Assist Reprod Genet*. 2023;40(9):2157–2173. doi:10.1007/s10815-023-02876-y
24. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405–424. doi:10.1038/gim.2015.30
25. Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet*. 2010;11(2):97–108. doi:10.1038/nrg2689
26. Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R. Gene duplication as a major force in evolution. *J Genet*. 2013;92(1):155–161. doi:10.1007/s12041-013-0212-8
27. Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol*. 1981;147(1):195–197. doi:10.1016/0022-2836(81)90087-5
28. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10(3):R25. doi:10.1186/gb-2009-10-3-r25
29. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760. doi:10.1093/bioinformatics/btp324
30. Hasan MS, Wu X, Zhang L. Uncovering missed indels by leveraging unmapped reads. *Sci Rep*. 2019;9(1):11093. doi:10.1038/s41598-019-47405-z
31. Baumgartner-Parzer S, Witsch-Baumgartner M, Hoeppner W. EMQN best practice guidelines for molecular genetic testing and reporting of 21-hydroxylase deficiency. *Eur J Hum Genet*. 2020;28(10):1341–1367. doi:10.1038/s41431-020-0653-5
32. New MI, Abraham M, Gonzalez B, et al. Genotype-phenotype correlation in 1,507 families with congenital adrenal hyperplasia owing to 21-hydroxylase deficiency. *Proc Natl Acad Sci U S A*. 2013;110(7):2611–2616. doi:10.1073/pnas.1300057110
33. Krone N, Braun A, Roscher AA, Knorr D, Schwarz HP. Predicting phenotype in steroid 21-hydroxylase deficiency? Comprehensive genotyping in 155 unrelated, well defined patients from southern Germany. *J Clin Endocrinol Metab*. 2000;85(3):1059–1065. doi:10.1210/jcem.85.3.6441

Risk Management and Healthcare Policy

Dovepress
Taylor & Francis Group

Publish your work in this journal

Risk Management and Healthcare Policy is an international, peer-reviewed, open access journal focusing on all aspects of public health, policy, and preventative measures to promote good health and improve morbidity and mortality in the population. The journal welcomes submitted papers covering original research, basic science, clinical & epidemiological studies, reviews and evaluations, guidelines, expert opinion and commentary, case reports and extended reports. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit http://www.dovepress.com/testimonials.php to read real quotes from published authors.

Submit your manuscript here: https://www.dovepress.com/risk-management-and-healthcare-policy-journal