

RESEARCH

Open Access



Genetic characterization and screening of congenital adrenal hyperplasia by long-read sequencing in a cohort of 21,239 newborns

Desheng Liang^{1*†}, Min Zhu^{2†}, Qiaowei Liang^{3†}, Rong Qiang^{4†}, Lei Yu^{5†}, Shiyi Xu^{6†}, Menglin Li¹, Jieping Song⁷, Yulin Zhou⁸, Xiaoyan He⁹, Yonglan Huang¹⁰, Hua Jin¹¹, Jianqiang Tan¹², Hui Liu⁶, Aihua Xia¹³, Yingdi Liu^{1,3}, Peisen Liu¹, Zhuo Li¹, Ruifang Wang¹⁴, Dongjuan Wang⁹, Ruixue Zhang⁴, Qian Pu⁵, Jinfu Zhou^{15,16}, Runhong Xu⁷, Xudong Wang¹⁷, Minyi Tan¹⁰, Dayu Chen¹², Chaoyan Wu¹³, Di Cui¹⁸, Aiping Mao^{18*}, Wenhao Zhou^{19*}, Wenjuan Qiu^{14*} and Lingqian Wu^{1,3*}

Abstract

Background Comprehensive genetic characterization and screening for congenital adrenal hyperplasia (CAH) have not yet been achieved at the population level because of the complexity of the *CYP21A2* locus. This prospective study incorporated long-read sequencing (LRS) into the current first-tier biochemical newborn screening (NBS) to comprehensively characterize the variant spectrum of *CYP21A2*, fully investigate the carrier frequency and expected incidence of classic and non-classic CAH (NCCA), and evaluate the clinical feasibility of genetic NBS for CAH.

Methods A total of 21,239 newborns were consecutively recruited from 11 centers across China between June 2023 and May 2024. All the participants underwent biochemical and genetic NBS. In vitro enzymatic activity and minigene assays were performed to determine the pathogenicity of novel variants. A 30.8-kb long amplicon, followed by LRS, was performed to determine the phasing of duplication chimera and single-nucleotide variations (SNVs) and indels in *CYP21A2*.

[†]Desheng Liang, Min Zhu, Qiaowei Liang, Rong Qiang, Lei Yu and Shiyi Xu contributed equally to this work.

*Correspondence:

Desheng Liang
liangdesheng@sklmg.edu.cn
Aiping Mao
maoaping@berrygenomics.com
Wenhao Zhou
zhouwenhao@fudan.edu.cn
Wenjuan Qiu
qiuwenjuan@xinhumed.com.cn
Lingqian Wu
wulingqian@sklmg.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Results Eligible genetic screening results were obtained for 21,234 (99.98%) newborns. The allele frequencies of duplications and deletions at the *CYP21A2* locus were 4.51% and 0.15%, respectively. In vitro functional analysis and LRS-based phasing were performed to precisely determine carrier alleles, setting an overall frequency of 1.67% (711/42468, 95% confidence interval (CI): 1.55–1.80%), with 0.75% (320/42468, 95% CI: 0.67–0.84%) and 0.92% (391/42468, 95% CI: 0.83–1.01%) for classic and NC carriers, respectively. Notably, hotspot variants including SNVs/indels caused by microgene conversion and 30-kb deletions caused by unequal crossover accounted for 84.0% (597/711), whereas rare variants comprised as high as 16.0% (114/711) of all variants. The expected incidence of classic CAH according to allele frequency was 1/17613. The expected incidence of NCCAH in Chinese population (1/4474) was significantly lower than that in US Ashkenazi Jews (1/133) and Caucasians (1/337), mainly owing to the different allele frequencies of the NC variant *CYP21A2*:c.844G > T. Biochemical NBS identified 106 (0.50%) positive samples with a positive predictive value of 0.94% (1/106). LRS accurately identified the one case of classic CAH, with no false positives.

Conclusions Our findings provide a population-level carrier frequency and incidence estimates with a comprehensive landscape of the *CYP21A2* locus, and demonstrate the effectiveness of first-tier LRS-based genetic NBS for CAH.

Keywords Congenital adrenal hyperplasia, Long-read sequencing, Newborn screening, *CYP21A2*, Carrier frequency, Disease incidence

Background

Congenital adrenal hyperplasia (CAH) is potentially life-threatening in its severe classic forms including salt-wasting (SW) and simple-virilizing (SV), and can be asymptomatic or late-onset in its mild non-classic (NC) form [1–4]. More than 95% of CAH cases are caused by 21-hydroxylase deficiency (21-OHD) with biallelic variants in *CYP21A2* [5]. The continuous phenotypic spectrum of CAH is generally correlated with the milder defected allele of *CYP21A2* and the expected residual 21-hydroxylase activity [6].

The epidemiology of the classic form of CAH is well established to be between 1/10000 and 1/20000 live births, owing to the screening of millions of newborns worldwide [7]. NCCAH is more common with an estimated incidence of 1/200 to 1/2000 [1, 3, 7, 8]. Notably, the disease rate of NCCAH in US Caucasian population has been estimated to be as high as 1/200, arguing for screening in the setting of female infertility regardless of ethnicity [8]. However, the incidence of NCCAH has only been estimated in small cohorts, limited to certain ethnicities and with techniques unable to comprehensively characterize all the disease-relevant variants. The presence of the *CYP21A1P* pseudogene, along with duplications, deletions, microgene conversions, and multiple variants of the same allele, makes it challenging to distinguish carrier alleles from non-carrier alleles using conventional molecular methods including short-read sequencing technologies [9–12]. Although 13 types of deletion chimeras have been reported [3, 6, 13, 14], the corresponding duplication chimeras remain uncharacterized. Comprehensive characterization of the variant spectrum of the *CYP21A2* gene locus is critical for a better understanding of CAH and for improved disease management. We recently reported a long-read

sequencing (LRS) assay, termed comprehensive analysis of CAH (CACAH) [9, 10, 12], which makes it methodologically possible to perform comprehensive and accurate genetic characterization of CAH at the population level.

Biochemical newborn screening (NBS), using 17-hydroxyprogesterone (17-OHP) as an indicator of classic CAH, has been routinely performed in over 40 countries [15]. Clinical diagnosis, follow-up and adequate treatment during the neonatal period are critical for achieving optimal long-term outcome [16]. However, the effectiveness of NBS for CAH has been hampered by the performance of biochemical assays using dried blood spot (DBS) samples. The positive predictive value (PPV) of the 17-OHP assays is low, especially in preterm and critically ill infants [17, 18]. With screening of more than 10 million newborns, false-negative results that lead to a missed diagnosis for 15–30% of classic CAH have been reported, particularly for less severe SV form, which can lead to failure or delays in the identification of infants with treatable conditions [18–22]. Additionally, newborns positive for the first 17-OHP screening need to be recalled for a second test a week or two weeks later, leading to a risk of delayed diagnosis or even missed diagnosis in newborns who failed to be recalled [23].

Measurement of steroid hormone is the standard method for diagnosing CAH. Genotyping has not used as a first-line diagnostic test because of the complexity of the *CYP21A2* locus [3]. In recent years, clinical studies have demonstrated improved screening efficacy by incorporating next-generation sequencing technologies as first-tier tests for a panel of diseases into current biochemical NBS programs [24–27]. However, genetic testing for CAH has not been studied as a first-tier screening method likewise owing to the complexity of the *CYP21A2* locus [3].

This prospective study applied CACAH for first-tier genetic screening of 21,239 newborns from 11 NBS centers, where routine biochemical 17-OHP screening was performed in parallel. The comprehensive data on *CYP21A2* locus-wide variant spectrum, carrier frequency, and expected incidence of both classic CAH and NCCAH in Chinese population, along with demonstration of the effectiveness of first-tier genetic NBS for CAH, would be insightful for clinicians and geneticists to improve their understanding and management of this serious and complex genetic disease.

Methods

Study design and participants

This prospective cross-sectional and observational study enrolled 21,239 newborns from 11 regional NBS centers across China (Fig. 1). The 11 centers (in alphabetical order) included Beihai People's Hospital, Children's

Hospital of Chongqing Medical University, Fujian Maternity and Child Health Hospital, Guangzhou Women and Children's Medical Center, Guiyang Maternal and Child Health Care Hospital, Jinan Maternal and Child Health Hospital, Liuzhou Maternity and Child Healthcare Hospital, Maternal and Child Health Hospital of Hubei Province, Northwest Women's and Children's Hospital, Women and Children's Hospital Xiamen University, and Xinhua Hospital. The samples were consecutively enrolled in each center for a two-month period between June 2023 and May 2024. The inclusion criteria were as follows: (1) written informed consent provided by the legal guardians of the participants; (2) DBS samples were collected on the third day after birth; and (3) at least one 8 × 8 mm DBS sample was available for genetic analysis. The exclusion criteria were as follows: (1) newborns who received blood transfusions; and (2) the specimens were unqualified for genetic testing. DBS samples from

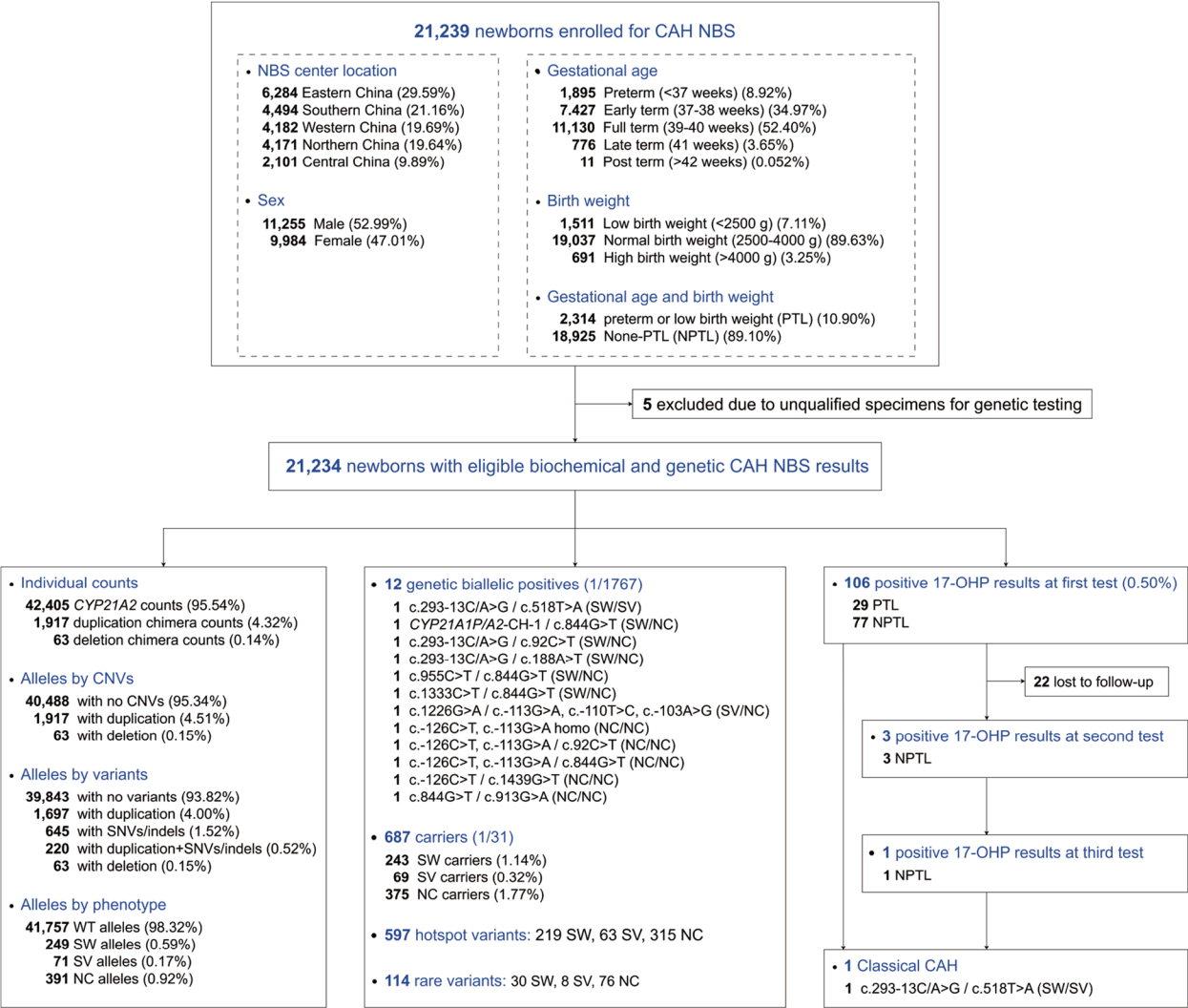


Fig. 1 Flow chart of study design, population characteristics, and summary of results

newborns were collected from newborns via heel prick using Whatman 903 filter paper for three 8×8 mm spots at approximately 72 h post-delivery. Each center collected the clinical information of the newborns and performed biochemical screening of 17-OHP locally. And one 8×8 mm spot of each newborn was sent to the clinical laboratory of Berry Genomics Corporation for CAH genetic testing. The sex of the newborns was determined based on their phenotypic presentation. There were 11,255 (52.99%) males and 9984 (47.01%) females. Of the 21,239 newborns, 1895 (8.92%) and 1511 (7.11%) were born prematurely and with low birth weight, respectively (Fig. 1).

NBS Biochemical Screening of 17-OHP

DBS samples were collected from newborns via heel prick using Whatman 903 filter paper at approximately 72 h post-delivery. The 17-OHP levels were measured by time-resolved fluoroimmunoassay using the GSP Neonatal 17-OHP Progesterone Kit (Perkin Elmer, Turku, Finland) in nine centers, and the Neonatal 17-OHP Screening Kit (Fenghua Biotech, Guangzhou, China) in two centers. For the Perkin Elmer kit, the 17-OHP cut-off values for newborns with preterm or low birth weight (PTL) and non-PTL (NPTL) were 25 and 12 nmol/L, respectively. For the Fenghua kit, the 17-OHP cutoff values for newborns with PTL and NPTL were 60 and 30 nmol/L, respectively. Infants with an initial positive result were re-screened within two weeks, and would be referred to pediatric endocrinologists for clinical evaluation if the second 17-OHP test was positive.

Genetic screening of CAH by long-read sequencing

Genomic DNA was extracted from DBS using MagMAX DNA Multi-Sample 96-Well kits (Thermo Fisher Scientific, USA). A previously described long-range multiplex PCR approach termed CACAH was used to specifically target *CYP21A2*, *CYP11B1*, *CYP17A1*, *STAR*, and *HSD3B2* genes [9]. CACAH specifically amplified *CYP21A2-TNXB* and *CYP21A1P-TNXA* with specific forward and reverse primers designed in the promoter regions of *CYP21A2* and *CYP21A1P*, and on the unique sequence of *TNXB* and *TNXA*, respectively. Therefore, CACAH can distinguish among *CYP21A2-TNXB*, *CYP21A1P-TNXA*, deletion chimeras, and duplication chimeras. The PCR amplification of target genes and construction of PacBio SMRTbell libraries were performed as previously described [9]. Briefly, the PCR product of each sample was ligated to a unique hairpin barcoded-adaptor to form a dumbbell-shaped pre-library using a one-step end-repair and ligation reaction. The reaction mix was then treated with exonucleases (Enzymatics, USA) to digest unligated DNA. The pre-libraries of each specific sample were purified, quantified and then pooled

with equal mass for up to 384-plex. The pooled library was then purified again with $0.6 \times$ Ampure PB beads (Pacific Biosciences, USA) and converted to the SMRTbell library using the Sequel II Binding Kit 3.2 (Pacific Biosciences, USA). The SMRTbell library was sequenced on the Sequel IIe platform in circular consensus sequencing (CCS) mode for 30 h (Pacific Biosciences, USA).

The raw sequencing subreads in bam files were processed to obtain high-quality CCS reads by ccs version 8.0.1 available at <https://github.com/PacificBiosciences/ccs> (Pacific Biosciences, USA), split into individual samples by debarcoding by lima version 2.13.0 available at <https://github.com/PacificBiosciences/barcoding> (Pacific Biosciences, USA), and then aligned to the genome reference GRCh38/hg38 by pbmm2 version 1.17.0 available at <https://github.com/PacificBiosciences/pbmm2> (Pacific Biosciences, USA) to specifically obtain the target reads for the *CYP21A2* loci, *CYP11B1*, *HSD3B*, *CYP17A1* and *STAR*. The CCS reads in the *CYP21A2* loci were assigned as *CYP21A2-TNXB*, *CYP21A1P-TNXA*, deletion chimera and duplication chimera according to the primer pair combinations using BLAST+ version 2.9.1 available at https://github.com/ncbi/blast_plus_docs [28]. The reads of a specific primer pair were considered as background if the ratio was less than 2% of the total number of *CYP21A2* and *CYP21A1P* reads. The pseudogene *CYP21A1P-TNXA* reads were excluded from further analysis. FreeBayes version 1.3.4, available at <https://github.com/freebayes/freebayes> [29] was employed to identify single-nucleotide variations (SNVs) and indels in the other three groups of *CYP21A2* CCS reads, and in *CYP11B1*, *CYP17A1*, *STAR*, and *HSD3B2* CCS reads. The 13 types of reported deletion chimeras (*CYP21A1P/A2-CH-1-10* and *TNXA/B-CH-1-3*) were distinguished by their junction sites [3, 6, 13, 14], and the 13 types of duplication chimeras (counterparts of deletion chimeras) with the same junction sites were named as *CYP21A2/A1P-CH-1-10* and *TNXB/A-CH-1-3* (Fig. 2A and B). The CCS reads of representative samples were displayed in the Integrative Genomics Viewer (IGV) to show *CYP21A2* variants. The reference build *CYP21A1P-TNXA* was aligned to reference build *CYP21A2-TNXB* and vice versa to better demonstrate the junction sites of chimeras and microgene conversions. WhatsHap v2.2 available at <https://github.com/whatsHap/whatsHap> [30] was used to determine the haplotypes of *CYP21A2-TNXB*, deletion and duplication chimeras across all the samples. The pathogenicity and predicted phenotype of *CYP21A2* SNVs/indels were interpreted according to the EMQN best practice guidelines for molecular genetic testing and reporting of 21-OHD [31]. Exonic and intronic variants of *CYP21A2* not listed in the EMQN guidelines were filtered with an allele frequency $\geq 1\%$, as shown in the gnomAD [32] or in this study cohort. Filtered exonic variants

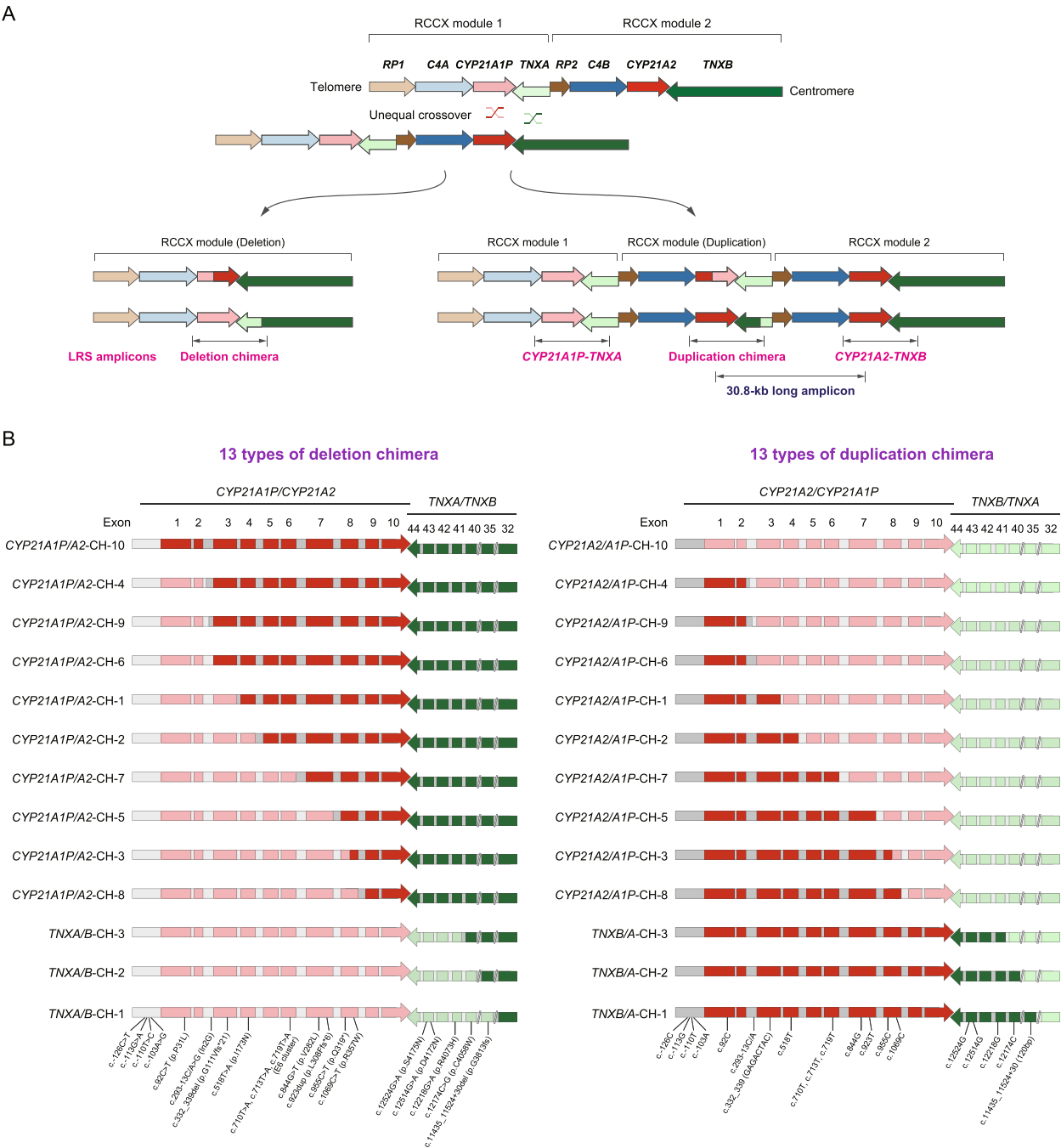


Fig. 2 Diagram showing the deletion and duplication chimeras of *CYP21A2* locus. **A** Diagram illustrating unequal crossover between two chromosomes with bimodular *RP-C4-CYP21A-TNX* (RCCX) that leads to formation of monomodular RCCX with deletion chimera *CYP21A1P/A2* or *TNXA/B*, and trimodular RCCX with duplication chimera *CYP21A2/A1P* or *TNXB/A*. **B** Diagram showing the 13 types of reported deletion chimera termed *CYP21A1P/A2-CH-1*–*10* and *TNXA/B-CH-1*–*3* as distinguished by recombination junction sites (adapted from Fig. 1B of Merke and Auchus, N Engl J Med. 2020 Sep 24;383(13):1248–1261.) (left panel) and the 13 types of duplication chimeras termed *CYP21A2/A1P-CH-1*–*10* and *TNXB/A-CH-1*–*3* (right panel), which have corresponding junction sites as the 13 types of deletion chimeras

were subjected to an in vitro enzymatic activity assay to determine their pathogenicity. Filtered intronic variants were predicted for their potential splicing-altering functions using SpliceAI [33]. Intronic variants with potential splicing defects or those within 20 bp flanking the exons

were subjected to a minigene assay to determine their pathogenicity.

For the long amplicon to discriminate cis- and trans-configuration between duplication chimera and SNVs/indels in *CYP21A2*, the primer pair Long-F

(5-CAGTCTCCATGTCSCAAAACAGTTC-3) and Long-R (5-GATGGTGGCATTGAGCAAGGGCAG-3) was used and the PCR was performed with KOD FX Neo (TOYOBO, Japan) according to the instruction manual with an extension time of 30 min. The SMRTbell library preparation and sequencing were performed as described for the CACAH assay.

In vitro expression of CYP21A2 and enzymatic activity assay

The wild-type (WT) expression plasmid CYP21A2-pcDNA3.1-3*Flag (Unibio Biotechnology, China) and Mut Express II Fast Mutagenesis Kit V2 (Vazyme Biotechnology, China) were used to generate the mutant plasmids. HEK293T cells were transfected with WT or mutant plasmids using the FuGENE HD Transfection Reagent (Promega, USA). For western blot analysis, HEK293T cells were seeded in 12-well plates at a level of 15×10^4 /well and transfected with 1 μ g of plasmid per well. Anti-CYP21A2 antibody (#ab230327, Abcam, UK) and the internal reference Vinculin Monoclonal antibody (#66,305-1-Ig, Proteintech, China) were used as the primary antibodies. For the CYP21A2 enzymatic activity assay, HEK293T cells were seeded in 24-well plates (1.5×10^5 cells/well) and transfected with the WT or mutant plasmids 12 h after cell culture. Then, 48 h after transfection, the cells were incubated for 1 h at 37 °C with 500 μ L of DMEM medium containing 2 μ mol/L of 17-hydroxyprogesterone (17-OHP, CAS#68-96-2) or progesterone (CAS#57-83-0), and 8 mmol/L of nicotinamide adenine dinucleotide phosphate. After 2 h of incubation, the supernatant was subjected to liquid chromatograph-mass spectrometer (AB Sciex Triple Quad 6500+ System, USA) to quantify 17-OHP and its metabolite 11-Deoxycortisol (CAS#152-58-9), as well as progesterone and its metabolite 11-Deoxycorticosterone (CAS#64-85-7). In transfected cell lines, the enzymatic activity of exonic CYP21A2 variants was measured as the relative conversion rate of 17-OHP to 11-Deoxycortisol and progesterone to 11-Deoxycorticosterone, as compared to that of WT CYP21A2. The CYP21A2 variant was classified as SW, SV, NC, and unaffected, with enzymatic activity of 0–1%, 1–10%, 10–80%, and >80%, respectively. CYP21A2 Q319*, I173N, and V282L were used as established SW, SV, and NC, respectively.

Minigene splicing assay

WT and mutant minigenes were constructed using the pcDNA3.1 vector. The minigene pcDNA3.1 vector, which encompassed the sequences from exons 1 to 4 of CYP21A2 (hg38 chr6:32,038,423–32,039,457), was synthesized by Tsingke (Tsingke Biotechnology, China). Mutant vectors for CYP21A2 variants c.202+5G>C, c.292+4A>C, and c.293-4G>A were constructed from

the WT vectors via site-directed mutagenesis using the Mut Express II Fast Mutagenesis Kit V2 (Vazyme Biotechnology, China). WT and mutant vectors were independently transfected into HEK293T cells using FuGENE® HD Transfection Reagent (Promega, USA). Cells were harvested using trypsin 48 h after transfection, and the total RNA was then extracted and subjected to RT-PCR and subsequent gel analyses. The DNA bands of interest were recovered and analyzed using Sanger sequencing. The CYP21A2 variant was classified as SW, SV, NC, and unaffected, with undisrupted transcript ratios of 0–1%, 1–10%, 10–80%, and >80%, respectively.

Estimation of the population disease incidence and carrier frequency

Hardy–Weinberg equilibrium was employed to estimate the incidence and carrier frequency of both classic CAH and NCCAH. The allele frequency of CAH (p) was defined as the sum of the allele frequencies for classic CAH (p_1) and NCCAH (p_2). The estimated incidence for classic CAH and NCCAH was calculated as $(p_1)^2$ and $p^2 - (p_1)^2$, respectively. The estimated carrier frequencies of classic CAH and NCCAH were calculated as $2 \times p_1 \times (1 - p_1 - p_2)$ and $2 \times p_2 \times (1 - p_1 - p_2)$, respectively.

Statistical analysis

Fisher's exact test was used to compare the distribution of haplotypes and single nucleotide polymorphisms (SNPs) between the different groups. A chi-square test was performed to analyze the differences in allele frequencies among different populations. The two-tailed Wilcoxon rank-sum test was used to analyze the differences in 17-OHP levels among the different genetic groups. Receiver operating characteristic (ROC) curve analysis was performed to analyze the sensitivity and specificity of 17-OHP screening using area under the curve (AUC) values. P value set at 0.05 was considered statistically significant.

Results

Study participants and NBS screening

A total of 21,239 newborns were recruited from 11 NBS centers across China, with 11,255 (52.99%) males and 9984 (47.01%) females. The mean (SD) gestational age at birth was 38.4 (± 1.7) weeks, and the mean (SD) birth weight was 3191 (± 494) grams (Fig. 1). The preterm birth and low birth weight rates were 8.92% (1895/21239) and 7.11% (1511/21239), respectively. In total, 10.89% (2314/21239) of infants were born prematurely or had low birth weight. All the 21,239 samples had biochemical NBS screening results for 17-OHP, with 21,234 (99.98%) samples yielding valid CAH genetic test results [34]. Five samples that failed to achieve quality control for CACAH were excluded from subsequent analyses (Fig. 1).

Characteristics of CYP21A2 alleles

Using the CACAH assay, 1917 (4.51%) alleles with duplication chimera, 63 (0.15%) alleles with deletion chimeras, and 40,488 (95.34%) alleles without duplication or deletion chimeras were detected among the 21,234 samples (Fig. 1), resulting in a total of 42,405 *CYP21A2* alleles. Haplotype analysis revealed 3122 different haplotypes for the 42,405 *CYP21A2* alleles, with the most frequent haplotype counted 3150 (Additional file1: Fig. S1). Ninety-eight high-frequency haplotypes had more than 50 *CYP21A2* alleles for each haplotype, and had cumulative count number of 33,119, accounting for 78.1% of the total number. Nine types of duplication chimeras (*CYP21A2/A1P-CH-1/2/4/5/7/8* and *TNXB/A-CH-1/2/3*) and ten types of deletion chimeras (*CYP21A2/A2-CH-1/3/4/5/6/7/8* and *TNXA/B-CH-1/2/3*) were identified (Table 1, Additional file1: Fig. S2). There were 223 different haplotypes for the 1917 duplication chimera alleles, and 53 different haplotypes for the 63 deletion chimera alleles (Additional file1: Fig. S2). These results demonstrated the overall diversity of *CYP21A2* and chimera alleles, as well as the conservative characteristics of a small subset of haplotypes.

LRS detected 71 different potentially disease-relevant SNVs/indels in *CYP21A2* (Additional file1: Table S1). Of the 71 variants, 47 were classified as SW, SV, or NC according to the EMQN guidelines, 13 were novel exonic variants, eight were exonic variants reported by the EMQN guidelines but with only rare cases, and three were intronic variants. In vitro function assays determined the residual enzymatic activity and classified the 21 exonic variants (Fig. 3A, Additional file1: Fig. S3). Notably, the variant *CYP21A2:c.371C>T* (T124I), which is highly prevalent in East and South Asia but not in other populations (Fig. 3B), was reclassified from SW to unaffected according to in vitro functional analysis. *CYP21A2:c.371C>T* had an allele frequency of 0.57% in this cohort, and was enriched in three haplotypes that accounted for 88.3% (212/240) of all the alleles (Fig. 3C). The minigene splicing assay demonstrated that the splicing of *c.292+4A>C*, *c.202+5G>C*, and *c.293-4G>A* was completely disrupted (SW), disrupted by 60% (NC), and unaffected, respectively (Fig. 3D and E).

CYP21A2 SNVs/indels and duplication chimeras, including 244 *TNXB/A-CH* and 12 *CYP21A2/A1P-CH*, were simultaneously identified in 256 samples. LRS with specifically designed 30.8-kb long amplicon identified duplication chimeras with SNVs/indels in the cis- and trans-configurations in 220 and 36 samples, respectively (Fig. 4A and B). Among the 220 cis samples, 214 harbored *CYP21A2:c.955C>T* and six had other SNVs/indels in *CYP21A2*. And additional 24 samples harbored *CYP21A2:c.955C>T* without any duplication chimeras. Among alleles carrying *CYP21A2:c.955C>T*, 17 SNPs

were significantly enriched in those with a duplication in cis (Fig. 4B, Additional file1: Table S2). These included the previously reported *c.293-79G>A* (rs114414746) in intron 2 and *c.*12C>T* (rs150697472) in the 3'-UTR [35]. For the duplication chimeras in cis with *c.955C>T*, *TNXB/A-CH-2* was significantly enriched (213/214), compared with that in the control group (43/1697) ($P<0.00001$). The distribution of 126 SNPs was significantly different between these two groups (Additional file1: Table S3), as represented by rs150697472, rs62402686, rs62402693, rs2471810, rs6466, and rs6462 (Fig. 4B).

Carrier allele frequency and disease incidence by Genetic NBS

Combined in vitro function and haplotype analysis enabled the precise identification of carrier alleles. Among the 21,234 samples, LRS identified 1697 (4.00%) alleles with duplications, 645 (1.52%) alleles with *CYP21A2* SNVs/indels, 220 (0.52%) alleles with both duplications and *CYP21A2* SNVs/indels, and 63 (0.15%) alleles with deletions (Fig. 1, Table 1). In total, 711 (1.67%) carrier alleles were successfully identified, including 249 (0.59%) SW, 71 (0.17%) SV, and 391 (0.92%) NC alleles. Hotspot variants including SNVs/indels caused by microgene conversion and 30-kb deletions caused by unequal crossover accounted for 84.0% (597/711) of the variants, whereas rare variants accounted for the remaining 16.0% (114/711) (Fig. 1). Twelve (1/1770) newborns harbored biallelic *CYP21A2* variants (Additional file1: Table S4), one of which had the classic SV type with *CYP21A2:c.293-13C/A>G/c.518 T>A* (Additional file1: Fig. S4), and the other 11 (1/1930) were NCCAH with at least one NC allele. The carrier frequency of *CYP21A2* was 3.23% (687/21234), with classic CAH and NCCAH carrier frequencies of 1.47% (312/21234) and 1.77% (375/21234), respectively (Fig. 1). Based on Hardy-Weinberg equilibrium calculations, the estimated incidence of classic CAH was $(0.75\%)^2 = 0.0057\%$ (1/17613), and the estimated incidence of NCCAH was $(1.67\%)^2 - 0.0057\% = 0.022\%$ (1/4474) (Additional file1: Table S5).

The allele frequencies of *CYP17A1*, *CYP11B1*, *STAR*, and *HSD3B2*, were 0.16%, 0.073%, 0.073%, and 0.038%, respectively (Additional file1: Table S6 to S9). No samples with biallelic variants were identified in any of these four genes.

CAH allele frequency and disease incidence among different ethnic groups

The carrier frequency in the studied Chinese population (3.23%) was lower than that reported in various European population-based studies (ranging from 4.0% to 7.5%), US Caucasians (11.0%), and Ashkenazi Jews (17.5%). This difference was particularly notable because these

Table 1 List of *CYP21A2* variants identified in the 21,234 samples

Type	Variants	Protein	Region	Carrier allele	Classification	Alleles	Alleles	Allele frequency
SNV/indel							645	1.52%
Nonsense	c.65G > A	p.W22*	Exon1	Y	SW	1	188	0.44%
Splicing	c.274A > G	p.R92G	Exon2	Y	SW	1		
Splicing	c.292 + 1G > A	NA	Intron2	Y	SW	3		
Splicing	c.292 + 4 A > C	NA	Intron2	Y	SW	1		
Splicing	c.293-13C/A > G	NA	Intron2	Y	SW	105		
Frameshift	c.332_339del	p.G111Vfs*21	Exon3	Y	SW	3		
Nonsense	c.377C > G	p.S126*	Exon3	Y	SW	1		
Nonsense	c.409G > T	p.E137*	Exon3	Y	SW	2		
Missense	c.535G > A	p.G179R	Exon4	Y	SW	1		
Frameshift	c.704_707dup	p.H236Qfs*61	Exon6	Y	SW	1		
Missense	E6 cluster(c.710 T > A, c.713 T > A, c.719 T > A)	p.I237N, p.V238E, p.M240K	Exon6	Y	SW	1		
Frameshift	c.740del	p.E247Gfs*11	Exon8	Y	SW	1		
Frameshift	c.923dup	p.L308Ffs*6	Exon7	Y	SW	1		
Nonsense	c.949C > T	p.R317*	Exon8	Y	SW	1		
Nonsense	c.955C > T	p.Q319*	Exon8	Y	SW	14		
Missense	c.1069C > T	p.R357W	Exon8	Y	SW	18		
Missense	c.1126G > A	p.G376S	Exon9	Y	SW	2		
Missense	c.1225C > T	p.R409C	Exon10	Y	SW	1		
Nonsense	c.1272C > A	p.C424*	Exon10	Y	SW	1		
Nonsense	c.1333C > T	p.R445*	Exon10	Y	SW	1		
Missense	c.1450C > T	p.R484W	Exon10	Y	SW	1		
Frameshift	c.1450dup	p.R484Pfs*40	Exon10	Y	SW	2		
Frameshift	c.1451_1452delinsC	p.R484Pfs*58	Exon10	Y	SW	8		
Frameshift	c.1454_1455del	p.G485Dfs*38	Exon10	Y	SW	1		
Complex	c.[−126C > T, −113G > A, 293-13C/A > G]	p.I173N	5'UTR, Intron2	Y	SW	1		
Complex	c.[−126C > T, −113G > A, −110T > C, −103A > G, 92 C > T, 293-13C/A > G, 332_339del, 518 T > A, E6_cluster, 923dup]	NA	5'UTR, Exon1/3/4/6/7, Intron2	Y	SW	1		
Complex	c.[293-13C/A > G, 332_339del]	NA	Intron2, Exon3	Y	SW	1		
Complex	c.[293-13C/A > G, 332_339del, 1069 C > T]	NA	Intron2, Exon3/8	Y	SW	1		
Complex	c.[E6 cluster, 844G > T, 923dup, 1069 C > T]	p.I237N, p.V238E, p.M240K, p.V282L, p.L308Ffs*6	Exon6/7/8	Y	SW	1		
Complex	c.[844G > T, 923dup]	p.V282L, p.L308Ffs*6	Exon7	Y	SW	1		
Complex	c.[844G > T, 923dup, 955 C > T]	p.V282L, p.L308Ffs*6	Exon7	Y	SW	1		
Complex	c.[844G > T, 923dup, 955 C > T, 1069 C > T]	p.V282L, p.L308Ffs*6	Exon7/8	Y	SW	1		
Complex	c.[923dup, 955 C > T]	p.L308Ffs*6	Exon7	Y	SW	3		
Complex	c.[923dup, 955 C > T, 1069 C > T]	p.L308Ffs*6	Exon7/8	Y	SW	2		
Complex	c.[955C > T, 1069 C > T]	p.Q319*	Exon7/8	Y	SW	3		
Missense	c.518 T > A	p.I173N	Exon4	Y	SV	60	69	0.16%
Missense	c.1070G > A	p.R357Q	Exon8	Y	SV	1		
Missense	c.1226G > A	p.R409H	Exon10	Y	SV	1		
Missense	c.1273G > A	p.G425S	Exon10	Y	SV	1		
Missense	c.1280G > A	p.R427H	Exon10	Y	SV	1		
Missense	c.1451G > A	p.R484Q	Exon10	Y	SV	2		

Table 1 (continued)

Type	Variants	Protein	Region	Carrier allele	Classification	Alleles	Alleles	Allele frequency
Complex	c.[−126C > T, −113G > A, 518 T > A]	p.I173N	5'UTR, Exon4	Y	SV	1		
Complex	c.[188A > T, 208G > T]	p.H63L, p.V70L	Exon1	Y	SV	2		
Promoter	c.−126C > T	NA	5'UTR	Y	NC	26	388	0.91%
Promoter	c.[−126C > T, −113G > A]	NA	5'UTR	Y	NC	132		
Promoter	c.[−126C > T, −113G > A, −110 T > C]	NA	5'UTR	Y	NC	17		
Promoter	c.[−113G > A, −110 T > C, −103A > G]	NA	5'UTR	Y	NC	6		
Promoter	c.[−126C > T, −113G > A, −110 T > C, −103A > G]	NA	5'UTR	Y	NC	8		
Promoter	c.−113G > A	NA	5'UTR	Y	NC	1		
Promoter	c.−103A > G	NA	5'UTR	Y	NC	7		
Missense	c.92C > T	p.P31L	Exon1	Y	NC	7		
Missense	c.143A > G	p.Y48C	Exon1	Y	NC	7		
Missense	c.188A > T	p.H63L	Exon1	Y	NC	13		
Splicing	c.202 + 5G > C	NA	Intron1	Y	NC	1		
Missense	c.373C > T	p.R125C	Exon3	Y	NC	1		
Missense	c.374G > A	p.R125H	Exon3	Y	NC	3		
Missense	c.797C > T	p.A266V	Exon7	Y	NC	7		
Missense	c.844G > T	p.V282L	Exon7	Y	NC	106		
Missense	c.913G > A	p.V305M	Exon7	Y	NC	23		
Missense	c.1019G > A	p.R340H	Exon8	Y	NC	2		
Missense	c.1099C > T	p.R367C	Exon8	Y	NC	2		
Missense	c.1100G > A	p.R367H	Exon8	Y	NC	1		
Missense	c.1108C > T	p.R370W	Exon8	Y	NC	1		
Missense	c.1174G > A	p.A392T	Exon9	Y	NC	2		
Missense	c.1175C > T	p.A392V	Exon9	Y	NC	1		
Missense	c.1195 T > C	p.W399R	Exon9	Y	NC	1		
Missense	c.1298C > T	p.P433L	Exon10	Y	NC	1		
Missense	c.1306C > T	p.R436C	Exon10	Y	NC	5		
Missense	c.1360C > T	p.R454C	Exon10	Y	NC	1		
Missense	c.1379C > A	p.R460H	Exon10	Y	NC	1		
Missense	c.1439G > T	p.R480L	Exon10	Y	NC	1		
Missense	c.1447C > T	p.P483S	Exon10	Y	NC	1		
Complex	c.[−126C > T, −113G > A, −110 T > C, −103A > G, 188 A > T]	p.H63L	5'UTR, Exon1	Y	NC	2		
Complex	c.[−126C > T, −113G > A, −110 T > C, −103A > G, 92 C > T]	p.P31L	5'UTR, Exon1	Y	NC	1		
Deletion							63	0.15%
	CYP21A1P/A2-CH-1	NA		Y	SW	23	61	0.14%
	CYP21A1P/A2-CH-3	NA		Y	SW	5		
	CYP21A1P/A2-CH-5	NA		Y	SW	3		
	CYP21A1P/A2-CH-6	NA		Y	SW	1		
	CYP21A1P/A2-CH-7	NA		Y	SW	1		
	CYP21A1P/A2-CH-8	NA		Y	SW	1		
	TNXA/B-CH-1	NA		Y	SW	13		
	TNXA/B-CH-2	NA		Y	SW	10		
	TNXA/B-CH-3	NA		Y	SW	4		
	CYP21A1P/A2-CH-4	NA		Y	SV	2	2	0.01%
Duplication							1697	4.00%
	CYP21A2/A1P-CH-1	NA		N	Normal	64	1697	4.00%
	CYP21A2/A1P-CH-2	NA		N	Normal	2		
	CYP21A2/A1P-CH-4	NA		N	Normal	9		

Table 1 (continued)

Type	Variants	Protein	Region	Carrier allele	Classification	Alleles	Alleles	Allele frequency
Duplication + SNV/indel	CYP21A2/A1P-CH-5	NA		N	Normal	3		
	CYP21A2/A1P-CH-7	NA		N	Normal	389		
	CYP21A2/A1P-CH-8	NA		N	Normal	40		
	TNXB/A-CH-1	NA		N	Normal	1135		
	TNXB/A-CH-2	NA		N	Normal	44		
	TNXB/A-CH-3	NA		N	Normal	11		
							220	0.52%
	CYP21A2/A1P-CH-1 + c.844G > T	p.V282L	Exon7	Y	NC	1	220	0.52%
	CYP21A2/A1P-CH-7 + c.844G > T	p.V282L	Exon7	Y	NC	1		
	CYP21A2/A1P-CH-7 + c.1307G > A	p.R436H	Exon10	Y	NC	1		
	TNXB/A-CH-1 + c.[−126C > T, −113G > A, −110T > C, −103A > G]	NA	NA	N	Normal	1		
	TNXB/A-CH-1 + c.518T > A	NA	NA	N	Normal	1		
	TNXB/A-CH-2 + c.−126C > T	NA	NA	N	Normal	1		
	TNXB/A-CH-2 + c.955C > T	NA	NA	N	Normal	213		
	TNXB/A-CH-3 + c.955C > T	NA	NA	N	Normal	1		
Total							2625	6.18%

previous studies primarily analyzed hotspot *CYP21A2* variants (Additional file1: Table S5) [8]. While the allele frequency of classic CAH was comparable among different populations, the allele frequency of NCCAH was significantly lower in Chinese population (0.92%) than in Ashkenazi Jewish (7.50%, $P<0.00001$) and US Caucasian populations (4.75%, $P<0.00001$). By Hardy–Weinberg equilibrium, the expected NCCAH incidence was also significantly lower in Chinese (1/4474) than in Ashkenazi Jewish (1/133) and US Caucasian populations (1/337) (Additional file1: Table S5). This difference was mainly caused by the hotspot NC variant *CYP21A2*:c.844G>T, with allele frequencies of 0.25% (106/42468), 7.00% (28/400), and 4.25% (17/400) in Chinese population (Fig. 5, Additional file1: Table S10), Ashkenazi Jewish, and US Caucasian populations [8], respectively. The allele frequencies of the three other NC variants (*CYP21A2*:c.188A>T, c.1174G>A, and c.1439G>T) were also lower in the studied Chinese population than in the other populations (Fig. 5). These results suggested that the incidence of NCCAH differs across populations, which may explain its lower incidence of NCCAH observed in this cohort.

Correlation between biochemical and genetic screening

Newborns with NPTL and PTL had significantly different basal 17-OHP levels and cutoffs ($P<0.0001$, Additional file1: Fig. S5). Therefore, the two groups were analyzed separately for genotype–phenotype correlations. No significant differences in 17-OHP levels were not observed among the NCCAH, SW carrier, SV carrier, NC carrier, and genetically negative subgroups in either the NPTL

(Fig. 6A) or PTL groups (Fig. 6B). ROC analysis indicated that 17-OHP levels could not distinguish NCCAH and carriers from the true genetic negatives in either the NPTL (Fig. 6C) or PTL groups (Fig. 6D).

Both first-tier biochemical and genetic NBS successfully identified a single case of a newborn patient with classic CAH (Fig. 1). Genetic screening did not generate any false positive results, whereas biochemical screening yielded 125 false positive results. The PPVs for first- and second-tier biochemical screening of classic CAH were only 0.94% (1/106, 95% CI: 0.83–1.21%) and 33.33% (1/3, 95% CI: 33.33–33.33%), respectively. Twenty-two newborns (20.75%, 22/106) with positive first-tier screening results were lost to follow-up. The recall rate for first-tier biochemical screening in the PTL groups was 1.25% (29/2314), and all the 29 recalled cases were false positives.

Discussion

Comprehensive genetic analysis of CAH in large population cohorts has been hindered by the complexity of the *CYP21A2* locus [3]. In this study, an LRS-based genetic approach was applied to CAH NBS in 21,239 newborns. Comprehensive characterization of the *CYP21A2* genetic landscape in a large cohort was achieved through precise genotyping combined with in vitro functional analysis of novel variants and phasing validation allowed for. We obtained the detected and estimated carrier frequency and incidence of classical CAH and NCCAH in Chinese population by precisely discriminating between carrier and non-carrier alleles. These results demonstrated the capability of LRS-based NBS as an effective first-tier screening method for CAH.

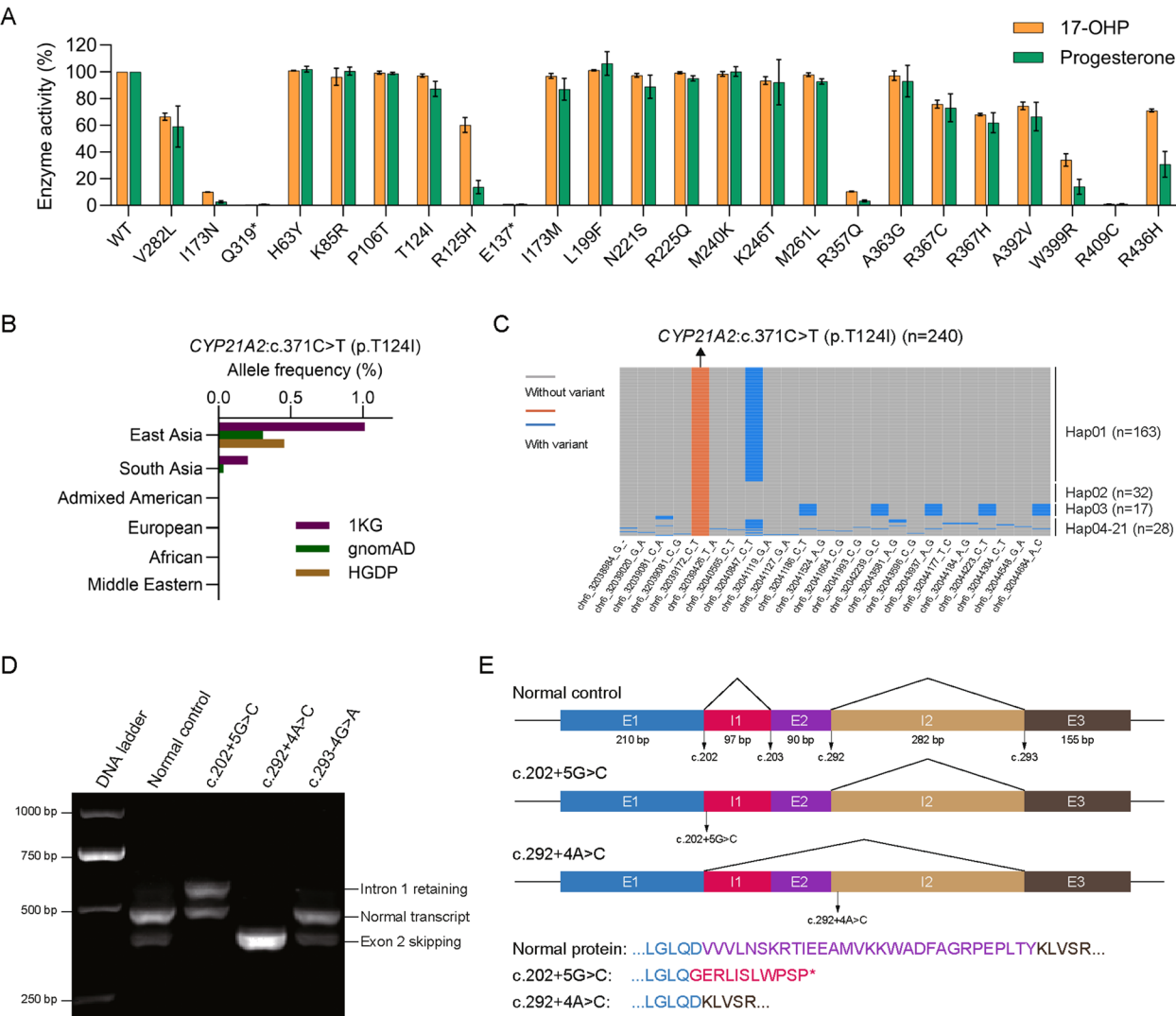
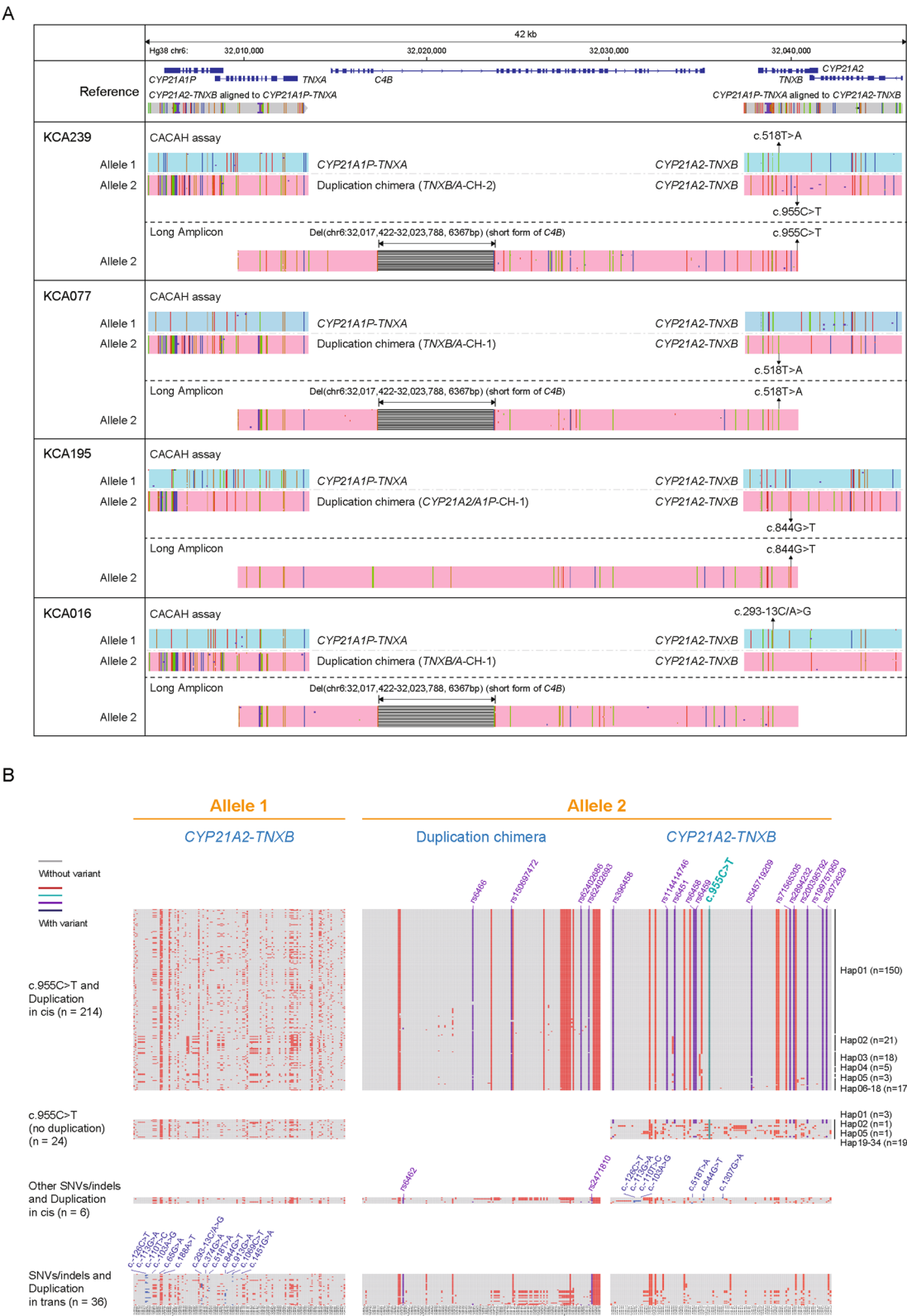


Fig. 3 Function and haplotype analysis to identify *CYP21A2* carrier alleles. **A** The in vitro functional analysis to detect residual enzymatic activity of mutant protein. Enzymatic activity was expressed as a percentage of conversion of 17-OHP and progesterone to their metabolites, with the specific activity of the WT *CYP21A2* as 100%. The experiments were repeated three times for each substrate. *CYP21A2* Q319*, I173N, and V282L were used as known controls for SW, SV, and NC, respectively. **B** The allele frequency of *CYP21A2*:c.371C>T across different populations from 1000 Genomes (1 KG), gnomAD, and Human Genome Diversity Project (HGDP). **C** The SNPs of 240 samples with *CYP21A2*:c.371C>T with heatmap. In total, 21 types of haplotypes were identified for the 240 samples. **D** Amplification of cDNA showing the transcripts of *CYP21A2* with intronic variants. **E** Diagram showing the change of protein sequence by c.202+5G>C and c.292+4A>C in *CYP21A2*

The complexity of *CYP21A2* lies in the difficulty of discriminating among functional gene *CYP21A2*, pseudogene *CYP21A1P*, recombination-caused microgene conversion, and deletion and duplication chimeras. We previously reported that the LRS-based CACAH assay could precisely and comprehensively differentiate all genotypes of the *CYP21A2* locus [9]. In this study, the classification of novel SNVs/indels identified by LRS was determined by in vitro functional analysis, whereas the phasing of SNVs/indels and duplication chimeras was determined by specifically designed long amplicon sequencing, which together allowed for comprehensive analysis of carrier alleles. Phasing between

duplication chimeras and SNVs/indels was necessary to identify whether the alleles with SNVs/indels were carrier alleles. This is particularly important for those with TNXB/A-CH chimeras because of a remaining intact *CYP21A2* gene. Notably, although the probability is low, the co-existence of c.955C>T and a duplication chimera does not definitively exclude carrier status. The co-existence could be in the trans configuration, given that the allele frequency of the duplication chimera was as high as 4.51%. Our identification of signature SNPs linked to c.955C>T and the duplication chimeras enabled direct phasing analysis by CACAH without the need for subsequent long-amplicon verification. The imbalance of



(See figure on previous page.)

Fig. 4 Determination of in cis/trans configurations between duplication chimeras and *CYP21A2* gene with SNVs/indels. **A** The CCS reads for CACAH assay and Long Amplicon for samples with both SNVs in *CYP21A2* and duplication chimera with IGV plots. The blue and pink blocks represent two different alleles. In sample KCA239, c.518T>A is located in allele 1, while c.955C>T and duplication chimera *TNXB/A-CH-2* are located in allele 2. In sample KCA077, c.518T>A and duplication chimera *TNXB/A-CH-1* are both located in allele 2 (in cis). In sample KCA195, the variant c.844G>T was in cis with the duplication chimera *CYP21A2/A1P-CH-1*. In sample KCA016, the variant c.293-13C/A>G was in trans with the duplication chimera *TNXB/A-CH-1*. **B** The SNPs in samples with both SNVs in *CYP21A2* and duplication chimera with heatmap. The blue blocks highlight disease-causing SNVs/indels of *CYP21A2*, and the purple blocks highlight represented different SNPs. Among them, rs6466, rs150697472, rs62402686, and rs62402693 represent SNPs enriched in duplication chimeras in-cis with *CYP21A2*:c.955C>T, rs6462 and rs2471810 represent SNPs enriched in other duplication chimeras, while rs396458, rs114414746, rs6451, rs6458, rs6459, rs545719209, rs71565305, rs2894232, rs200395792, rs199757950, and rs2072629 represent SNPs linked with *CYP21A2*:c.955C>T that is in-cis with duplication chimera

duplications to deletions found in this study is consistent with previous studies [3, 36, 37]. By leveraging advanced sequencing technology and a large cohort, our study provides the first accurate analysis of the population-level allele frequency of duplications and deletions in the *CYP21A2* locus.

More than 200 pathogenic variants of *CYP21A2* have been reported [31]. However, most CAH screening studies have only focused on hotspot PVs, including SNVs/indels caused by microgene transfer from the pseudo-gene *CYP21A1P* and the 30-kb deletion chimera, which were previously reported to account for approximately 95% of all variants [8]. By comprehensively analyzing the *CYP21A2* locus-wide variants, this population-based

study showed that common variants only accounted for only 84.0% of all the variants, whereas rare variants comprised as much as 16.0%. Among the variants of classic CAH (SW and SV), common and rare variants accounted for 88.1% (282/320) and 11.9% (38/320), respectively. These results highlight the necessity of including rare variants in genetic screening and diagnosis of CAH. Comprehensive and precise identification of carrier alleles is the foundation for accurately analyzing the expected incidence of both classic CAH and NCCAH, which are critical for better disease management.

The carrier frequency and disease incidence vary across different ethnicities. The carrier frequency in the studied Chinese population (3.28%) was lower than that

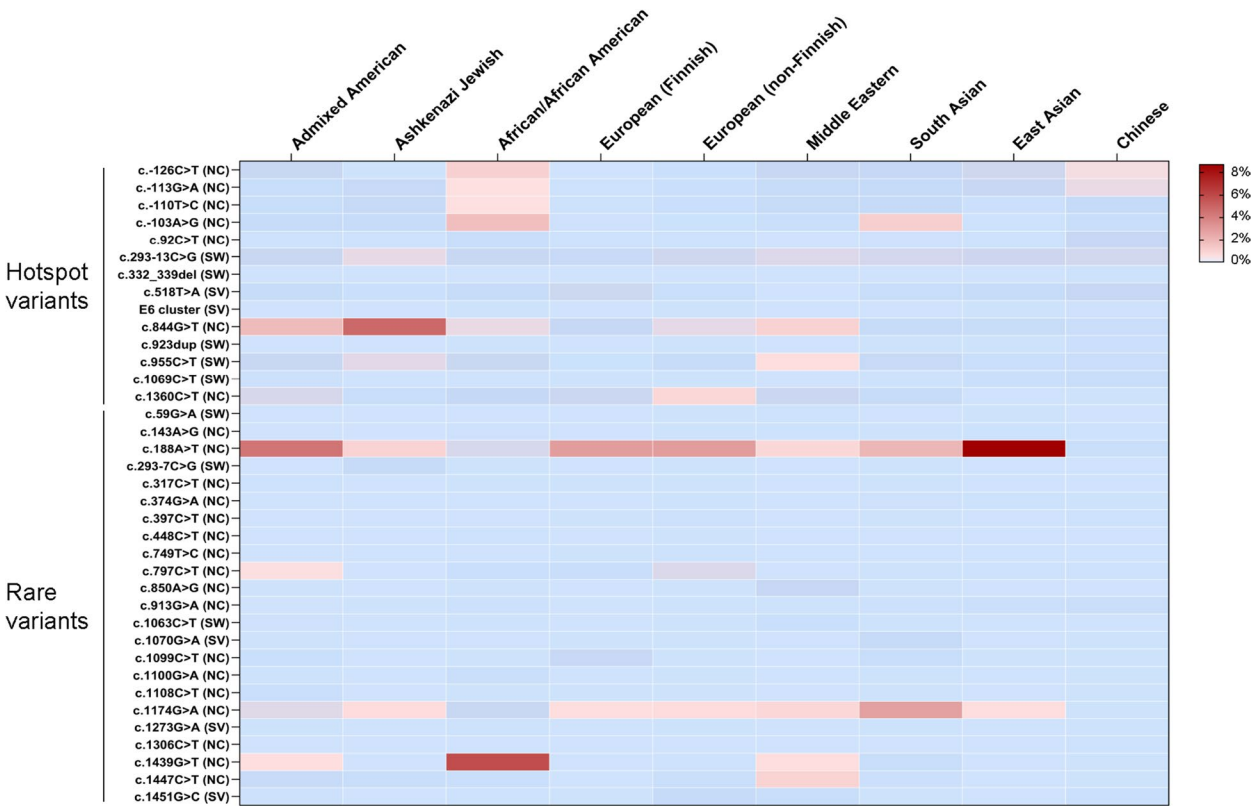


Fig. 5 Heatmap displaying the allele frequency of common and rare *CYP21A2* variants across different populations. The allele frequency of Chinese population was obtained from this study, and those of the other populations were from gnomAD. Variants with allele frequency of $\geq 0.01\%$ in at least one population were shown in the heatmap

reported in various European population-based studies across different ethnic groups (ranging from 4.0% to 7.5%), US Caucasians (11.0%), and Ashkenazi Jews (17.5%). This difference is particularly notable considering that these studies primarily analyzed hotspot *CYP21A2* variants [8]. The estimated incidence of classic CAH was 1/17613, which was in the reported range of 1/10000 to 1/20000 worldwide. The estimated incidence of NCCAH was much lower in Chinese cohort

(1/4474) than in Ashkenazi Jewish (1/133) and US Caucasian populations (1/337). This difference was mainly due to the much lower allele frequency of the hotspot NC variant *CYP21A2*:c.844G>T in Chinese population. The identified NCCAH incidence in this study was 1/1930, which was 2.3-fold higher than the expected frequency ($\chi^2=6.36$, $P=0.042$). Since this was a large cohort with no consanguineous marriages, the deviation between the expected and observed NCCAH incidence could be

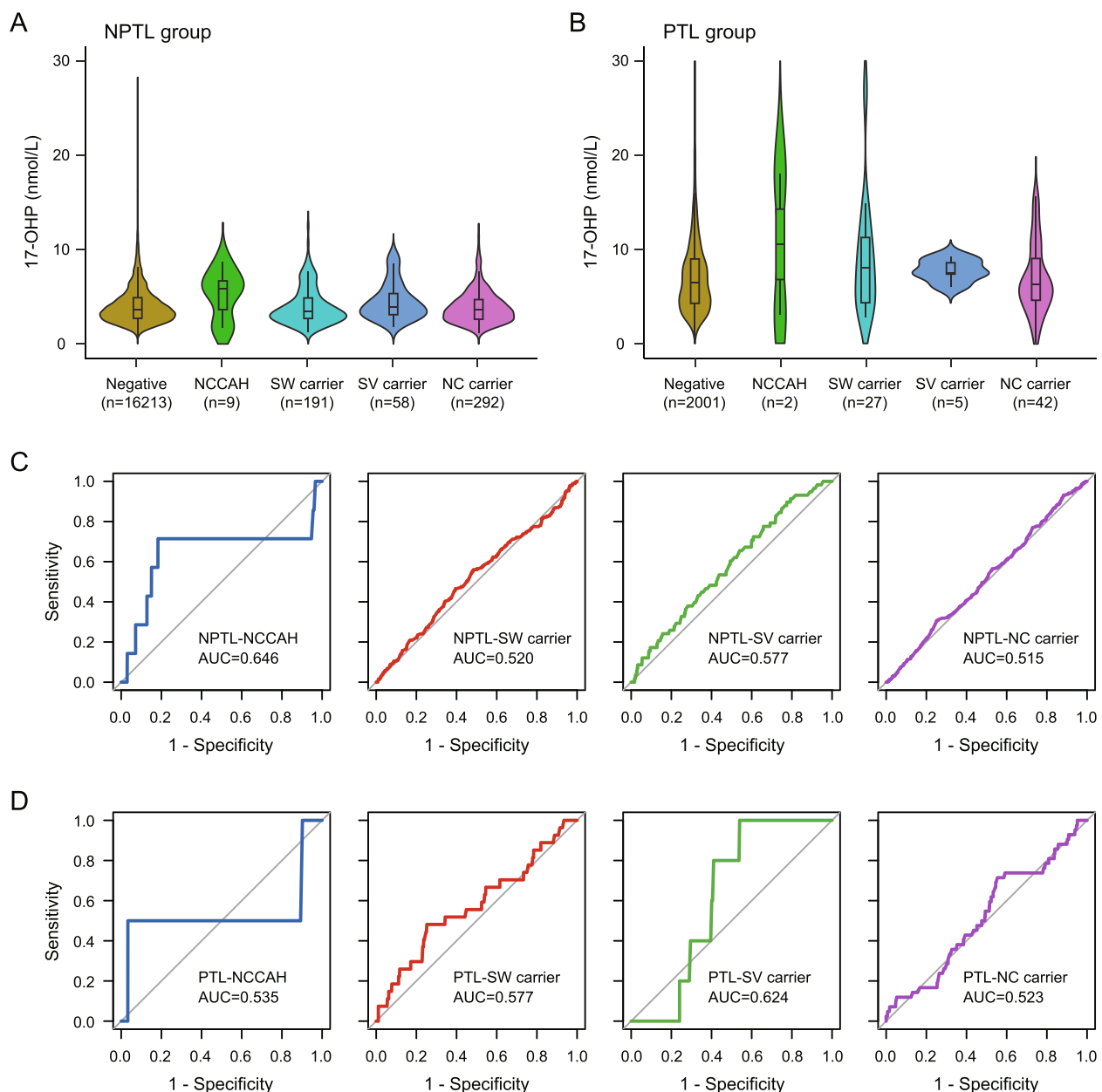


Fig. 6 Correlation between genetic and biochemical NBS screening. **(A and B)** The 17-OHP values among negative, NCCAH, SW carrier, SV carrier and NC carrier subgroups in NPTL **(A)** and PTL **(B)** newborn groups. Two-tailed Wilcoxon rank-sum test was performed for statistical analysis. **(C and D)** The sensitivity and specificity ROC curve analysis of 17-OHP screening to separate NCCAH, SW carrier, SV carrier and NC carrier subgroups from true genetic negatives in NPTL **(C)** and PTL **(D)** newborn groups

caused by de novo mutations, migration, or natural selection, as there is a possible survival advantage for fetuses carrying heterozygous *CYP21A2* alleles [8, 38–42].

Although first-tier biochemical NBS for CAH has substantially reduced classic CAH-related infant mortality, some clinical issues remain, including false negative results, low PPVs particularly in premature newborns, and incomplete recall [19, 21, 22]. Advancements in sequencing technologies offer a complementary approach to NBS [24, 43]. Combined first-tier LRS genetic screening for CAH/NCCAH detection has achieved almost 100% of sensitivity and specificity, while also providing critical information for genetic counseling. This is particularly valuable for families carrying classic CAH alleles, who face a significantly increased risk of classic CAH in future pregnancies [3]. However, reporting *CYP21A2* carrier status in a newborn screening context may face ethical issues, such as parental anxiety and uncertain implications for NCCAH, while it could be valuable to report carrier status of classic CAH for families planning for future pregnancies. Furthermore, the CACAH assay expanded the screening capabilities by detecting CAH-causing variants in genes beyond the current biochemical NBS coverage, including *CYP11B1*, *CYP17A1*, *STAR*, and *HSD3B2*.

Although underdiagnosed, NCCAH represents a prevalent global disorder with variable expressivity. Genetic screening now permits early diagnosis and timely intervention across pediatric and adult populations [7], a critical advance given the established links of NCCAH to treatable infertility and pregnancy loss [3], and NCCAH screening has recently been recommended in fertility evaluations regardless of ethnicity [8]. Although traditional NBS primarily targets classic CAH, biochemical assays detect only a fraction of NCCAH cases [18]. Our findings confirmed that first-tier LRS genetic NBS achieved NCCAH detection, enabling pre-symptomatic or timely management. However, the heterogeneous and late-onset manifestations of NCCAH necessitate careful counseling to manage the expectations of its clinical presentation.

In addition to accuracy, CACAH also demonstrated robust methodological feasibility as a first-line screening assay, achieving a 99.98% (21,234/21239) successful rate for DBS samples. Of these, 99.90% (21,217/21239) passed the initial quality control and only 0.08% (17/21239) required repeat testing. This approach eliminates the need for follow-up tests, which are typically required for biochemical screening. The current turnaround time for CACAH is six days, which could potentially be reduced to four days with further protocol optimization.

This study has some limitations. Owing to the insufficient sample size, only one newborn with classic CAH was identified, which made it challenging to fully determine and compare the sensitivities and PPVs of first-tier biochemical versus genetic screening for classic CAH.

However, the accuracy and efficiency of CACAH have been well-established in multiple cohort studies [9, 12, 13, 44–46], suggesting a minimal likelihood of false negatives or false positive results for genetic CAH screening. Implementing a combined first-tier biochemical and genetic screening strategy would further minimize the potential diagnostic errors.

Conclusions

Our study generated a comprehensive genetic landscape of the complex *CYP21A2* gene locus by integrating LRS technology with in vitro functional analysis and phasing validation in a cohort of 21,239 newborns. This approach provided a population-level carrier frequency and incidence estimates with a comprehensive landscape of the *CYP21A2* locus, while demonstrating the effectiveness of targeted LRS-based genetic NBS for CAH. First-tier CAH genetic screening offers a complementary method to significantly decrease both false-positive and false-negative rates, and to accelerate diagnostic timelines. These findings support the adoption of a combined first-tier genetic and biochemical screening approach for CAH, though further clinical validation is needed in the future.

Abbreviations

CAH	Congenital adrenal hyperplasia
CACAH	Comprehensive analysis of congenital adrenal hyperplasia
21-OHD	21-Hydroxylase deficiency
LRS	Long-read sequencing
NBS	Newborn screening
SW	Salt-wasting
SV	Simple-virilizing
NC	Non-classic
17-OHP	17-Hydroxyprogesterone
DBS	Dried blood spot
PPV	Positive predictive value
PTL	Preterm or low birth weight
NPTL	None-perterm or low birth weight
IGV	Integrative Genomics Viewer
ROC	Receiver operating characteristic
AUC	Area under the curve

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-025-01594-7>.

Supplementary Material 1

Acknowledgements

We thank all the families who participated in this study and the clinical teams involved in their care and project support.

Authors' contributions

D.L., L.W., W.Q., M.Z., W.Z., R.Q., L.Y., S.X., and A.M. conceptualized the research project. Y.Z., X.H., Y.H., H.J., J.T., H.L., and A.X. implemented the research. M.L., Y.L., Q.L., P.L., and Z.L. performed the function experiment and statistical analysis. R.W., D.W., R.Z., Q.P., J.Z., R.X., X.W., M.T., D.C., and C.W. did analysis for biochemical screening. A.M., Q.L., and D.C. did analysis for genetic screening and interpreted statistical findings. D.L., L.W., W.Q., W.Z., and M.Z. supervised the project. D.L., L.W., W.Q., M.Z., and A.M. drafted the original manuscript. All authors read and approved the final manuscript.

Funding

This research was supported by the National Key R&D Program of China (2022YFC2703400 and 2022YFC2703700), the National Natural Science Foundation of China (81770200), Shanghai Healthcare Commission Project (202340103), Clinical Innovation Project and Academic Climbing Project of Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University (23XHCR12B, XKPF2024A100, XKPF2024B100).

Data availability

The sequencing data from this study have been deposited securely in the Genome Sequence Archive (GSA) housed within the National Genomics Data Center (NGDC) at the China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences. All the datasets are in compliance with ethics approvals, patient consent and confidentiality agreements. The raw sequencing data are available under restricted access in GSA-Human (BioProject accession no.: PRJCA035456 (<https://ngdc.cncb.ac.cn/gsa-human/>)) [34] due to patient privacy and Regulations on the Management of Human Genetics Resources of China, following the GSA guidelines (<https://ngdc.cncb.ac.cn/gsa-human/document>). The detailed policies of data sharing and restrictions are listed in Principle for the Access of Human Genetic Resource Data in NGDC (https://ngdc.cncb.ac.cn/gsa-human/document/Principle_of_Accessing_Human_Genetic_Resource_Data_in_NGDC_V1.pdf). This dataset is restricted to research purposes only, and access is restricted to the specific research group and research collaborators who make the request. Dataset distribution to other people or groups is not allowed and the dataset cannot be used to identify individuals. Requests of data access can be sent to Dr. Desheng Liang (liangdesheng@sklmg.edu.cn). Upon approval and completion of a data use agreement, the dataset will be shared within three months. The allele frequency data from gnomAD v4 for different populations are available at <https://gnomad.broadinstitute.org/> [32].

Declarations

Ethics approval and consent to participate

The principal coordinating center was Central South University and study was approved by the institutional review board (2023–1–25). The other participating centers also obtained ethics approval from each of their institutional review boards. This research conformed to the principles of the Helsinki Declaration. Written informed consent to participate in this study was obtained from the parents of the participating newborns.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Center for Medical Genetics, Hunan Key Laboratory of Medical Genetics, School of Life Sciences, MOE KeyLab of Rare Pediatric Diseases, Central South University, 110 Xiangya Road, Changsha, Hunan 410078, China

²Department of Endocrinology and Genetic Metabolic Diseases, National Clinical Research Center for Child Health and Disorders, Children's Hospital of Chongqing Medical University, Chongqing 400013, China

³Department of Medical Genetics, Hunan Jiahui Genetics Hospital, Changsha, Hunan 410078, China

⁴Center of Neonatal Disease Screening, Department of Clinical Genetics, Northwest Women's and Children's Hospital, Xi'an 710061, China

⁵Prenatal Diagnosis Center, Guiyang Maternal and Child Health Care Hospital, Guiyang, Guizhou 550001, China

⁶Department of Endocrinology and Inborn Metabolic Diseases, Fujian Children's Hospital (Fujian Branch of Shanghai Children's Medical Center), College of Clinical Medicine for Obstetrics and Gynecology and Pediatrics, Fuzhou, Fujian 350014, China

⁷Medical Genetics Center, Maternal and Child Health Hospital of Hubei Province, Wuhan, Hubei 430070, China

⁸United Diagnostic and Research Center for Clinical Genetics, Women and Children's Hospital, School of Medicine and School of Public Health, Xiamen University, Xiamen 361102, China

⁹Clinical Molecular Medicine Center, Children's Hospital, Chongqing Medical University, National Clinical Research Center for Child Health and Disorders, Chongqing 400014, China

¹⁰Guangzhou Newborn Screening Center, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, Guangdong 510180, China

¹¹Department of Prenatal Diagnosis, Jinan Maternal and Child Health Hospital, Jinan, Shandong 250001, China

¹²Liuzhou Maternity and Child Healthcare Hospital, Liuzhou, Guangxi 545000, China

¹³Department of Obstetrics, Beihai People's Hospital, Beihai, Guangxi 536000, China

¹⁴Department of Pediatric Endocrinology and Genetic Metabolism, Xinhua Hospital, Shanghai Institute of Pediatric Research, School of Medicine, Shanghai Jiao Tong University, Shanghai 200092, China

¹⁵Medical Genetic Diagnosis and Therapy Center, Fujian Maternity and Child Health Hospital, College of Clinical Medicine for Obstetrics and Gynecology and Pediatrics, Fujian Medical University, Fuzhou, Fujian 350001, China

¹⁶Fujian Provincial Key Laboratory for Prenatal Diagnosis and Birth Defect, Fuzhou, Fujian 350001, China

¹⁷Xiamen Newborn Screening Center, Women and Children's Hospital, School of Medicine, Xiamen University, Xiamen, Fujian 361102, China

¹⁸Department of Research and Development, Berry Genomics Corporation, Beijing 102200, China

¹⁹Division of Neonatology and Center for Newborn Care, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangdong Provincial Clinical Research Center for Child Health, Guangzhou 510623, China

Received: 31 July 2025 / Accepted: 22 December 2025

Published online: 30 December 2025

References

1. Auer MK, Nordenström A, Lajic S, Reisch N. Congenital adrenal hyperplasia. *Lancet*. 2023;401(10372):227–44.
2. El-Maouche D, Arlt W, Merke DP. Congenital adrenal hyperplasia. *Lancet*. 2017;390(10108):2194–210.
3. Merke DP, Auchus RJ. Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *N Engl J Med*. 2020;383(13):1248–61.
4. Speiser PW, White PC. Congenital adrenal hyperplasia. *N Engl J Med*. 2003;349(8):776–88.
5. Hannah-Shmouni F, Chen W, Merke DP. Genetics of congenital adrenal hyperplasia. *Endocrinol Metab Clin North Am*. 2017;46(2):435–58.
6. New MI, Abraham M, Gonzalez B, Dumic M, Razzaghy-Azar M, Chitayat D, 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):2611–6.
7. Speiser PW, Arlt W, Auchus RJ, Baskin LS, Conway GS, Merke DP, et al. Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2018;103(11):4043–88.
8. Hannah-Shmouni F, Morissette R, Sinaii N, Elman M, Prezant TR, Chen W, et al. Revisiting the prevalence of nonclassic congenital adrenal hyperplasia in US Ashkenazi Jews and Caucasians. *Genet Med*. 2017;19(11):1276–9.
9. Liu Y, Chen M, Liu J, Mao A, Teng Y, Yan H, et al. Comprehensive analysis of congenital adrenal hyperplasia using long-read sequencing. *Clin Chem*. 2022;68(7):927–39.
10. Wang R, Luo X, Sun Y, Liang L, Mao A, Lu D, et al. Long-read sequencing solves complex structure of CYP21A2 in a large 21-hydroxylase deficiency cohort. *J Clin Endocrinol Metab*. 2024. <https://doi.org/10.1210/clinem/dgae519>.
11. Concolino P. Issues with the detection of large genomic rearrangements in molecular diagnosis of 21-hydroxylase deficiency. *Mol Diagn Ther*. 2019;23(5):563–7.
12. Yuan D, Cai R, Mao A, Tan J, Zhong Q, Zeng D, et al. Improved genetic characterization of congenital adrenal hyperplasia by long-read sequencing compared with multiplex ligation-dependent probe amplification plus Sanger sequencing. *J Mol Diagn*. 2024;26(9):770–80.

13. Zhang X, Gao Y, Lu L, Cao Y, Zhang W, Wu X, et al. Chimeric CYP21A1P/CYP21A2 genes in 21-hydroxylase deficiency detected by long-read sequencing and phenotypes correlation. *J Clin Endocrinol Metab.* 2025. <https://doi.org/10.1210/clinem/dgae819>.
14. Chen W, Xu Z, Sullivan A, Finkelstein GP, Van Ryzin C, Merke DP, et al. Junction site analysis of chimeric CYP21A1P/CYP21A2 genes in 21-hydroxylase deficiency. *Clin Chem.* 2012;58(2):421–30.
15. Heather NL, Nordenstrom A. Newborn screening for CAH-challenges and opportunities. *Int J Neonatal Screen.* 2021;7(1):11.
16. Lajic S, Karlsson L, Zetterström RH, Falhammar H, Nordenström A. The success of a screening program is largely dependent on close collaboration between the laboratory and the clinical follow-up of the patients. *Int J Neonatal Screen.* 2020;6(3):68.
17. Held PK, Bird IM, Heather NL. Newborn screening for congenital adrenal hyperplasia: review of factors affecting screening accuracy. *Int J Neonatal Screen.* 2020;6(3):67.
18. Gidlöf S, Wedell A, Guthenberg C, von Döbeln U, Nordenström A. Nation-wide neonatal screening for congenital adrenal hyperplasia in Sweden: a 26-year longitudinal prospective population-based study. *JAMA Pediatr.* 2014;168(6):567–74.
19. Sarafoglou K, Banks K, Kylo J, Pittock S, Thomas W. Cases of congenital adrenal hyperplasia missed by newborn screening in Minnesota. *JAMA.* 2012;307(22):2371–4.
20. Varness TS, Allen DB, Hoffman GL. Newborn screening for congenital adrenal hyperplasia has reduced sensitivity in girls. *J Pediatr.* 2005;147(4):493–8.
21. Held PK, Shapira SK, Hinton CF, Jones E, Hannon WH, Ojodu J. Congenital adrenal hyperplasia cases identified by newborn screening in one- and two-screen states. *Mol Genet Metab.* 2015;116(3):133–8.
22. Gidlöf S, Falhammar H, Thilén A, von Döbeln U, Ritzén M, Wedell A, et al. One hundred years of congenital adrenal hyperplasia in Sweden: a retrospective, population-based cohort study. *Lancet Diabetes Endocrinol.* 2013;1(1):35–42.
23. Li Z, Huang L, Du C, Zhang C, Zhang M, Liang Y, et al. Analysis of the screening results for congenital adrenal hyperplasia involving 7.85 million newborns in China: a systematic review and meta-analysis. *Front Endocrinol.* 2021;12:624507.
24. Ziegler A, Koval-Burt C, Kay DM, Suchy SF, Begtrup A, Langley KG, et al. Expanded newborn screening using genome sequencing for early actionable conditions. *JAMA.* 2025;333(3):232–40.
25. Roman TS, Crowley SB, Roche MI, Foreman AKM, O'Daniel JM, Seifert BA, et al. Genomic sequencing for newborn screening: results of the NC NEXUS project. *Am J Hum Genet.* 2020;107(4):596–611.
26. Adhikari AN, Gallagher RC, Wang Y, Currier RJ, Amatuni G, Bassaganyas L, et al. The role of exome sequencing in newborn screening for inborn errors of metabolism. *Nat Med.* 2020;26(9):1392–7.
27. Bodian DL, Klein E, Iyer RK, Wong WS, Kothiyal P, Stauffer D, et al. Utility of whole-genome sequencing for detection of newborn screening disorders in a population cohort of 1,696 neonates. *Genet Med.* 2016;18(3):221–30.
28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10.
29. Erik G, Gabor M. Haplotype-based variant detection from short-read sequencing. *arXiv preprint. arXiv:1207.3907* [q-bio.GN] 2012.
30. Patterson M, Marschall T, Pisanti N, van Iersel L, Stougie L, Klau GW, et al. Whatsap: weighted haplotype assembly for future-generation sequencing reads. *J Comput Biol.* 2015;22(6):498–509.
31. Baumgartner-Parzer S, Witsch-Baumgartner M, Hoepfner W. EMQN best practice guidelines for molecular genetic testing and reporting of 21-hydroxylase deficiency. *Eur J Hum Genet.* 2020;28(10):1341–67.
32. Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, et al. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature.* 2024;625(7993):92–100.
33. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell.* 2019;176(3):535–48.e24.
34. Liang D, Zhu M, Liang Q, Qiang R, Yu L, Xu S, et al. Data for: First-tier newborn screening of congenital adrenal hyperplasia by long-read sequencing. PRJCA035456. National genomics data center. 2025. Available from: <https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA035456>.
35. Doleschall M, Luczay A, Koncz K, Hadzsiev K, Erhardt É, Szilágyi Á, et al. A unique haplotype of RCCX copy number variation: from the clinics of congenital adrenal hyperplasia to evolutionary genetics. *Eur J Hum Genet.* 2017;25(6):702–10.
36. Finkelstein GP, Chen W, Mehta SP, Fujimura FK, Hanna RM, Van Ryzin C, et al. Comprehensive genetic analysis of 182 unrelated families with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *J Clin Endocrinol Metab.* 2011;96(1):E161–72.
37. Kleinle S, Lang R, Fischer GF, Vierhapper H, Waldhauser F, Föding M, et al. Duplications of the functional CYP21A2 gene are primarily restricted to Q318X alleles: evidence for a founder effect. *J Clin Endocrinol Metab.* 2009;94(10):3954–8.
38. Mantovani V, Dondi E, Larizza D, Cisternino M, Bragiani M, Viggiani M, et al. Do reduced levels of steroid 21-hydroxylase confer a survival advantage in fetuses affected by sex chromosome aberrations? *Eur J Hum Genet.* 2002;10(2):137–40.
39. Witchel SF, Lee PA, Suda-Hartman M, Trucco M, Hoffman EP. A favorable metabolic and antiatherogenic profile in carriers of CYP21A2 gene mutations supports the theory of a survival advantage in this population. *Horm Res.* 2009;72(6):337–43.
40. Nordenström A, Svensson J, Lajic S, Frisén L, Nordenskjöld A, Norrby C, et al. Carriers of a Classic CYP21A2 Mutation Have Reduced Mortality: A Population-Based National Cohort Study. *J Clin Endocrinol Metab.* 2002;104(12):6148–54.
41. Nordenström A, Butwick A, Lindén Hirschberg A, Almqvist C, Nordenskjöld A, Falhammar H, et al. Are carriers of CYP21A2 mutations less vulnerable to psychological stress? A population-based national cohort study. *Clin Endocrinol (Oxf).* 2017;86(3):317–24.
42. Livadas S, Dracopoulou M, Lazaropoulou C, Papassotiropoulos I, Sertedaki A, Angelopoulos GN, et al. A favorable metabolic and antiatherogenic profile in carriers of CYP21A2 gene mutations supports the theory of a survival advantage in this population. *Horm Res.* 2009;72(6):337–43.
43. Chen T, Fan C, Huang Y, Feng J, Zhang Y, Miao J, et al. Genomic sequencing as a first-tier screening test and outcomes of newborn screening. *JAMA Netw Open.* 2023;6(9):e2331162.
44. Wang Y, Zhu G, Li D, Pan Y, Li R, Zhou T, et al. High clinical utility of long-read sequencing for precise diagnosis of congenital adrenal hyperplasia in 322 probands. *Hum Genomics.* 2025;19(1):3.
45. Wang X, Lu X, Zheng F, Lin K, Liao M, Dong Y, et al. Assessment of long-read sequencing-based congenital adrenal hyperplasia genotyping assay for newborns in Fujian, China. *Int J Neonatal Screen.* 2025. <https://doi.org/10.3390/ijns11010022>.
46. Zhang R, Cui D, Song C, Ma X, Cai N, Zhang Y, et al. Evaluating the efficacy of a long-read sequencing-based approach in the clinical diagnosis of neonatal congenital adrenocortical hyperplasia. *Clin Chim Acta.* 2024;555:117820.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.