

Universidade Federal do Rio Grande do Sul
Centro de Biotecnologia
Programa de Pós-Graduação em Biologia Celular e Molecular

Título

Hiperplasia adrenal congênita: predição e elucidação do efeito de mutações no gene *CYP21A2* humano na estrutura e na atividade da enzima 21-hidroxilase

Tese de Doutorado

Mayara Jorgens Prado

Porto Alegre

20 de janeiro de 2022

Universidade Federal do Rio Grande do Sul
Centro de Biotecnologia
Programa de Pós-Graduação em Biologia Celular e Molecular

Título

Hiperplasia adrenal congênita: predição e elucidação do efeito de mutações no gene *CYP21A2* humano na estrutura e na atividade da enzima 21-hidroxilase

Aluna

Mayara Jorgens Prado

Tese submetida ao Programa de Pós-Graduação em Biologia Celular e Molecular do Centro de Biotecnologia da UFRGS como requisito parcial para a obtenção do título de Doutor em Ciências.

Orientadora: Maria Lucia Rosa Rossetti

Co-orientador: Arnaldo Zaha

Instituições e Fontes Financiadoras

O presente trabalho foi realizado em:

- Centro de Biotecnologia - Universidade Federal do Rio Grande do Sul.
- Centro de Desenvolvimento Científico e Tecnológico - Secretaria da Saúde do Estado do Rio Grande do Sul.
- *Department of Biomedical Research - Universität Bern.*

Agências financiadoras:

- Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)
- Secretaria da Saúde do Estado do Rio Grande do Sul (SES/RS)
- *Swiss Government Excellence Scholarship (ESKAS)*
- *Universität Bern (UniBe)*

Agradecimentos

Aos meus orientadores Prof^a. Lucia Rossetti e Prof. Arnaldo Zaha por terem aceitado me orientar e pelos valiosos ensinamentos.

Ao Prof. Amit Pandey por ter aceitado me receber em seu laboratório, pelos valiosos ensinamentos científicos e pelo suporte a mim dado durante minha estadia em Bern/CH.

Aos membros da banca desta tese, Dra. Fernanda C. Lopez, Dr. Pablo R. Arantes e Dra. Poli Marra Spritzer, pela disponibilidade que tiveram para contribuir e enriquecer este trabalho.

Ao Prof. Rodrigo Ligabue-Braun pelos ensinamentos e ajuda científica durante todas as etapas de bioinformática deste trabalho.

À Profa. Christa Frück pela receptividade e valiosos ensinamento e aos colegas e amigos de seu laboratório, Kay, Rawda, Emanuelle, Chrysanthi e Idoia.

Aos colegas e amigos dos laboratórios do CDCT/CEVS/SES-RS, em especial à Thaianne, Betânia, Grazi, Richard, Sun Hee, Tarciana, Cláudia, Regina e Anelise.

Aos colegas e amigos dos laboratórios 204/206/210 do Centro de Biotecnologia da UFRGS, em especial à Cristine, Bruna, Nathan, Lais, Karina e Henrique.

Aos colegas e amigos do laboratório LABPANDEY, por terem me recebido e dado suporte durante minha estada em Bern, em especial à Natalia, Shripriya e Katyayani.

Às minhas amigas da biotec, Amanda, Cristine, Luiza e Thaianne, pela amizade, apoio e ajuda para todas as horas desde o primeiro semestre de faculdade.

Às agências de fomento que viabilizaram a realização deste trabalho, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e *Swiss Government Excellence Scholarship* (ESKAS), e ao Programa de Pós-Graduação em Biologia Celular e Molecular (PPGBCM), em especial à Silvia Regina Centeno que cuida de todos os alunos do PPGBCM com muito carinho, bom humor e atenção.

À minha família, em especial à minha irmã Márcia e ao meu pai João por sempre acreditarem em mim, pelo incentivo de não desistir dos sonhos e por todo o suporte e amor incondicional, que foram ainda mais cruciais durante todo o ano que estive sozinha na Suíça. À minha Vó Tota, tia Otília e tia Marina pelo carinho, pelo amor e por sempre torcerem por mim.

Ao meu companheiro de vida, Mauro, por sempre me incentivar e me ajudar a concretizar os meus sonhos, pelo apoio e compreensão durante todo o período do meu doutorado, e por ter sido fundamental para eu terminar esse trabalho com a mente em paz.

Enfim, a todos que, de alguma forma, participaram da construção e realização deste trabalho.

MUITO OBRIGADA!!

Lista de abreviaturas, símbolos e unidades

Ac	Acurácia
ACTH	<i>Adrenocorticotropic hormone</i> (Hormônio adrenocorticotrófico)
CRH	<i>Corticotropin releasing hormone</i> (Hormônio liberador de corticotrofina)
CYPs	Superfamília de proteínas citocromo P450
DHEA	Dehidroepiandrosterona
DNA	Ácido desoxirribonucleico
Ep	Especificidade
HAC	Hiperplasia Adrenal Congênita
HAC-PS	Hiperplasia Adrenal Congênita, forma clássica perdedora de sal
HAC-VS	Hiperplasia Adrenal Congênita, forma clássica virilizante simples
HAC-NC	Hiperplasia Adrenal Congênita, forma não clássica
HGVS	<i>Human Genome Variation Society</i>
kb	Kilobases
MCC	<i>Matthews' correlation coefficient</i>
MLPA	<i>Multiplex Ligation-dependent Probe Amplification kit</i>
NADPH	Fosfato de dinucleotídeo de nicotinamida e adenina
NGS	<i>Next-Generation Sequencing</i> (sequenciamento de nova geração)
nt	Nucleotídeo
OMIM	<i>Online Mendelian Inheritance in Man</i>
pb	Par de bases
PCR	<i>Polymerase Chain Reaction</i> (Reação em Cadeia da Polimerase)
RCCX	Módulo composto pelos genes <i>RP1/RP2</i> , <i>C4A/C4B</i> , <i>CYP21A1P/CYP21A2</i> e <i>TNXA/TNXB</i>
RNA	Ácido ribonucleico
rs#	<i>Reference sequence</i> (sequência de referência)
Se	Sensibilidade
SNaPshot	<i>Single Nucleotide Primer Extension</i>
VPN	Valor Preditivo Negativo
VPP	Valor Preditivo Positivo

Lista de Figuras e Tabelas

FIGURA 1. CORTE HISTOLÓGICO DA GLÂNDULA ADRENAL HUMANA.....	15
FIGURA 2. ROTA DE BIOSÍNTESE DE ESTEROIDES HUMANA	19
FIGURA 3. REPRESENTAÇÃO ESQUEMÁTICA DA VIA DE BIOSÍNTESE DO CORTISOL.....	21
FIGURA 4. INCIDÊNCIA DE HIPERPLASIA ADRENAL CONGÊNITA POR ESTADO <i>BRASILEIRO</i> . 22	
FIGURA 5. REAÇÕES DE HIDROXILAÇÃO REALIZADAS PELA ENZIMA 21-HIDROXILASE (CYP21A2) HUMANA NO CÓRTEX ADRENAL	25
FIGURA 6. ESQUEMA DA FISIOLOGIA DE UMA CRISE DE PERDA DE SAL ADRENAL DEVIDO A COMPLETA PERDA DA ATIVIDADE DA ENZIMA 21-HIDROXILASE..	27
FIGURA 7. ESCALA PRADER (I a V) PARA A CLASSIFICAÇÃO DE GENITÁLIAS ATÍPICAS....	29
FIGURA 8. FREQUÊNCIA DA ESCALA PRADER OBTIDA NA ANÁLISE DE 154 INDIVÍDUOS 46,XX	33
FIGURA 9. ESQUEMA DA COMPOSIÇÃO DO MÓDULO RCCX NO CROMOSSOMO 6.....	36
FIGURA 10. ESQUEMA DAS MUTAÇÕES PATOGÊNICAS MAIS COMUNS ENCONTRADAS NO GENE CYP21A2 SÃO PROVENIENTES DE EVENTOS DE RECOMBINAÇÃO GÊNICA COM O PSEUDOGENE CYP21A1P.	38
FIGURA 11. CORRELAÇÃO DE GENÓTIPO E FENÓTIPO.....	42
FIGURA 12. VIRILIZAÇÃO DA GENITÁLIA EXTERNA CLASSIFICADA COM A ESCALA PRADER POR GRUPOS DE MUTAÇÃO.	47
FIGURA 13. QUADRO CLÍNICO DA HIPERPLASIA ADRENAL CONGÊNITA.....	52
FIGURA 14. ESTRUTURA EM FORMA DE PRISMA TRIANGULAR TÍPICA DA FAMÍLIA DE PROTEÍNAS CITOCROMO P450S.	57
FIGURA 15. ESTRUTURAS TRIDIMENSIONAIS DA PROTEÍNA 21-HIDROXILASE OBTIDAS POR CRISTALOGRAFIA DE RAIOS X.....	59
FIGURA 16. ESTRUTURA QUÍMICA DOS SUBSTRATOS DA 21-HIDROXILASE EM SEUS ESTADOS NATURAIS.	60
FIGURA 17. ESQUEMA DA REAÇÃO DE ELETRO-TRANSFERASE MEDIADA PELA ENZIMA P450 OXIDO-REDUTASE (POR)..	65
TABELA 1. COMPARAÇÃO ENTRE AS MUTAÇÕES MAIS COMUNS ENCONTRADAS NO GENE CYP21A2 EM DIFERENTES POPULAÇÃO	40

Resumo

Contexto A hiperplasia adrenal congênita (HAC) devido à deficiência da CYP21A2 é uma doença genética com um amplo espectro fenotípico e bom prognóstico quando o portador é acompanhado e tratado corretamente. A análise genética do gene *CYP21A2* é considerada uma peça-chave em casos complexos, principalmente em portadores assintomáticos e casos positivos. Portanto, ferramentas acessíveis para inferir a patogenicidade de novas variantes e a elucidação de variantes não caracterizadas são essenciais para um diagnóstico preciso. **Objetivo** Investigar a patogenicidade de variantes não caracterizadas no *CYP21A2* reportadas nas populações brasileiras ou portuguesas, além de analisar o desempenho de ferramentas de predição *in silico* para categorizar variantes *missense* do *CYP21A2*. **Métodos** Ferramentas computacionais de predição de dano e modelagem estrutural foram usadas para analisar a conservação de resíduos e alterações físico-químicas e estruturais de variantes não caracterizadas encontradas nas populações brasileira ou portuguesa. A atividade residual da enzima foi obtida por ensaio funcional através das proteínas *CYP21A2* do tipo selvagem e mutantes expressas em células HEK293. Ademais, nós acessamos o desempenho de preditores *in silico* através de variantes do *CYP21A2* caracterizadas *in vitro* por ensaio funcional. No total, 103 variantes foram usadas para testar dezessete programas de predição de dano, treze simples e quatro meta-preditores. **Resultados** As análises computacionais e funcionais de variantes não caracterizadas mostraram que todas as seis variantes selecionadas diminuem a atividade da *CYP21A2*. As variantes p.P35L e p.L199P causam uma atividade parcial da enzima, correspondente à forma tardia de HAC. Enquanto, as variantes

p.W202R, p.E352V p.P433L e p.R484L diminuem a atividade da enzima a um nível residual compatível com a forma clássica virilizante simples. Quanto ao desempenho das ferramentas de predição, todos os programas testados foram capazes de identificar variantes patogênicas. As acurácias obtidas variaram entre 0,69 e 0,97 e o MCC entre 0,49-0,90, sendo que os melhores valores foram obtidos para os programas CADD, ConSurf, DANN e ProlyPhen-2. O programa PANTHER-PSEP não foi capaz de identificar variantes neutras. **Conclusão** Elucidamos o impacto na estrutura e na funcionalidade de seis variantes com perfis desconhecidos, e mostramos boa correlação entre os estudos *in silico* e *in vitro*. Além disso, identificamos quatro programas de predição com bom desempenho para classificar variantes patogênicas e neutras na proteína CYP21A2. Portanto, esse estudo enfatiza a importância de se usar múltiplas ferramentas de predição em conjunto e contribui para futuras análises genéticas da deficiência da CYP21A2.

Abstract

Context Congenital adrenal hyperplasia (CAH) due to CYP21A2 deficiency is a genetic disease with a broad phenotypic spectrum and good prognosis when the carrier is followed-up and treated correctly. Genetic analysis of *CYP21A2* gene is considered a key component in complex CAH phenotypes, mainly with asymptomatic carriers and false positive cases. Therefore, accessible tools to infer the pathogenicity of new variants and the elucidation of uncharacterized variants are essential to a precise diagnosis. **Aim** Investigate the pathogenicity of uncharacterized variants in the CYP21A2 reported in the Brazilian or Portuguese populations and analyze the performance of *in silico* prediction tools to categorize missense variants of the CYP21A2. **Methods** Computational tools of damage prediction and structural modeling were used to analyze the conservation of residues and physicochemical and structural changes of uncharacterized variants found in the Brazilian or Portuguese populations. The residual enzyme activity was obtained by functional assay with wild-type and mutants CYP21A2 proteins expressed in HEK293 cells. Furthermore, we accessed *in silico* predictors performance through variants of the CYP21A2 characterized *in vitro* by functional assays. In total, 103 variants were used to test seventeen predictors' programs, thirteen single-predictors and four meta-predictors. **Results** Computational and functional analyzes of uncharacterized variants showed that all six selected variants decrease the CYP21A2 activity. The variants p.P35L and p.L199P cause partial enzyme activity, corresponding to the late form of CAH. Whereas the variants p.W202R, p.E352V, p.P433L, and p.R484L decrease the enzyme activity to a residual level compatible with the classic simple virilizing form. Concerning the

performance of the *in silico* predictor tools, all programs were able to identify pathogenic variants. The accuracies obtained ranged between 0.69 and 0.97 and the MCC between 0.49-0.90, and the better accuracy and MMC were obtained with CADD, ConSurf, DANN and ProlyPhen-2. The program PANTHER-PSEP was not able to identify neutral variants. **Conclusion** We elucidated the enzyme structure and functionality impact of six variants with unknown profiles, and we showed good correlation between *in silico* and *in vitro* studies. Furthermore, we identified four predictors' programs with good performance to discriminate pathogenic and neutral variants of the CYP21A2 protein. Therefore, this study emphasizes the relevance of using multiple algorithms together and contributes to future genetic analysis of the CYP21A2 deficiency.

Sumário

Agradecimentos.....	iv
Lista de abreviaturas, símbolos e unidades	vi
Lista de Figuras e Tabelas	vii
Resumo	viii
Abstract	x
Capítulo I - Introdução.....	14
1. Revisão Bibliográfica.....	15
1.1 Glândula adrenal e sua síntese de esteroides	15
1.2 A Hiperplasia adrenal congênita.....	20
1.3 A enzima 21-hidroxilase (CYP21A2).....	23
1.3.1 Categorias fenotípicas	31
1.4 O gene codificador da proteína CYP21A2.....	34
1.5 Relação genótipo e fenótipo.....	41
1.5.1 Variações na categorização das mutações	44
1.5.2 Influência de fatores externos no fenótipo de HAC.....	46
1.6 Diagnóstico e terapia de HAC	49
1.6.1 Pré-natal	49
1.6.2 Pós-natal.....	50
1.6.3 Genético	54
1.7 Elucidação de novas variantes	55
1.7.1 Caracterização <i>in silico</i>	55
1.7.2 Caracterização <i>in vitro</i>	66
2. Justificativa	69
Capítulo II: Objetivos	71
Objetivo Geral.....	72
Objetivos específicos	72
Capítulo III – Artigo 1: <i>Characterization of mutations causing CYP21A2 deficiency in Brazilian and Portuguese populations</i>	73
Capítulo IV – Artigo 2: <i>Meta-analysis of variant predictions in congenital adrenal hyperplasia caused by mutations in CYP21A2</i>	100
Capítulo V - Discussão.....	148

Capítulo VI: Conclusões	163
Capítulo VII – Referências Bibliográficas	165
Apêndice I – Certificado de aceite do artigo 1	176
Apêndice II – Submissão do artigo 2	177
Apêndice III – Colaboração em artigo	178
CURRICULUM VITÆ resumido	185

Capítulo I - Introdução

1. Revisão Bibliográfica

1.1 Glândula adrenal e sua síntese de esteroides

A Glândula adrenal tem um formato triangular que contorna o polo dorsal dos rins. Em mamíferos, ela é composta de dois tipos de tecidos organizados em uma conexão bidirecional, a medula e o córtex (Figura 1). A origem embrionária desses dois tecidos é distinta, as células da medula são derivadas da crista neural, enquanto o córtex é originado no mesoderma intermediário, assim como o ovário e testículo (MESIANO *et al.*, 1997). Ademais, a formação da medula adrenal é influenciada pelo córtex, que posiciona suas células primeiro, delimitando a região medular (BECHMANN *et al.*, 2021). Essa separação entre esses tecidos não é bem definida, havendo ainda uma conexão entre eles pelo sistema vascular adrenal e uma forte interação endócrino e parácrino (MESIANO *et al.*, 1997).

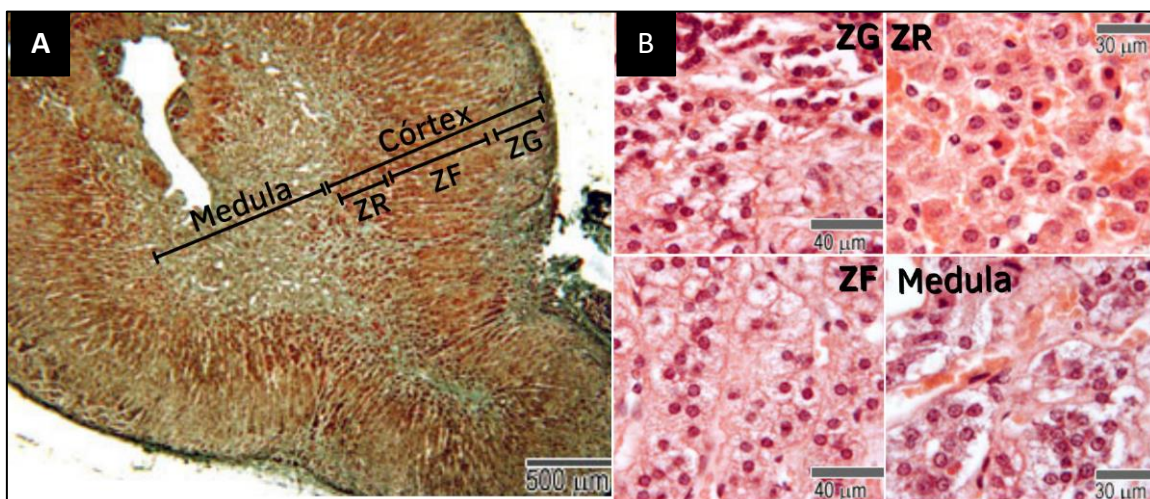


Figura 1. Corte histológico da glândula adrenal humana. (A) Localização das regiões do córtex e da medula. (B) Aumento óptico das células adrenais. ZG, zona glomerulosa; ZF, zona fasciculada; ZR, zona reticulada. Figura adaptada de PECKHAM *et al.* (2021).

As células da medula adrenal são caracterizadas por seu citoplasma granuloso e sem lipídios esteroides. Elas são altamente vascularizadas e secretam os hormônios catecolaminas (adrenalina, noradrenalina e a dopamina) (BECHMANN *et al.*, 2021). Esses hormônios são controlados pelo sistema nervoso simpático e atuam no coração, vasos sanguíneos, bronquíolos, músculo visceral, músculo esquelético e fígado (MESIANO *et al.*, 1997).

Já o córtex é classificado em três zonas em humanos, zona glomerulosa, zona fasciculada e zona reticulada (Figura 1). A zona glomerulosa apresenta células densamente compactadas com a área citoplasmática menor em relação a outras zonas, com aparência de bolas. A zona fasciculada é composta por células com citoplasmas ricos em gotículas lipídicas dispostas em colunas, fascículos. Por fim, a zona reticulada que apresenta células menores que a fasciculada, com menos gotículas lipídicas citoplasmáticas e organizadas na forma de uma rede que separa o córtex da medula (PIGNATTI *et al.*, 2017; SECCIA *et al.*, 2018).

No córtex da adrenal, a molécula de colesterol dá origem a três tipos de hormônios secretados: mineralocorticoides, glicocorticoides e andrógenos (Figura 2). Esses hormônios conhecidos como hormônios esteroides, corticosteroides ou apenas esteroides têm uma estrutura química semelhante à da molécula de colesterol e uma biossíntese tecido-específica, que é realizada majoritariamente no córtex adrenal (PIGNATTI *et al.*, 2017).

Mineralocorticoides são importantes hormônios para a regulação homeostática e tem sua secreção realizada na zona glomerulosa. O mineralocorticoide aldosterona tem um papel importante na regulação dos níveis

de íons sódio e potássio nos rins. Enquanto isso, a maioria dos glicocorticoides são excretados na zona fasciculada, sendo apenas alguns deles na zona reticulada. Os glicocorticoides são importantes para o metabolismo de carboidratos, proteínas e lipídeos. Seu esteroide mais sintetizado, o cortisol, é modulador de processos fisiológicos no metabolismo energético, na resposta ao estresse, na reprodução, na imunidade e na cognição, além de interagir com receptores de mineralocorticoides e outros de glicocorticoides (HAHNER *et al.*, 2021). Por fim, os andrógenos, hormônios sexuais, são majoritariamente secretados na zona reticulada (MESIANO *et al.*, 1997). Os esteroides sintetizados em maior volume em humanos são: cortisol (zona fasciculada), DHEA (zona reticulada) e aldosterona (zona glomerulosa) (SECCIA *et al.*, 2018).

Um complexo enzimático de citocromos P450 tecido-específico é responsável pela conversão do colesterol nas diferentes moléculas de esteroide, como é apresentado na Figura 2 (SECCIA *et al.*, 2018). A síntese de aldosterona a partir do colesterol, por exemplo, envolve cinco enzimas e duas zonas do córtex. As três primeiras enzimas envolvidas na conversão do colesterol em 11-desoxicorticosterona (DOC) são expressas apenas em células das zonas glomerulosa e fasciculada, são elas: enzima de clivagem de cadeia lateral do colesterol (P450scc; SCC), 3 β -hidroxisteroide desidrogenase tipo 2 (3 β -HSD2) e 21-hidroxilase (P450c21; CYP21A2; 21OH). Na sequência, a enzima 11 β -hidroxilase (P450c11; 11 β OH), que converte DOC em corticosterona, é expressa apenas na zona fasciculada. Por último, a enzima aldosterona sintase (P450aldo; Star), que converte corticosterona em aldosterona, tem sua expressão limitada na zona glomerulosa (SECCIA *et al.*, 2018).

Já a enzima 17 α -hidroxilase (P450c17; 17 α OH/17,20-liase) converte duas moléculas, sendo cada uma delas em uma zona diferente. A conversão da pregnenolona em dehidroepiandrosterona (DHEA), precursor do androstenediol, é realizada na zona reticulada. Enquanto isso, a mesma enzima converte progesterona em androstenediona, precursor da testosterona, na zona fasciculada (SECCIA *et al.*, 2018).

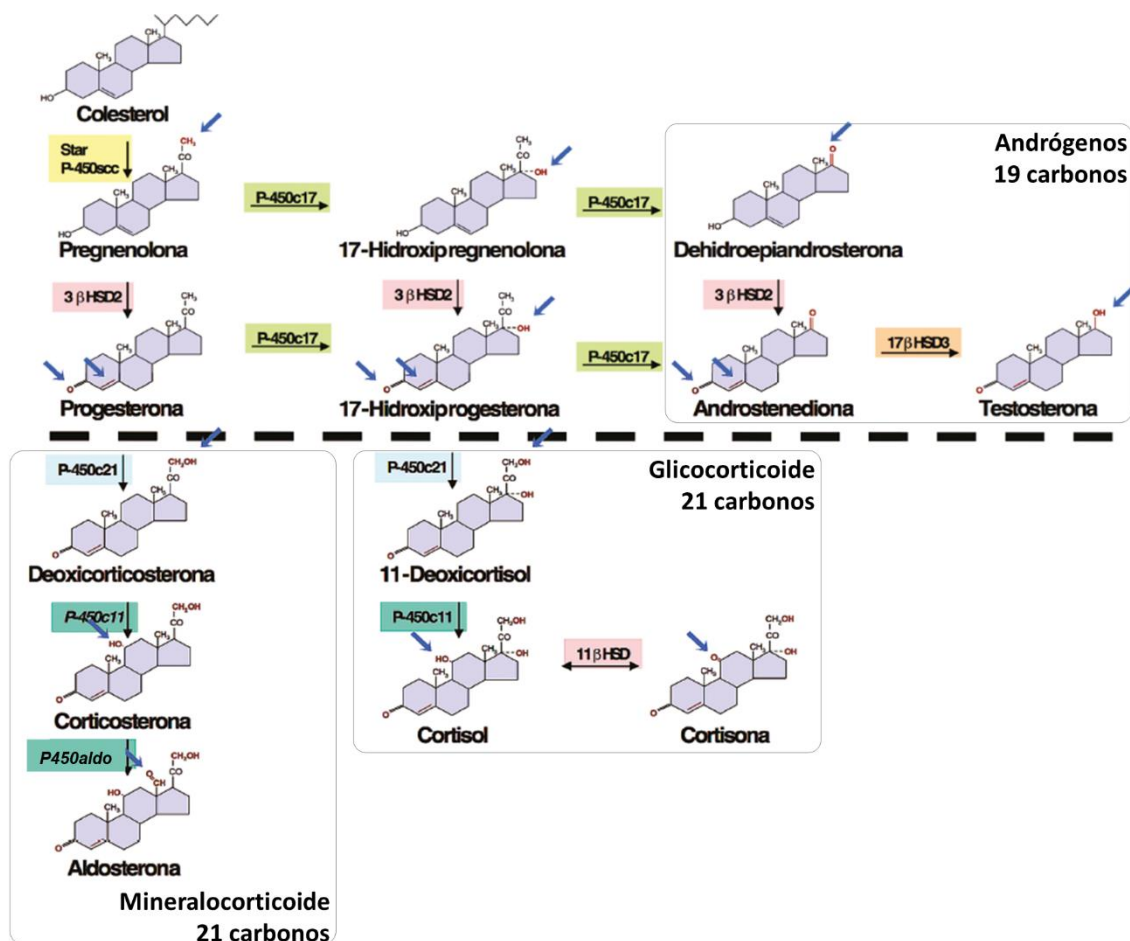


Figura 2. Rota de biossíntese de esteroides humana. A rota inicia com a conversão de colesterol em pregnenolona, que pode acontecer em todos os tecidos esteroidogênicos. Na sequência, já dentro do córtex da adrenal, há a conversão da pregnenolona em duas vias: uma que vai direcionar a síntese de cortisol e aldosterona, e a outra a síntese de hormônios sexuais. Apenas as enzimas P450scc, P450c11 e P450aldo (localizadas no lado esquerdo no esquema) são mitocondriais, o restante é localizado no retículo endoplasmático rugoso. A ausência da atividade da enzima 21-hidroxilase (P450c21) impacta a via de biossíntese de glicocorticoides e mineralocorticoides – a parte da rota bloqueada está abaixo da linha tracejada. Figura adaptada de ALVES et al. (2019).

1.2 A Hiperplasia adrenal congênita

A hiperplasia adrenal congênita (HAC) é um grupo de doenças autossômicas recessivas que impacta a via de biossíntese do cortisol. Esse grupo de doenças acarreta uma superprodução do hormônio liberador de corticotrofina (CRH), no hipotálamo, e do hormônio adrenocorticotrófico (ACTH), na glândula pituitária. Ambos os hormônios induzem a via dos corticosteroides adrenais e são regulados negativamente pelo cortisol (Figura 3) (EI-MAOUCHE *et al.*, 2017; ZHOU *et al.*, 2017). Portanto, além da HAC causar acúmulo de esteroides precursores da enzima com defeito da rota, ela também aumenta a síntese destes gerando hiperplasia da glândula adrenal.

Devido às diferentes desregulações hormonais que podem ser geradas, a HAC apresenta uma grande variabilidade bioquímica, podendo afetar as sínteses de glicocorticoides, mineralocorticoides e andrógenos. Conseqüentemente, são observadas diversas manifestações clínicas, como: insuficiência adrenal, hiperplasia da glândula adrenal, distúrbio no desenvolvimento sexual (DDS), infertilidade, comprometimento do crescimento, hipertensão, síndromes metabólicas na puberdade, entre outras. A gravidade da doença e suas características dependem basicamente de três fatores, a enzima afetada, o grau de dano na sua atividade e o genótipo do paciente (EI-MAOUCHE *et al.*, 2017; ZHOU *et al.*, 2017).

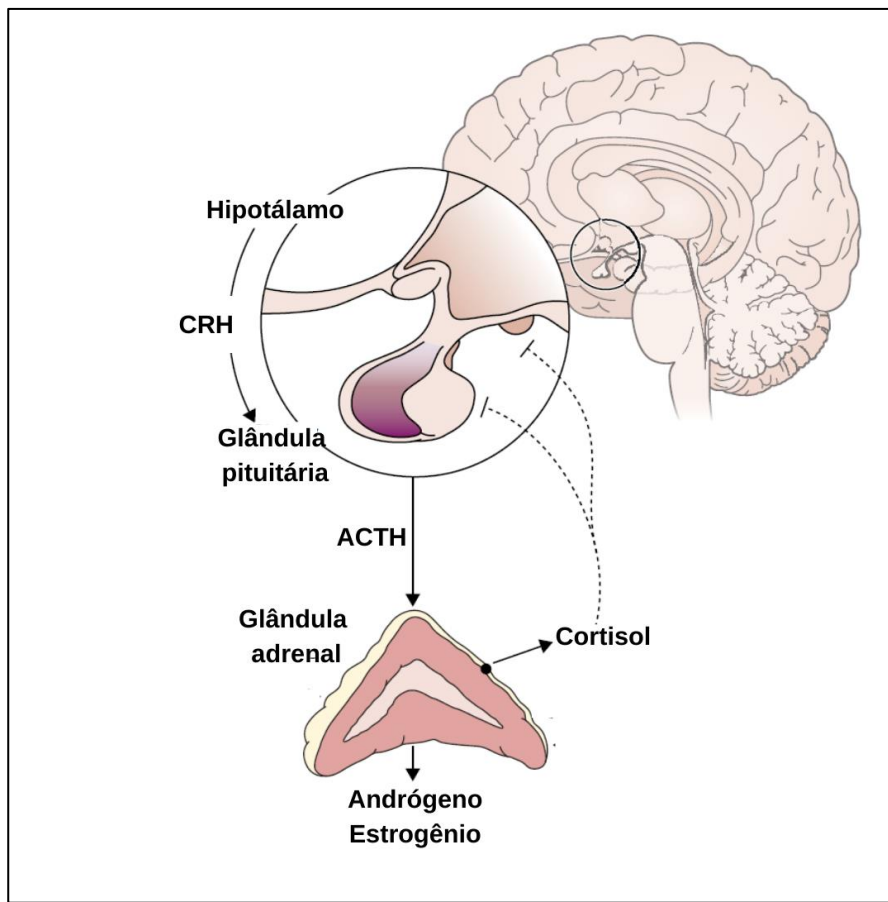


Figura 3. Representação esquemática da via de biossíntese do cortisol. A flecha preta pontilhada mostra a ação do cortisol de regular negativamente sua síntese no hipotálamo. Figura adaptada de EI-MAOUCHE *et al.* (2017).

A HAC devido à deficiência na enzima CYP21A2 representa 90-95% dos casos de HAC (MILLER & AUCHUS, 2011). Sua forma grave ou clássica apresenta uma incidência entre 1:10.000 a 1:20.000 nascidos vivos (n.v.), tendo seu rastreamento populacional realizado por programas nacionais de triagem de recém-nascidos (KOPACEK *et al.* 2017; ALI *et al.*, 2021).

A Figura 4 apresenta os dados reportados por estados brasileiros a partir da implementação do programa nacional brasileiro de triagem neonatal de HAC. Por outro lado, a forma leve ou não clássica (NC) da HAC está presente em ~1:200 n.v. (EI-MAOUCHE *et al.*, 2017; HANNAH-SHMOUNI *et al.*, 2017a; NEW *et al.*, 2017). A HAC-NC é uma das doenças autossômicas recessivas mais comum em humanos, podendo ser ainda mais frequente em certos grupos populacionais (HANNAH-SHMOUNI *et al.*, 2017a; NEW *et al.*, 2017).

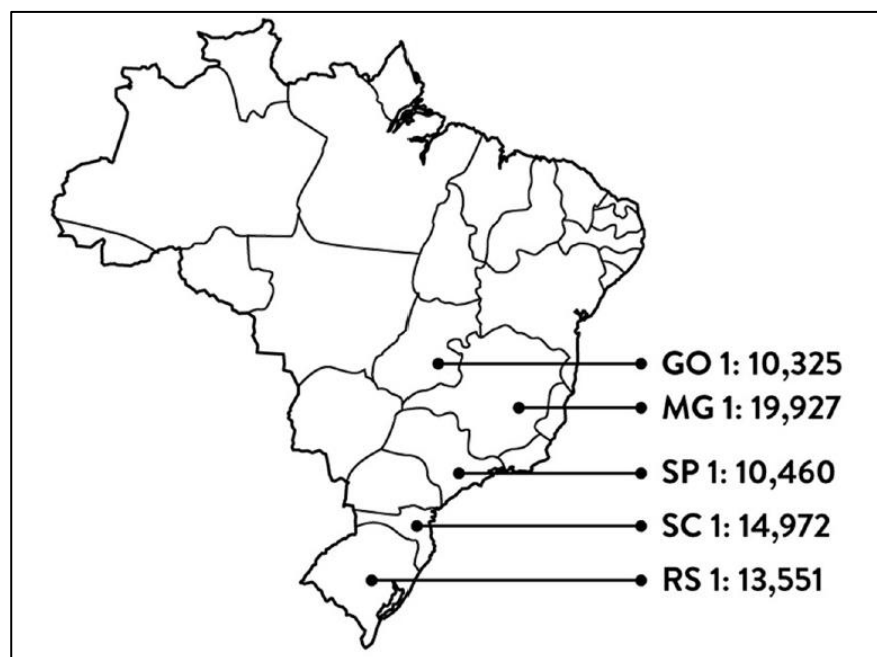


Figura 4. Incidência de hiperplasia adrenal congênita por estado brasileiro. Dados reportados por estados Brasileiros até o ano de 2017. Fonte da figura KOPACEK *et al.* (2017).

Em geral, a segunda enzima do grupo da HAC com deficiência mais prevalente é P450c11, sendo estimada em 5-8% dos casos de HAC. Sua forma clássica é registrada em 1:100.000 n.v. caucasianos e em 1:6.000 n.v. marroquinos e judeus (MENABÒ *et al.*, 2014). Já a deficiência na P450c17 é relatada estar mais presente apenas no Brasil e em uma minoria descendente de Holandeses (HANNAH-SHMOUNI *et al.*, 2017b). No Brasil, a incidência de HAC devido a 17OH é de 1:50.000 n.v. (COSTA-SANTOS *et al.*, 2004). As outras deficiências enzimáticas do grupo de HAC são mais raras e não têm uma estimativa populacional, sendo elas nas enzimas 3 β -HSD2, proteína de regulação aguda da esteroidogênese (StAR), P450scc e P450 oxidorreductase (POR) (EI-MAOUCHE *et al.*, 2017; FLÜCK *et al.*, 2008).

Neste trabalho será abordada a deficiência na enzima CYP21A2, que afeta a biossíntese de glicocorticoide e mineralocorticoides, através do total ou parcial bloqueio dessas vias.

1.3 A enzima 21-hidroxilase (CYP21A2)

A proteína CYP21A2, também conhecida como P450c21, P450 21A2 ou CYP21A2, é membro da superfamília de citocromos P450 (CYPs). Ela apresenta duas isoformas funcionais, uma com 485 e outra com 484 aminoácidos [dado fornecido pelo RefSeq julho 2008]. A diferença entre as isoformas é a ausência do aminoácido 10 na proteína mais curta.

A sequência mais curta tem sido utilizada desde o início das descrições, pois foi baseada no gene de um indivíduo com essa deleção. No entanto, a partir das análises populacionais, foi observado que a proteína mais longa é a mais frequente.

Recentemente, a *Human Genome Variation Society* (HGVS) determinou a substituição da sequência referência para a nomenclatura da CYP21A2 para a da proteína mais longa. Ambas as nomenclaturas coexistem em trabalhos publicados atualmente, mas a proteína longa já está implementada em praticamente todos os bancos de dados (ex., Ensembl e GeneCard). Neste trabalho a nomenclatura adotada foi a da proteína longa (NP_000491.4).

A enzima CYP21A2 tem a função de converter 17-hidroxiprogesterona (17OHP) em 11-deoxicortisol e progesterona em DOC, nas biossínteses de mineralocorticoides e glicocorticoides, respectivamente (Figura 5) (PALLAN *et al.*, 2015). Sua expressão é predominantemente nas células das zonas glomerulosa e fasciculada da glândula adrenal, sob controle do ACTH, onde se encontra ancorada na membrana do retículo endoplasmático (MILLER & AUCHUS, 2011).

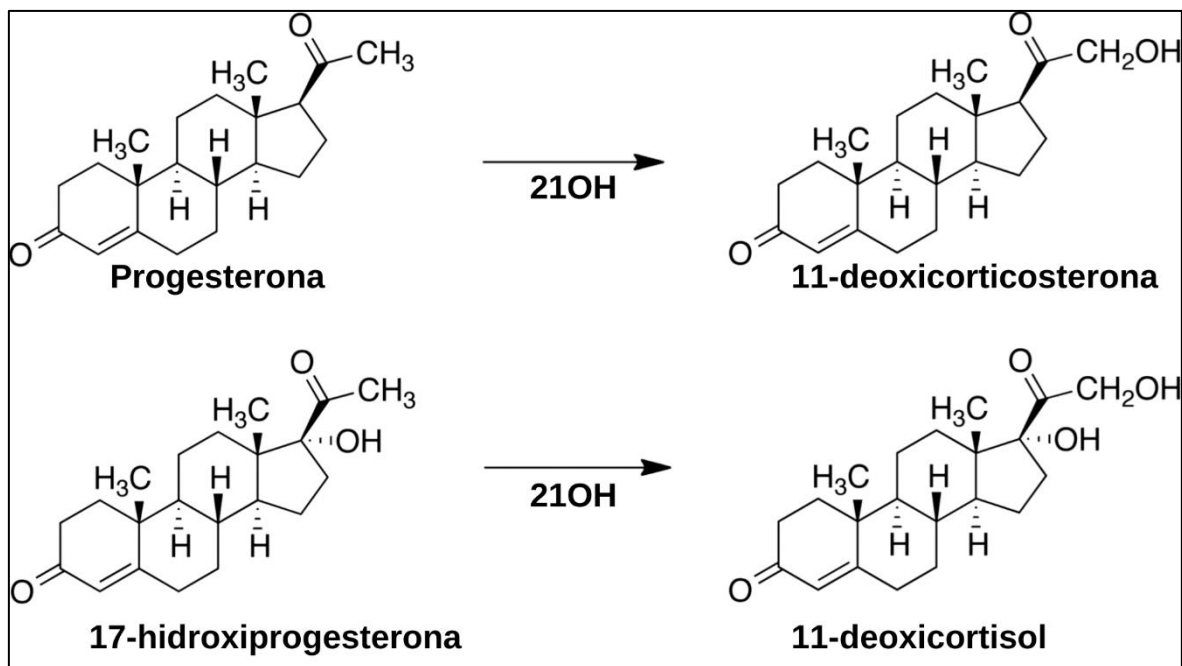


Figura 5. Reações de hidroxilação realizadas pela enzima 21-hidroxilase (CYP21A2) humana no córtex adrenal. Adição de uma hidroxila no carbono 21 das moléculas precursoras. Fonte da figura PALLAN *et al.* (2015).

A reação de hidroxilação da 17OHP é exclusivamente realizada pela CYP21A2, diferentemente da hidroxilação da progesterona, que pode ser realizada por outras enzimas hepáticas, como a CYP2C19 e CYP3A4, que em parte podem ser responsáveis pela menor necessidade de mineralocorticoides adrenais em adultos (GOMES *et al.*, 2009).

A progesterona hidroxilada fora do tecido adrenal foi descrita estar presente em uma ampla gama de tecidos adulto e fetal. Entretanto, pela oxidação da progesterona não ser mediada pela mesma enzima encontrada no córtex adrenal, a CYP21A2, o produto não atua da mesma forma (MILLER & AUCHUS, 2011).

Como exemplo, no fígado humano, a enzima CYP2C19 converte progesterona em 21-hidroxiprogesterona, 16 α - hidroxiprogesteronas e 17 α -hidroxiprogesteronas, além de oxidar também a testosterona em androstenediona (YAMAZAKI & SHIMADA, 1997).

A síntese de glicocorticoides, como o cortisol, também foi observada em outros tecidos, como do intestino, pele, cérebro e pulmão (AHMED *et al.*, 2019). No cérebro humano já foi identificada a síntese de glicocorticoides e mineralocorticoides, onde a presença de RNA mensageiro (RNAm) referentes a CYP21A2 foi muito baixo ou abaixo do limite de detecção. Isso provou que a mesma hidroxilação da CYP21A2 é realizada por uma enzima alternativa no cérebro, a CYP2D6 (TAVES *et al.*, 2011). Esses fatos, justificam a presença de concentrações razoáveis de esteroides hidroxilados semelhante aos produtos da CYP21A2 no plasma de pacientes com ausência de atividade da CYP21A2 (MILLER & AUCHUS, 2011).

A deficiência na enzima CYP21A2, além de ser a forma prevalente de HAC, é um dos erros inatos do metabolismo mais frequentes (HANNAH-SHMOUNI *et al.*, 2017a; NEW *et al.*, 2017). A perda completa da atividade da CYP21A2 impacta a síntese de mineralocorticoides resultando na deficiência de aldosterona, bem como, a característica síntese de cortisol da via glicolítica.

A aldosterona é um importante regulador eletrolítico com ação na reabsorção do sódio pelos túbulos renais distais (NEW *et al.*, 2017). Sua ausência causa grave hiponatremia, hipercalemia e acidose, que podem gerar um quadro clínico de hipotensão, choque e colapso cardiovascular, com evolução à óbito do recém-

nascido não tratado. Ademais, crianças com perda de sal renal apresentam pouco apetite, perda de peso, vômito, desidratação e hipotensão (MILLER & AUCHUS, 2011; NEW *et al.*, 2017). O esquema da Figura 6 resume a fisiologia acarretada de uma crise de perda de sal adrenal.

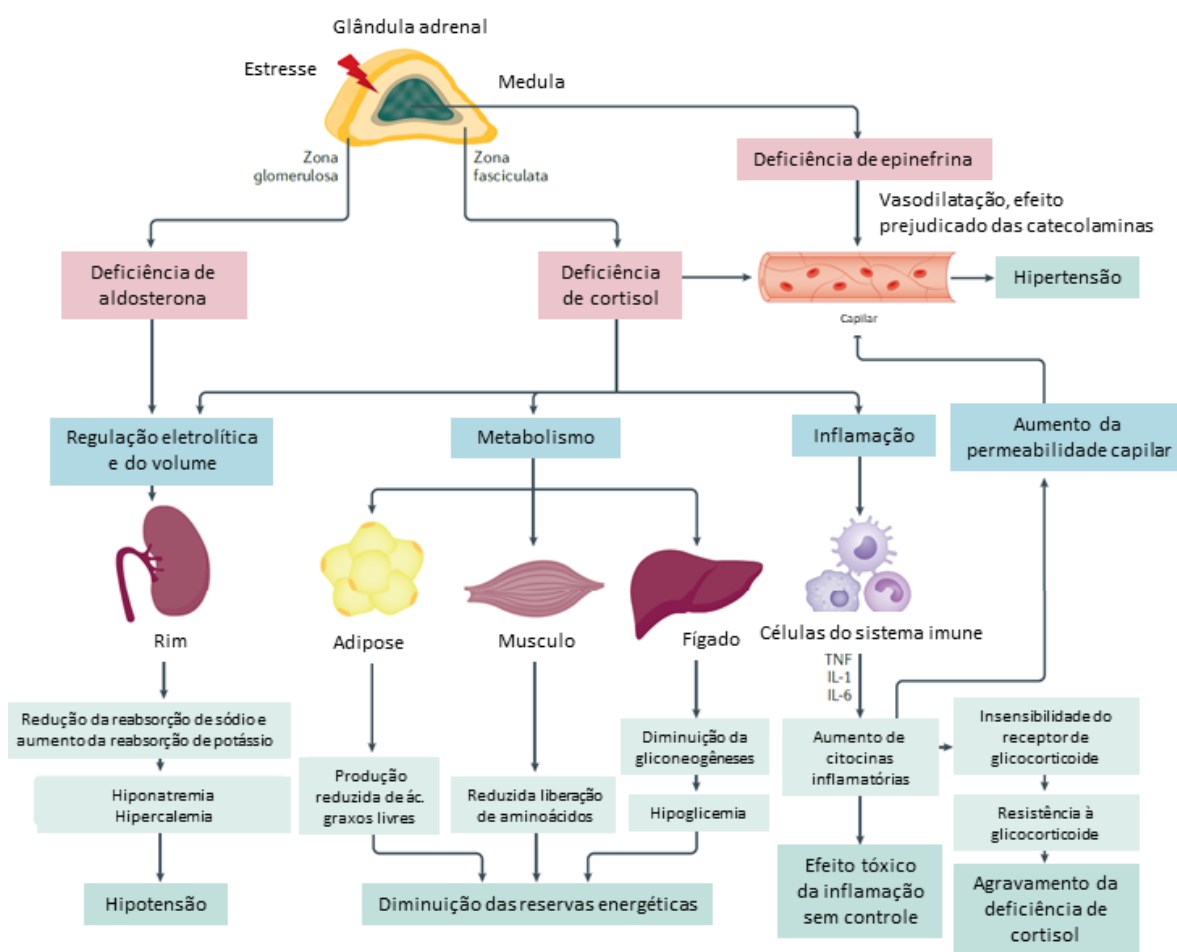


Figura 6. Esquema da fisiologia de uma crise de perda de sal adrenal devido a completa perda da atividade da enzima 21-hidroxilase. O desequilíbrio entre a demanda e a disponibilidade de glicocorticoides afeta os sistemas de regulação eletrolítico, metabólico, inflamatório e circulatório. Figura adaptada de HAHNER *et al.* (2021).

Por sua vez, o cortisol atua em diferentes vias (Figura 6). Esse esteroide é necessário em altas concentrações na medula adrenal para a transcrição do gene que codifica a feniletanolamina N-metiltransferase. Essa enzima é responsável por catalisar a conversão de norepinefrina em epinefrina, conhecidas como noradrenalina e adrenalina, respectivamente. Uma baixa concentração de epinefrina aumenta o risco de hipoglicemia associado à deficiência do cortisol (TUTUNCULER *et al.*, 2009). O déficit do nível de epinefrina está diretamente associado ao grau de disfunção adrenocortical, não sendo observado na forma leve de HAC (VERMA *et al.*, 2010). Por outro lado, pacientes com a forma grave de HAC apresentam aumento da secreção de norepinefrina de forma compensatória ao déficit de epinefrina (TUTUNCULER *et al.*, 2009).

Ademais, devido à síntese do cortisol ser regulada negativamente pelos CRH e ACTH na sua presença, a baixa concentração de cortisol causa o estímulo da secreção de ambos os hormônios, desde o período fetal (Figura 6). Como consequência, há a indução da hiperplasia adrenal e transcrição dos genes envolvidos na esteroidogênese, gerando acúmulo dos substratos da enzima CYP21A2 (progesterona e 17OHP) (MILLER & AUCHUS, 2011). O esteroide 17OHP convertido pela enzima CYP21A2 em 11-deoxicortisol (precursor do cortisol) também é convertido, na mesma etapa da via de biossíntese de esteroides, pela enzima P450c17 em androstenediona (precursora da testosterona/dihidrotestosterona e do estriol). Sendo assim, o esteroide 17OHP acumulado na rota devido a deficiência da CYP21A2 acaba, por conseguinte, sendo direcionado para a síntese do andrógeno testosterona (CHOI *et al.*, 2016).

Naturalmente, a diferenciação da genitália masculina mediada pelos andrógenos sintetizados no testículo do feto ocorre no primeiro trimestre gestacional. Da mesma forma, também é nesse período que o córtex adrenal apresenta um rápido desenvolvimento, com um pico na síntese de cortisol na oitava semana da gestação (GOTO *et al.*, 2006). Portanto, no primeiro trimestre de gestação o feto feminino, 46,XX, torna-se mais suscetível à virilização provocada pelo aumento dos níveis de andrógenos adrenais e tornam essa etapa a mais crítica para o dimorfismo sexual do feto feminino (GOTO *et al.*, 2006).

A virilização da genitália externa é categorizada em cinco graus de acordo com a escala de Prader, descrita na Figura 7, que tem um escore entre I e V (PRADER, 1954; JORGE *et al.*, 2008).

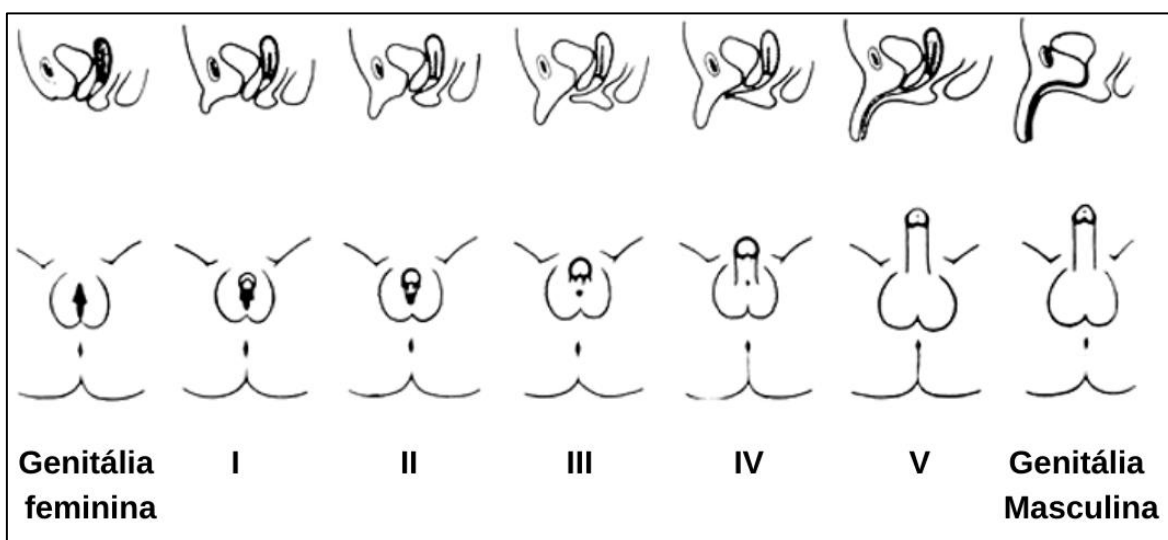


Figura 7. Escala Prader (I a V) para a classificação de genitálias atípicas. À esquerda está representada a genitália feminina típica e à direita a masculina, entre elas está representada a escala Prader. Tipo I, clitomegalia sem fusão labial; tipo

II, clitomegalia com fusão labial posterior; tipo III, alto grau de clitomegalia, ausência de orifício vaginal; e fusão labial quase completa; tipo IV, clitóris mais flácido, seio urogenital semelhante a uretra e completa fusão labial; tipo V, clitóris peniano, canal uretral finalizando na ponta do falo e lábios semelhantes ao escroto. Fonte da figura YAU & NEW (2019).

Por outro lado, os órgãos genitais internos femininos têm o seu aspecto normal preservado, pois não são influenciados pelos andrógenos. A genitália interna é regida pelos ductos de Müller, responsáveis pelo desenvolvimento do útero, tubos uterinos, cérvix e a parte interna da genitália feminina. Apenas na presença do hormônio anti-Mülleriano cuja síntese ocorre nos testículos fetais a ação dos ductos de Müller é inibida (WILSON & BORDONI, 2021).

Quanto aos recém-nascidos do sexo masculino, 46,XY, o montante sintetizado no córtex adrenal devido à deficiência da CYP21A2 não é significativo em relação à abundância normal de testosterona testicular (MILLER&AUCHUS, 2011; CHOI *et al.*, 2016).

Outra característica dos portadores de HAC, tanto nas formas clássica quanto NC, é a de crianças altas que se tornam adultos baixos. Isso se deve à aceleração do crescimento somático, gerando um rápido ganho de estatura durante a infância, acarretando prematuro fechamento das esfinges. Por conseguinte, há o comprometimento da altura adulta final, que é em torno de 10 cm menor do que a estimada pela altura parental (LIN-SU *et al.*, 2011; NEW *et al.*, 2017). Esse mesmo quadro também foi relatado em pacientes que receberam excesso de glicocorticoides no tratamento (LIN-SU *et al.*, 2011). No entanto, em pacientes sob

tratamento de HAC pode ser realizada a administração de hormônios reguladores do crescimento para auxiliar a reduzir o comprometimento do crescimento (LIN-SU *et al.*, 2011; NEW *et al.*, 2017).

1.3.1 *Categorias fenotípicas*

A alta variabilidade nos níveis hormonais é responsável pelo grande espectro de fenótipos visto na deficiência da CYP21A2, que variam de acordo com o grau de comprometimento da enzima CYP21A2. Os casos graves ou clássicos, apresentam deficiência de glicocorticoides e podem apresentar ou não deficiência na síntese de mineralocorticoide. A forma clássica de HAC que apresenta deficiência em ambas as vias é denominada perdedora de sal (HAC-PS), já a que tem o nível de aldosterona suficiente para regular o balanço eletrolítico, é denominada virilizante simples (HAC-VS). A forma leve da HAC é denominada NC (HAC-NC), sendo distinguida pela ausência da perda de sal e da virilização.

A forma clássica HAC-PS é caracterizada pela ausência da atividade da enzima CYP21A2 (< 1% em relação a selvagem), comprometendo a síntese de glicocorticoides e mineralocorticoides (NEW *et al.*, 2013; XU *et al.*, 2019). Nesse caso, não há aldosterona suficiente para haver a regulação eletrolítica através da absorção de sódio na adrenal. Crises de perda de sal renal devido à deficiência de aldosterona são observadas entre o 5º e o 15º dia de vida de vida do recém-nascido (MILLER & AUCHUS, 2011). A prevenção da crise adrenal em recém-nascidos do sexo feminino (46,XX) pode ser mais facilmente realizada com a identificação precoce da patologia através da avaliação visual da genitália externa e então prosseguimento do tratamento mais adequado. Por outro lado, recém-nascidos do

sexo masculino (46,XY) apresentam sutis alterações físicas que podem apenas sugerir a deficiência do cortisol, como hiperpigmentação e grande falo com pequeno testículo (MILLER & AUCHUS, 2011). Dessa forma, meninos com HAC-PS são identificados pela triagem pré-natal, pela triagem de recém-nascidos ou devido a crises de perda de sal presenciadas em casa nas primeiras semanas de vida (NEW *et al.*, 2017).

A forma clássica HAC-VS é caracterizada pela ausência da perda de sal renal, mas com elevada concentração dos substratos da CYP21A2. Bioquimicamente, o que distingue a forma HAC-PS da -VS é a atividade da enzima CYP21A2. Na HAC-VS, a enzima CYP21A2 tem uma atividade residual entre 1 e 10% em relação à selvagem, sendo esta suficiente para a síntese de aldosterona e prevenção significativa da perda de sal renal (NEW *et al.*, 2013; SIMONETTI *et al.*, 2018). Meninas (46,XX) apresentam virilização da genitália externa semelhante à forma HAC-PS (Figura 8) que, se não iniciado o tratamento com glicocorticoides logo após o nascimento, apresentam uma progressiva masculinização da genitália pós-natal devido ao contínuo excesso de andrógenos adrenais (NEW *et al.*, 2017). De outra forma, a maioria dos meninos recém-nascidos (46,XY) com a forma HAC-SV passam despercebidos pela avaliação clínica, sendo identificados na triagem de recém-nascidos ou pré-natal. Se não tratados, crianças de ambos os sexos apresentam sinais de hiperandrogenismo, como o início precoce de pelos nas axilas, faciais e pubianos, odor corporal de adulto, avanço da idade óssea levando ao fechamento precoce das esfinges e baixa estatura.

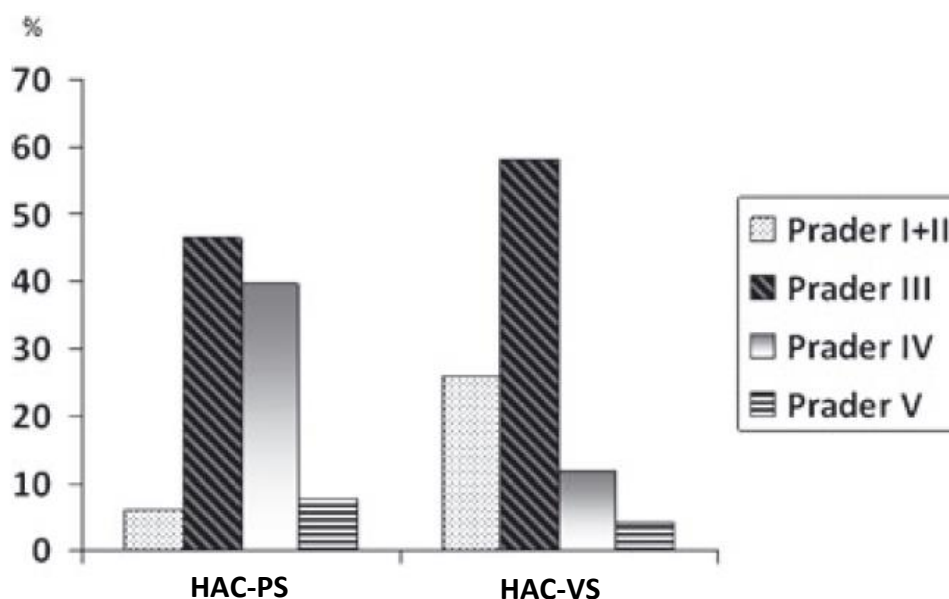


Figura 8. Frequência da escala Prader obtida na análise de 154 indivíduos 46,XX. Genótipos associados à atividade da 21-OH < 2% em relação a selvagem (HAC-PS) apresentaram predomínio da virilização Prader III, seguido da IV. Genótipos associados à atividade enzimática entre 3 e 7% em relação à selvagem (HAC-VS) mostraram predomínio da virilização Prader III, seguido pela II. Fonte da figura KAUPERT *et al.* (2013).

Por fim, a HAC-NC é conhecida pelo início tardio dos sintomas em relação às formas clássicas, sendo estes moderados ou leves. O comprometimento enzimático da CYP21A2 na HAC-NC permite uma atividade parcial de hidroxilação, que varia entre 10 e 78% em relação à enzima selvagem, permitindo uma concentração sérica de cortisol quase normal (NEW *et al.*, 2017; SIMONETTI *et al.*, 2018). Essa grande faixa de atividade da enzima está associada a sintomas de hiperandrogenismo que podem aparecer logo após o nascimento, sendo

amenizadas com o tempo ou até não serem notadas em indivíduos assintomáticos (NEW *et al.*, 2017). Os sinais de hiperandrogenismo na forma HAC-NC são mais amenos do que nas formas clássicas não tratadas. Entre os principais sintomas, em ambos os sexos, estão o início precoce de pelos pubianos, avanço da idade óssea levando à baixa estatura e acne cística grave (HANNAH-SHMOUNI *et al.*, 2017a). Além disso, mulheres adultas podem apresentar outros sintomas atípicos a indivíduos 46, XX causados pelo excesso de andrógenos, como hirsutismo, calvície temporal, infertilidade, menarca tardia e amenorreia secundária. Sintomas semelhantes à síndrome do ovário policístico (*like-SOP*) também pode ocorrer devido aos elevados níveis de andrógenos. Duas possibilidades para a *like-SOP* devido a HAC são: (I) a interrupção da liberação de gonadotrofina e com consequente formação de cistos e (II) reprogramação do eixo hipotálamo – hipófise - gônada no período pré e pós-natal (NEW *et al.* 2017; HANNAH-SHMOUNI *et al.*, 2017a). Quanto aos sintomas da HAC-NC em homens adultos estão calvície precoce, baixa fertilidade, oligozoospermia e testículos pequenos em relação ao tamanho do falo. Essa última característica é relacionada a supressão do eixo hipotálamo- hipófise – gônada pelos andrógenos adrenais (NEW *et al.*, 2017; HANNAH-SHMOUNI *et al.*, 2017a).

1.4 O gene codificador da proteína CYP21A2

A proteína CYP21A2 é codificada pelo gene *CYP21A2* (*Cytochrome P450 Family 21 Subfamily A Member 2*; GeneID 1589; NG_007941.3), também conhecido como *CYP21*, *CYP21B*, *P450c21B*. Esse gene possui 3,44 kb organizados em 10 exons e 9 introns situado no braço curto do cromossomo 6,

dentro do complexo principal de histocompatibilidade (MHC – *major histocompatibility complex*) (Figura 9) (RODRIGUES *et al.*, 1987).

Além disso, o gene *CYP21A2* faz parte da composição do módulo RCCX, uma região que apresenta uma organização complexa, com uma grande variabilidade de tamanho de genes e número de cópias (RODRIGUES *et al.*, 1987). Esse módulo engloba os genes RPs (*RP1* e *RP2*), C4s (*C4A* e *C4B*), CYP21s (*CYP21A2* e *CYP21A1P*) e TNXs (*TNXA* e *TNXB*) em uma organização tandem bimodular (*RP1-C4A-CYP21A1P-TNXA-RP2-C4B-CYP21A2-TNXB*). Os genes *RP2*, *CYP21A1P* e *TNXA* não são funcionais (pseudogenes) (Figura 9). Os dois genes *C4* se distinguem por apenas 4 aminoácidos e apresentam alta variabilidade no número de cópias (LI *et al.*, 2017). Os genes RPs compartilham uma identidade na sequência de 48%. Os genes TNX têm a orientação transcricional oposta aos demais, sendo o seu 3'-UTR sobreposto ao 3'-UTR e exon 10 do gene *CYP21A2* (MILLER & MERCKE, 2018). Por fim, os CYP21s compartilham um alto nível de identidade na sequência, 98% nos exons e 96% nos introns, e estão distantes por aproximadamente 30 kb (RODRIGUES *et al.*, 1987). No entanto, o *CYP21A1P* tornou-se inativo devido a presença de múltiplas mutações que rangem de substituição pontual de nucleotídeo a pequenas deleções e inserções (RODRIGUES *et al.*, 1987).

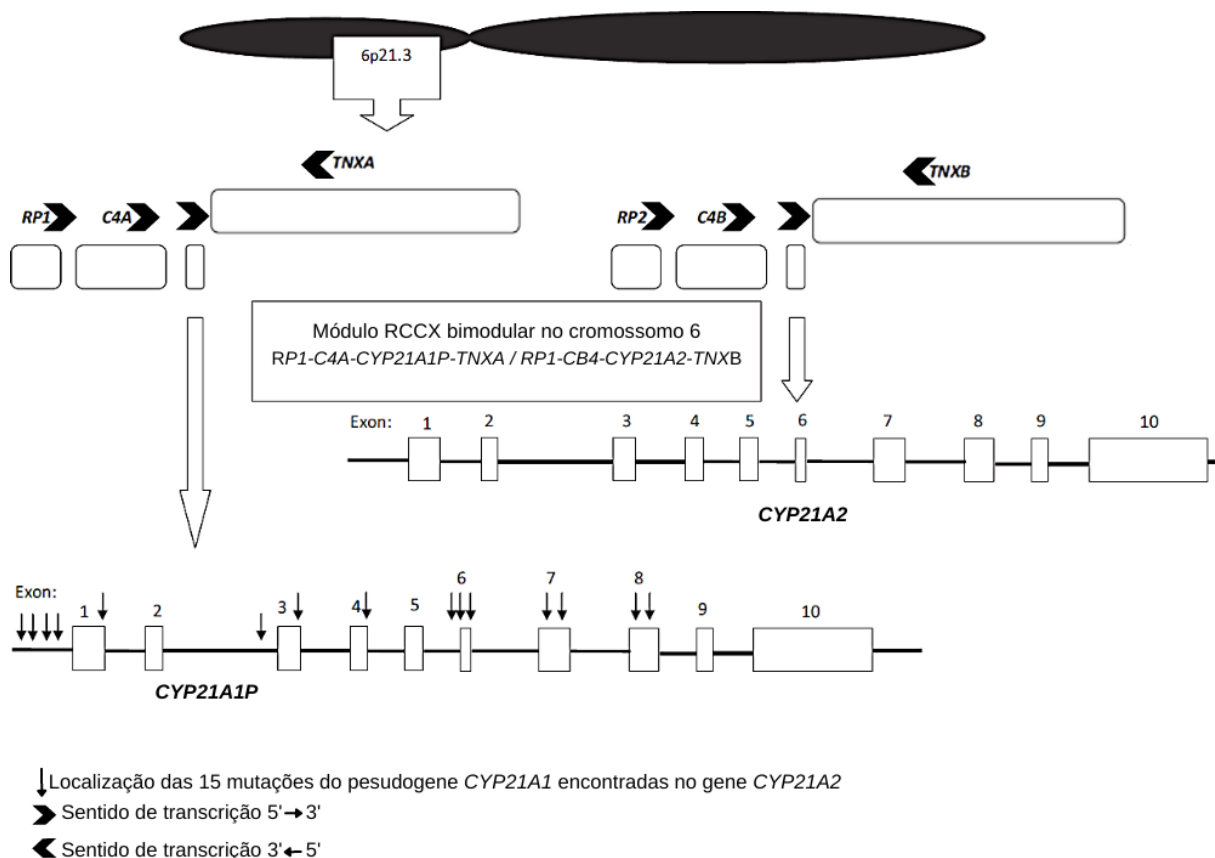


Figura 9. Esquema da composição do módulo RCCX no cromossomo 6. O módulo grande (à esquerda) é composto pelos genes RP1-C4A-CYP21A1P-TNXA. O módulo pequeno (à direita) é composto por RP2-C4B-CYP21A2-TNXB. Os genes RPs, C4s e CYPs estão orientados no sentido 5'→3', sentido telômero para o centrômero. Os genes TNXs estão orientados no sentido inverso, 3'←5', do centrômero para o telômero. Figura adaptada de NARASIMHAN & KATTAB (2019).

A proximidade entre os genes CYP21s, a localização em um módulo com quatro genes organizados em tandem e a alta homologia entre eles geram a alta taxa de eventos de recombinação gênica observados no gene *CPY21A2*. Esses eventos podem ocorrer durante a duplicação cromossômica, causando grandes

perdas gênicas, ou em eventos de *crossing-over* desigual, causando pequenas trocas nucleotídicas (TUSIÉ-LUNA & WHITE, 1995). Por fim, recombinações entre os genes CYP21s são responsáveis por ~95% das mutações patogênicas conhecidas do gene CYP21A2 (Figura 9).

Em toda a extensão do gene CYP21A2 já foram descritas mutações patogênicas, desde a região 5'-UTR até a 3'-UTR. No entanto, dez dessas mutações representam de 80-90% das mutações patogênicas em pacientes com HAC e são provenientes do pseudogene (Figura 10) (NEW *et al.*, 2013; CARVALHO *et al.*, 2016; XU *et al.*, 2019). Cabe ressaltar que há variabilidade nas frequências dessas mutações em diferentes grupos étnicos e regiões (Tabela 1), como mostrado por vários grupos (NEW *et al.*, 2013; PRADO *et al.*, 2017; RIEDL *et al.*, 2019). Entre as mutações conhecidas provenientes do pseudogene CYP21A1P estão (Figura 9 e Figura 10):

- Região promotora: g.-103A> G, g.-110T> C, g.-113G> A e g.-126C> T - associadas à redução de 20% da atividade transcricional do gene CYP21A2 (NARASIMHAN & KATTAB, 2019).
- Exon 1: p.P31L - mutação leve associada a HAC-NC com uma atividade residual da CYP21A2 de 30-60%, quando representa um evento de microconversão envolvendo o exon 1 do pseudogene. No entanto, quando provém de um evento de microconversão envolvendo a região promotora + exon 1 do pseudogene, as mutações citadas no item anterior são incorporadas, reduzindo o nível transcricional do gene e resultando no fenótipo HAC-SV (TUSIÉ-LUNA *et al.*, 1991; ARAUJO *et al.*, 2005).

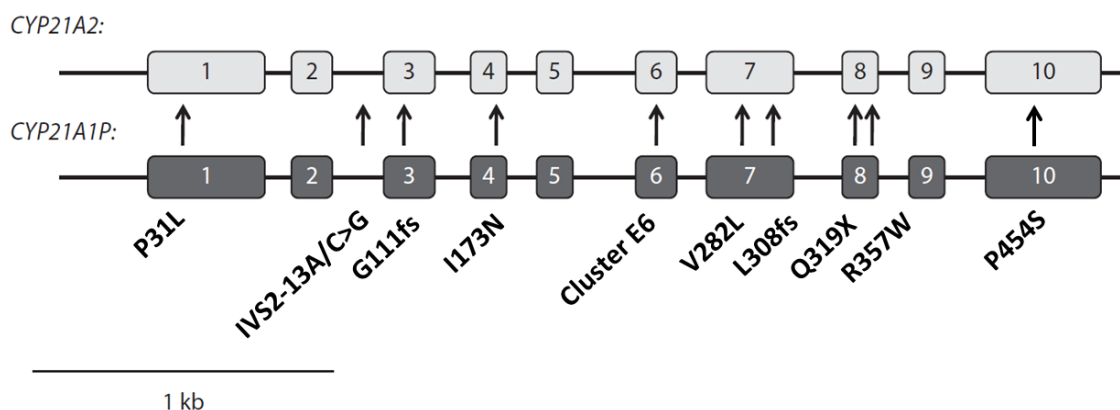


Figura 10. Esquema das mutações patogênicas mais comuns encontradas no gene *CYP21A2* são provenientes de eventos de recombinação gênica com o pseudogene *CYP21A1P*. Figura adaptada de WEDELL (2011).

- Íntron 2: IVS2-13C/A> G (também conhecida como I2G) – mutação severa de um dos sítios *aceptores* de *splice* associada a HAC-PS com uma atividade mínima da *CYP21A2* (0-1%) (HIGASHI *et al.*, 1988).
- Exon 3: p.G111fs (também conhecida como Δ 8pb no exon 3) - mutação do tipo *frameshift* que causa a deleção de oito pares de bases e a um códon de terminação prematura associada a HAC-PS com perda completa da atividade da *CYP21A2* (HAIDER *et al.*, 2013).
- Exon 4: p.I173N - mutação grave que altera a estrutura da proteína *CYP21A2*. Está associada a HAC-SV devido a redução da atividade da *CYP21A2*, entre 1-2% em relação à selvagem (HAIDER *et al.*, 2013).
- Exon 6: Cluster E6 [p.I237N, p.V238E e p.M240K] – grupo de mutações que normalmente ocorrem juntas e está associado a HAC-PS devido a

um dano relacionado à ligação da CYP21A2 ao substrato que resulta na completa perda de atividade da enzima (HAIDER *et al.*, 2013).

- Exon 7: p.V282L – mutação leve associada a HAC-NC com uma atividade parcial da CYP21A2, entre 20 e 50% da selvagem (SPEISER *et al.*, 1992).
p.L308fs, também conhecida como L308+T- mutação do tipo *frameshift* que insere uma base única, gerando um códon que causa a terminação prematura e a perda completa da atividade da CYP21A2, sendo associada a HAC-PS (HAIDER *et al.*, 2013).
- Exon 8: p.Q319X e p.R357W – ambas mutações estão associadas a HAC-PS devido à ruptura da ligação H em um módulo relacionado com a estabilidade da ligação heme com a CYP21A2, sendo associada à perda completa da atividade da CYP21A2 (HAIDER *et al.*, 2013).
- Exon 10: p.P454S – mutação leve com uma atividade parcial da CYP21A2 entre 36 e 44% em relação a selvagem, associada a HAC-NC (RIEPE *et al.*, 2008).

Além dessas mutações provenientes de eventos de *crossing-over* desiguais, as deleções chamadas de 30-kb del são provenientes de quimeras entre CYP21A2/CYP21A1P e causam perda total da atividade da CYP21A2. Em torno de 20% dessas grandes deleções, uma parte do gene TNXB é deletada em conjunto com a remoção do exon 10 do gene *CPY21A2*. Essa mutação está relacionada à doença genética denominada de X-CAH (MILLER & MERCKE, 2018; LAO *et al.*, 2020).

Tabela 1. Comparação entre as mutações mais comuns encontradas no gene *CYP21A2* em diferentes populações.Fonte da tabela Prado *et al.* (2017).

Country (state or region)	No of alleles	Number of mutant alleles (frequency %)					IVS2-13A/C > G	p.Leu308PhefsTer6	Cluster E6	p.Val282Leu*	p.Ile173Asn	<i>CYP21A2</i> (LD, LGC)
		p.Pro31Leu*	p.Gln319Ter	p.Arg357Trp	p.Pro454Ser	p.Gly111ValfsTer21						
Brazil (RS) present study	96	0	11 (11.5)	7 (7.3)	2 (2.1)	1 (1.0)	20 (20.8)	2 (2.1)	3 (3.1)	26 (27.1)	3 (3.1)	9 (9.4)
Brazil (SP + NE) [34]	856	5 (0.6)	52 (6.1)	46 (5.4)	12 (1.4)	15 (1.8)	181 (21.1)	19 (2.2)	10 (1.2)	228 (26.6)	64 (7.5)	77 (9.0)
Brazil (GO) [35]	56	0	9 (16.1)	7 (12.5)	0	0	16 (28.6)	2 (3.6)	0	5 (8.9)	6 (10.7)	9 (16.1)
Brazil (AR) [33]	92	6 (6.5)	11 (12.0)	1 (1.1)	–	1 (1.1)	33 (35.9)	8 (8.7)	6 (6.5)	11 (12.0)	2 (2.2)	–
Argentina [30]	866	6 (0.7)	58 (6.7)	36 (4.2)	12 (1.4)	7 (0.8)	178 (20.6)	–	17 (2.0)	227 (26.2)	71 (8.2)	97 (11.2)
Chile [31, 32]	164	0	15 (9.1)	16 (9.8)	–	–	26 (15.9)	1 (0.6)	3 (1.8)	4 (2.4)	21 (12.8)	32 (19.5)
Mexico [36]	94	8 (8.5)	4 (4.2)	7 (7.4)	2 (2.1)	2 (2.1)	45 (47.9)	1 (1.1)	0	8 (8.5)	11 (11.7)	1 (1.0)
USA [37]	3005	78 (2.6)	105 (3.5)	108 (3.6)	0	63 (2.1)	688 (22.9)	0	63 (2.1)	718 (23.9)	246 (8.2)	601 (20.0)
Portugal [38]	112	2 (1.8)	7 (6.3)	2 (1.8)	2 (1.6)	3 (2.7)	11 (9.8)	5 (4.4)	0	29 (25.9)	11 (9.8)	29 (25.9)
Italy [39]	106	0	5 (4.7)	6 (5.7)	0	1 (0.9)	23 (21.7)	–	–	34 (32.1)	14 (13.2)	3 (2.8)
Spain [40]	354	–	33 (9.3)	0	0	12 (3.3)	62 (17.5)	5 (1.4)	–	120 (33.9)	13 (3.7)	68 (19.2)
Denmark [41]	136	3 (2.2)	12 (8.8)	3 (2.2)	0	0	46 (33.8)	1 (0.7)	2 (1.5)	6 (4.4)	14 (10.3)	49 (36.0)
Sweden [42]	186	3 (1.6)	6 (3.2)	8 (4.3)	1 (0.5)	2 (1.2)	56 (30.1)	1 (0.5)	2 (1.1)	13 (7.0)	38 (20.4)	55 (29.6)
Middle Europe [43]	864	32 (3.7)	22 (2.5)	21 (2.4)	6 (0.7)	9 (1.0)	270 (31.2)	14 (1.6)	3 (0.3)	29 (3.3)	125 (14.5)	264 (30.5)
Germany [44]	310	8 (2.6)	15 (4.8)	14 (4.5)	1 (0.3)	5 (1.6)	94 (30.3)	1 (0.3)	3 (1.0)	9 (2.9)	61 (19.7)	85 (27.4)
Croatia [13]	186	11 (5.9)	9 (4.8)	31 (16.7)	–	4 (2.2)	65 (34.9)	–	4 (2.2)	0	21 (11.3)	–
China [45]	460	1 (0.2)	21 (4.6)	27 (5.9)	–	20 (4.3)	161 (35)	8 (1.7)	6 (1.3)	1 (0.2)	66 (14.3)	90 (19.6)
Greek-Cypriot [46]	64	2 (3.1)	2 (3.1)	–	–	–	14 (21.9)	1 (1.6)	–	36 (56.2)	2 (3.1)	5 (7.8)
Average (%)		2.4	6.7	5.6	0.8	1.6	26.7	2.0	1.6	16.8	10.3	17.8
SD		2.5	3.7	4.3	0.8	1.1	9.1	2.2	1.6	15.4	5.3	10.3

AR Amazon region (Northern region of Brazil), GO Goiás (Central Western region of Brazil), LD large *CYP21A2* gene deletion, LGC large gene conversion, RS Rio Grande do Sul state (Southern Brazil region), SD standard deviation, SP + NE São Paulo (Southeastern region of Brazil) and Northeastern region of Brazil, – not studied

* The statistical analysis performed for frequencies found in South American studies showed a significant difference ($p < 0.001$)

1.5 Relação genótipo e fenótipo

De modo geral, a deficiência da CYP21A2 apresenta uma boa relação entre o genótipo e o fenótipo, mostrando-se mais efetiva para as formas HAC-PS e -NC (NEW et al, 2013). No trabalho descrito em SPEISER *et al.* (1992) foi realizado um dos estudos pioneiros relacionando grupos de mutações aos fenótipos de pacientes com a CYP21A2, mostrando haver uma significativa correlação dos grupos de mutações com os níveis basais de 17OHP no soro dos pacientes, comprovando *in vivo* essa relação. Sua categorização de mutações de acordo com a atividade remanescente da CYP21A2 e, então, associação entre as três formas de HAC é utilizado até hoje. A Figura 11 apresenta um resumo esquemático da correlação entre os grupos de mutações com a respectiva atividade da CYP21A2 e o fenótipo esperado. O estudo de WEDELL *et al.* (1994), com uma amostra maior de pacientes com HAC, ampliou o número de combinações entre mutações dos quatro grupos e acrescentou um grupo para as mutações de efeito não conhecido, o Grupo D.

O Grupo de Mutações Nulo compreende mutações que comprometem totalmente a atividade da CYP21A2 como deleções, rearranjos e conversões gênicas, mutações que causam códon de parada prematura e aquelas em resíduos chaves para a estrutura da proteína e/ou ligação da CYP21A2 com o substrato ou grupamento heme (HAIDER *et al.*, 2013). As mutações mais comuns deste grupo são: 30-kb del, p.G111fs, Cluster E6, p.L308fs, p.Q319X e p.R357W (Figura 11) (KRONE *et al.*, 2013; NEW *et al.*, 2013). Já o Genótipo Grupo Nulo é dito quando ambos os alelos apresentam mutações do grupo nulo (nulo/nulo), sendo estas em homozigose ou heterozigose composta com outra mutação do mesmo grupo. O genótipo nulo é relacionado ao fenótipo HAC-PS com uma correlação de acerto de

99.6% (n=236) segundo dados extraídos do estudo realizado com mais de 1.500 indivíduos de diferentes etnias em NEW *et al.* (2013).

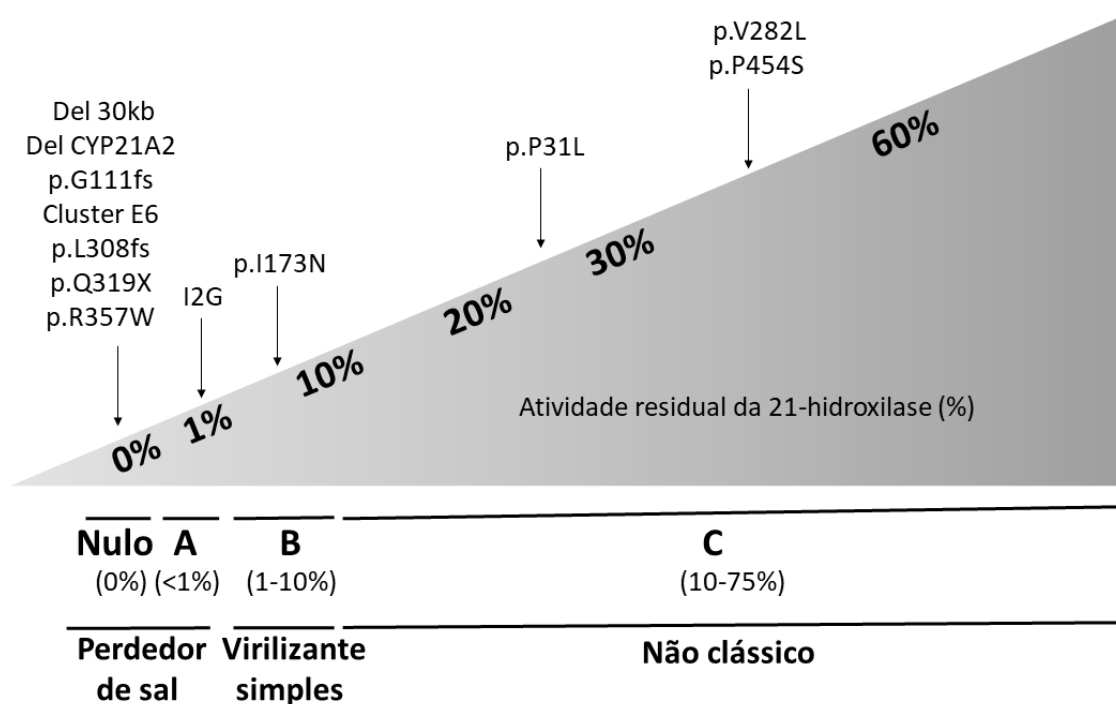


Figura 11. Correlação de genótipo e fenótipo. Relação entre os grupos de mutações (Nulo, A, B e C) e a atividade da enzima 21-hidroxilase. O fenótipo esperado é referente ao alelo com a mutação mais leve (menos comprometido), uma vez que essa deficiência enzimática é recessiva.

O Grupo de Mutações A corresponde à mutação no íntron 2 IVS2-13A/C>G que permite uma mínima atividade da CYP21A2, entre 0-1% em relação à atividade normal (Figura 11). Essa mutação no final do íntron 2 cria um sítio acceptor de *splice* adicional, causando a retenção de 19 nucleotídeos do íntron 2 no transcrito

(HIGASHI *et al.*, 1988). O Genótipo Grupo A é dito para as combinações A/A ou A/nulo e está associado ao fenótipo HAC-PS em 91% (n=317) dos casos, segundo dados extraídos do estudo de New *et al.* (2013). No entanto, em 8% (n=28) foram observados em HAC-SV.

O Grupo de Mutações B compreende mutações relacionadas a uma atividade residual da CYP21A2 que varia entre 1-10% da atividade normal da CYP21A2 (Figura 11) (SIMONETTI *et al.*, 2017). A mutação desse grupo mais comumente encontrada é a p.I173N que apresenta uma atividade entre 1-2% em relação ao normal. O Genótipo Grupo B é dito para as combinações B/B, B/A e B/nulo e está associado ao fenótipo HAC-SV, sendo observado em 76% (n=138) dos casos, segundo dados extraídos do estudo de NEW *et al.* (2013). No entanto, em 23% (n=42) desses, foi observado o fenótipo HAC-PS.

O Grupo de Mutações C é composto por mutações que permitem uma atividade parcial da CYP21A2 (Figura 11). Esse é o grupo que tem o maior espectro de atividade da CYP21A2, que varia de 10 a 78%, em relação à atividade normal da CYP21A2 (SIMONETTI *et al.*, 2017). A mutação mais comum deste grupo é a p.V282L, que tem uma atividade em torno de 20-30% para progesterona e 50-60% para a 17OHP em relação a CYP21A2 normal (SPEISER *et al.* 1992; KRONE *et al.*, 2013). Outras mutações bem conhecidas são P31L e P454S (NEW *et al.*, 2013). O Genótipo Grupo C é dito para as combinações C/C, C/B, C/A ou C/nulo, e é relacionado em 95% dos casos ao fenótipo HAC-NC, segundo dados extraídos do estudo de New *et al.* (2013). Os outros 5,4% (n=31) foram identificados com as outras formas de HAC (3% VS e 2% PS).

O Grupo de Mutações D é composto por variantes ainda não caracterizadas. A elucidação dessas novas mutações é feita através de ensaios funcionais da atividade de CYP21A2 que avaliam a taxa de conversão de progesterona em DOC e/ou 17OHP em 11-desoxicortisol da enzima mutada em relação à selvagem.

1.5.1 Variações na categorização das mutações

A busca e análise de mutações no gene *CYP21A2* realizada por SIMONETTI *et al.* (2018) encontrou 122 variantes com atividade da CYP21A2 testada *in vitro* depositadas em bancos de dados. Destas variantes, 95.4% apresentam atividade menor que 10% em relação à selvagem e são associadas à forma clássica de HAC (-PS e -SV). Enquanto isso, CYP21A2 mutadas com atividade residual entre 10 e 78% são associadas à forma NC em 90%.

Os Grupos de Mutações nulo e A englobam as mutações que afetam a atividade da enzima em $\geq 99\%$, sendo a atividade da CYP21A2 mutada entre 0-1% em relação à selvagem (KRONE *et al.*, 2013; RIEDL *et al.*, 2019). O grupo de mutação A também é classificado em outros estudos com a atividade residual da CYP21A2 sendo 0-2% (CARVALHO *et al.*, 2016; SANTOS-SILVA *et al.*, 2019).

Por outro lado, os Grupos de Mutação B e C apresentam a porcentagem de atividade da enzima CYP21A2 entre diferentes estudos ainda mais variáveis. O grupo B pode variar a atividade residual na faixa entre 1 e 10% e o grupo C entre 5 e 85%, ambos em relação à enzima selvagem (SIMONETTI *et al.*, 2017). Ademais, em alguns trabalhos, uma faixa de atividade entre esses grupos não é contemplada.

- O trabalho de SPEISER *et al.* (1992) inicialmente categorizou o grupo de mutação B com a mutação p.1173N cuja atividade da CYP21A2

observada em estudos *in vitro* foi de ~2% em relação à selvagem, e o grupo de mutação C com a atividade entre 20 e 50% baseado em p.V282L e p.P31L.

- O trabalho de WEDELL *et al.*, (1994) caracterizou um grupo de mutação B com atividade da CYP21A2 entre 2 e 10% em relação à selvagem, enquanto para o grupo C entre 7 e 75%.
- O estudo de KRONE *et al.*, (2013) classificou as mutações do grupo B com a atividade da CYP21A2 semelhante à p.I173N ou entre 1-10% em relação à selvagem em estudos *in vitro*. Enquanto, mutações do grupo C foram descritas como aquelas com a atividade da CYP21A2 semelhante às mutações p.P31L, p.V282L e p.P454S ou maiores que 20-60% em relação à CYP21A2 selvagem em ensaios *in vitro*.
- O estudo de RIEDL *et al.* (2019) classificou as mutações do grupo B com atividade da CYP21A2 entre 1-2% em relação à selvagem, incluindo p.I173N, e do grupo C com atividade entre 20-60%, incluindo p.P31L, p.V282L e p.P454S.
- O trabalho de SANTOS-SILVA *et al.* (2019) classificou no grupo B mutações com atividade da CYP21A2 de ~2% em relação à selvagem, incluindo p.I173N, e o grupo C mutações com atividade 20-60%, incluindo p.P31L, p.V282L, p.P454S e c.13*G>A.
- O estudo de CARVALHO *et al.* (2016) classificou o grupo B mutações com atividade entre 3 e 7% em relação à selvagem, numa faixa que foi englobada a p.I173N. O grupo C incluiu p.P31L, p.V282L e p.P454S e as mutações com atividade entre 20 e 50% relativa à selvagem.

- Já o estudo de BARBARO *et al.* (2015) estabeleceu uma faixa de atividade da CYP21A2 baseada em mutações associadas ao fenótipo HAC-NC, grupo de mutações C. A atividade da CYP21A2 mutada relacionada aos fenótipos HAC-NC graves com clitomegalia ao NC leves foram obtidos entre 5 e 85% para a conversão de 17OHP em 11-desoxicortisol e entre 4 e 66% para progesterona em DOC, ambas atividades relacionadas a enzima selvagem. Interessante realçar, que a mutação p.P31L característica do grupo C, mostrou valor abaixo desta faixa, sendo de 1.3% para a 17OHP e 2% para a progesterona.

1.5.2 Influência de fatores externos no fenótipo de HAC

A variação entre o fenótipo esperado e o observado tem sido relatada nos genótipos A, B e C, havendo inclusive diferenças entre estudos populacionais (NEW *et al.*, 2013; CARVALHO *et al.* 2016). Essas variações entre o genótipo e fenótipo foram observadas tanto com mutações em homozigose, destaque para a IVS2-13A/C>G e p.P31L, quanto em heterozigose composta.

Em parte, uma explicação seria a perda de mutações menos frequentes nos rastreios populacionais genéticos que, na maioria dos casos, envolvem entre 10 e 14 mutações pontuais específicas e grandes deleções (CARVALHO *et al.*, 2016). A utilização da análise de todo o gene por sequenciamento Sanger ou, mais recentemente, o sequenciamento de nova geração (sigla em inglês, NGS) são abordagens para minimizar essa variável de erro.

Outra linha estudada é a influência fenotípica de fatores externos. Apesar da HAC ser uma doença genética monogênica, fatores externos à doença podem

influenciar nas manifestações clínicas dentro do mesmo grupo genético, como no grau de virilização (Figura 12) e de perda de sal adrenal. A presença da CYP21A2 fora da adrenal, mutações nos promotores do gene *CYP21A2*, variações no metabolismo de andrógenos, sensibilidade aos andrógenos e influência de outras enzimas envolvidas no metabolismo de esteroides são alguns dos fatores estudados (MULLER & AUCHUS 2011; KAUPER *et al.*, 2012).

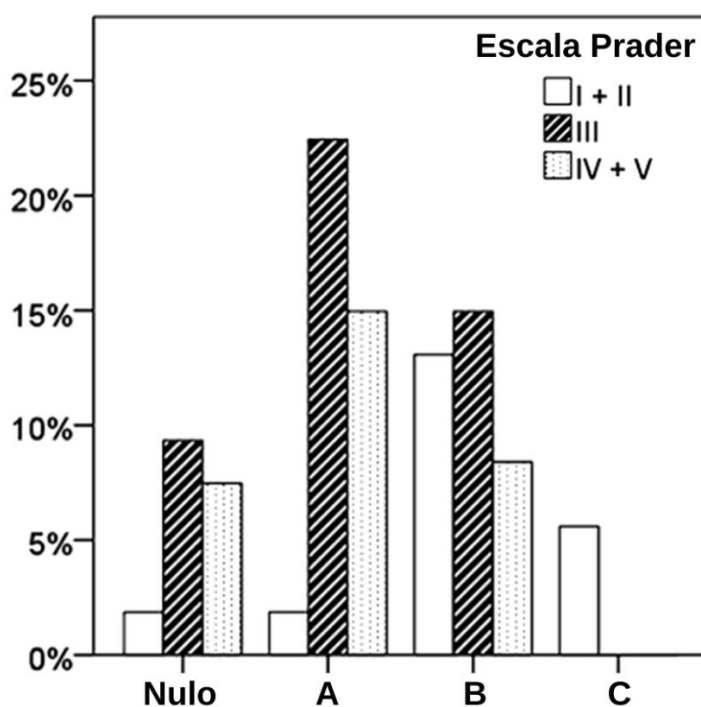


Figura 12. Virilização da genitália externa classificada com a escala Prader por grupos de mutação. No grupo estudado por Carvalho *et al.*, (2016) foram avaliadas 359 mulheres 46,XX com HAC, e destas 101 apresentaram virilização. Em relação ao total de mulheres com a forma leve de HAC (Genótipo C), 8% apresentaram virilização Prader I. Fonte da figura CARVALHO *et al.* (2016).

No estudo realizado por Kauper *et al.* (2012), por exemplo, foi observado que o Receptor Constitutivo de Androstano (CAR) têm efeito modulador indireto no grau de virilização genital. O CAR é um dos fatores de transcrição da principal enzima metabolizadora de esteroides com 19 carbonos (gene *CYP17A1*), como a androstenediona. Essa enzima está presente no período gestacional da diferenciação sexual e mostra alta variabilidade da sua atividade entre indivíduos.

Em relação à variabilidade no fenótipo de pacientes com HAC-NC, a presença de virilização observada em alguns casos tem sido relacionada à presença de outros genes moduladores. Um dos candidatos apontado por MOURA-MASSARI *et al.*, (2016) foi o receptor de andrógenos. Esse estudo com pacientes HAC-NC com e sem clitomegalia, mostrou associação entre o número de cópias do polimorfismo (nCAG) no gene receptor de andrógeno com a presença de virilização. O número de repetições bi-alélicas pequenas mostrou ser estatisticamente diferente em HAC-NC com clitomegalia (18CAG) em relação aos sem clitomegalia (21CAG), indicando uma possível modulação do fenotípica na HAC-NC.

A associação de outras doenças também pode modular o fenótipo. Como exemplo, o paciente neonato com o genótipo grupo C observado por WADELL *et al.* (1994). Esse paciente recém-nascido apresentou infecção gastrointestinal e fenótipo HAC-PS. Entretanto, após o término da infecção os sintomas de HAC-PS também se extinguiram, havendo a descontinuação do uso de glicocorticoides administrados para a regulação eletrolítica.

1.6 Diagnóstico e terapia de HAC

O diagnóstico hormonal e os tratamentos serão abordados nas etapas em que eles podem ser realizados, no pré-natal e/ou no pós-natal, uma vez que não apresentam o mesmo fluxo de procedimentos. Já a análise genética será apresentada na sequência deste, visto que em ambas as etapas pode ser solicitada.

1.6.1 Pré-natal

A triagem de doenças genéticas no período pré-natal pode ser realizada a partir da 5^a-6^a. semana de gestação através da coleta de sangue periférico materno para a análise genética pela técnica mais recente de diagnóstico pré-natal não invasivo (NIPD, sigla em inglês) que utiliza uma amostra de sangue periférico da mãe para a determinação do genótipo e cariótipo do feto. Além da NIPD, a análise genética pode ser realizada através da amostragem de vilosidades coriônicas com 10 a 12 semanas de gestação, a mais frequente técnica utilizada, ou por amniocentese com 15 a 16 semanas de gestação (BARONIO *et al.*, 2019; BAUMGARTNER-PARZER *et al.*, 2020).

A terapia para HAC, no período pré-natal, visa prevenir a virilização de genitália de fetos femininos pela supressão do ACTH fetal e do hiperandrogenismo adrenal. Entretanto, ela ainda é considerada experimental devido a dados controversos quanto ao resultado do tratamento e a segurança materna e fetal (BARONIO *et al.*, 2019).

1.6.2 Pós-natal

A triagem hormonal de HAC de recém-nascidos estabelecida mundialmente é realizada pela dosagem da 17OHP em soro ou gota de sangue seco em papel filtro, através de radioimunensaio (RIAs), ensaios de imunoabsorção enzimática (ELISA), fluoroimunoensaio (DELFIAs) ou cromatografia acoplada à espectrometria de massas em tandem (LC-MS/MS) (ALVES *et al.*, 2019; MILLER, 2019). Atualmente, devido ao número de casos falso positivos (FP) detectados nas triagens neonatais, tem se discutido a substituição da dosagem do esteroide 17OHP pelo 21-desoxicortisol, numa tentativa de minimizar esses casos FP que, além dos custos ao sistema, causam estresse e ansiedade desnecessários às famílias destes (MILLER, 2019). Os casos FP ocorrem principalmente em casos de recém-nascidos com nascimento prematuro, baixo peso, desidratação, icterícia ou submetidos a estresse neonatal. Pode haver também casos falso negativos (FN), quando a coleta do exame foi realizada antes do segundo dia de vida ou quando o recém-nascido (ou a mãe durante a gravidez) foi tratado com corticosteroide (KOPACEK *et al.*, 2017; ALVES *et al.*, 2019). Adicionalmente ao 17OHP, a análise de outros esteroides é realizada a fim da confirmação hormonal de HAC. No Brasil, os testes confirmatórios laboratoriais são as dosagens séricas de cortisol, androstenediona, testosterona, sódio e potássio, podendo ser solicitado também uréia, creatinina, glicemia e gasometria (ALVES *et al.*, 2019).

Tendo em vista que, entre 7-17% dos casos FP não se obtém a normalização dos analitos nas recoletas e testes confirmatórios séricos, a realização do teste genético é recomendada (ALVES *et al.* 2019; BAUMGARTNER-PARZER *et al.*, 2020). A indicação de teste genético no Brasil é principalmente para identificar

HAC-VS e HAC-NC em meninos e descartar a doença em pacientes FP assintomáticos com valores persistentemente elevados, diminuindo o tempo de acompanhamento desnecessário. Casos FN para HAC podem demorar até 6-12 meses para normalizar os valores de 17OHP (ALVES *et al.* 2019).

O tratamento pós-natal de todas as formas de HAC apresenta um relativo aumento na qualidade de vida dos portadores, apesar de alguns sintomas gerais poderem permanecer em adolescentes e adultos, como calvície temporal, acne e risco de tumor testicular em homens 46,XY, e hirsutismo, irregularidade menstrual e síndrome do ovário policístico (SOP) em mulheres 46,XX (NEW *et al.*, 2017). O tratamento adequado está diretamente relacionado à amenização dos sintomas num nível geral, enquanto, a administração inadequada de glicocorticoides ou ausência de tratamento geram o agravamento do quadro clínico. A Figura 13 resume algumas das principais manifestações da HAC e os efeitos colaterais dos tratamentos.

O principal objetivo no tratamento de reposição de cortisol na HAC é o restauro da fisiologia normal através da dosagem ideal hormonal, evitando-se superdosagens, síndromes metabólicas secundárias, doenças cardiovasculares e a osteoporose (HAHNER *et al.*, 2021). A terapia de reposição de cortisol através de sua forma sintética (hidrocortisona ou acetato de cortisona) geralmente é baseada na forma de HAC, nos sintomas e na idade do paciente. A hidrocortisona também exerce atividade mineralocorticoide. Terapias com baixas doses de hidrocortisona causam fadiga, perda de peso e náusea, podendo levar a crises adrenais. Por outro lado, altas dosagens podem causar o excesso de cortisol

gerando características da síndrome de Cushing, como ganho de peso, aumento da gordura abdominal, “corcunda de búfalo”, afinamento da pele, hematomas faciais, hipertensão e até diabetes mellitus tipo 2 (HAHNER *et al.*, 2021). Hormônio para a regulação do crescimento também pode ser administrado.

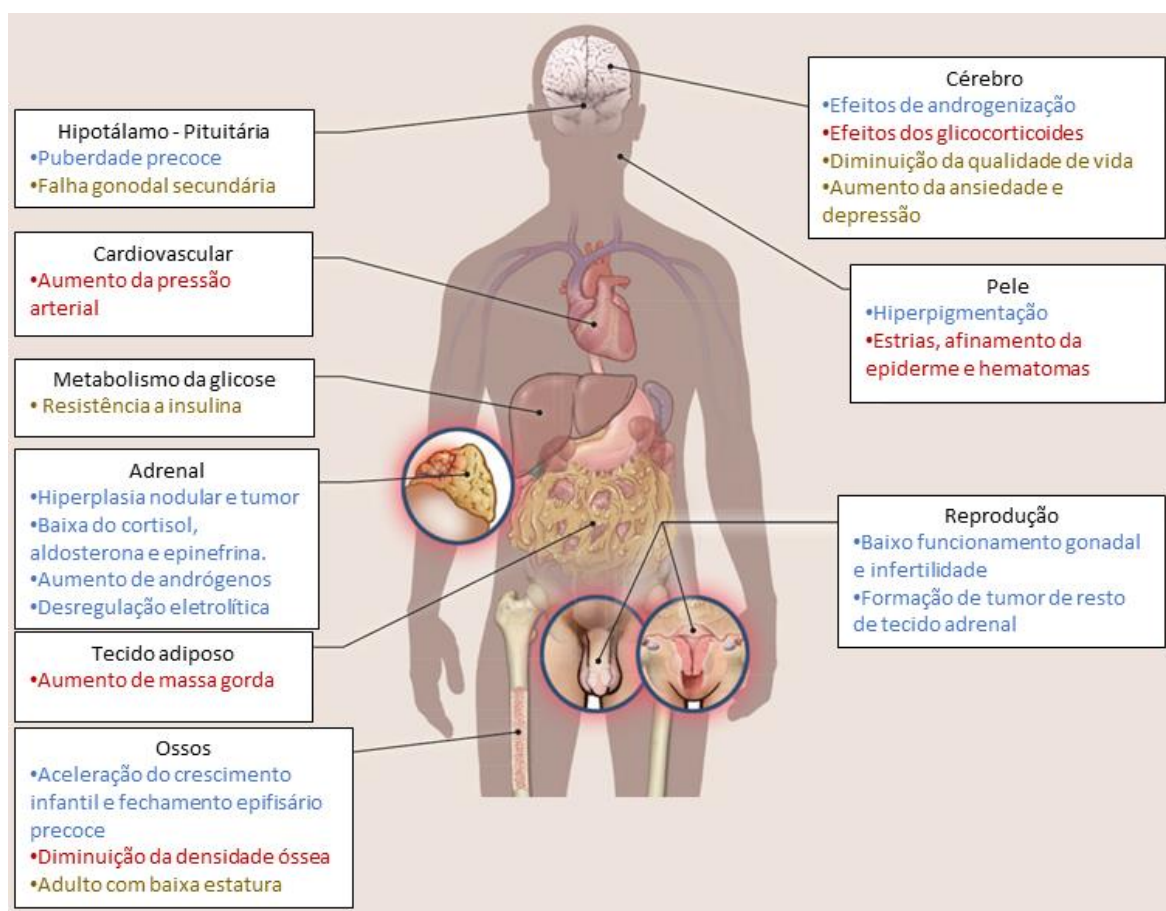


Figura 13. Quadro clínico da hiperplasia adrenal congênita. Características da doença (em azul), do tratamento (vermelho) e da combinação da doença com o tratamento (marrom). Figura adaptada de MERKE & AUCHUS (2020).

A terapia de reposição de mineralocorticoides é utilizada para a prevenção de crises de perda de sal e restauração da pressão sanguínea normal (HAHNER

et al., 2021). Uma vez que a molécula de aldosterona tem a meia-vida muito curta e rápido metabolismo hepático após a ingestão, ela não é usada diretamente para este fim. Atualmente, a terapia para a reposição da aldosterona é realizada pelo substituto do deoxicorticosterona, a fludrocortisona (9-alfa-fluorocortisol) (HAHNER *et al.*, 2021). Essa molécula mostra entre 200 e 400 vezes mais efeito mineralocorticoide do que a hidrocortisona. Altas doses podem ser percebidas pelo rápido ganho de peso, pressão alta, edema e hipocalcemia (HAHNER *et al.*, 2021). Adicionalmente, de 1-2 gramas de cloreto de sódio deve ser consumido ao longo das refeições do dia para repor a perda (ALVES *et al.*, 2019)

Por fim, visto que os substratos da CYP21A2 na rota de biossíntese do cortisol continuam sendo acumulados pela deficiência da enzima CYP21A2, a regulação dos hormônios sexuais se faz necessária, principalmente em pacientes do sexo feminino com alto nível de andrógenos adrenais. A hidrocortisona exerce atividade também de prevenir o progresso da virilização pela inibição da hiperprodução de andrógenos pela redução da secreção de ACTH, porém níveis acima do fisiológico deveriam ser mantidos para a obtenção de resultados (BARONIO *et al.*, 2019). Dependendo do grau de virilização, com a administração de medicamentos é possível reverter o quadro. No entanto, nas escalas Prader referente aos estágios mais avançados, a cirurgia de correção da genitália feminina, clitoroplastia e/ou vaginoplastia, pode ser sugerida. Ademais, a distrofia de gênero é observada em meninas com virilização, sendo a causa de distúrbios emocionais e comportamentais (NARASIMHAN & KATTAB, 2019). No Brasil, a cirurgia pode ser realizada até os dois anos de idade do paciente apenas pelo consentimento dos pais. Entretanto, após os 2 anos de idade, um exame

psicológico para avaliação da identidade de gênero deve ser procedido para a autorização cirúrgica (ALVES *et al.*, 2019).

1.6.3 Genético

A análise genética do gene *CYP21A2*, cobrindo mutações pontuais (troca de base única, pequenas deleções ou inserções), número de cópias e grandes eventos deletérios, é indicada, a fim de cobrir as mutações patogênicas mais comum nesse gene (BAUMGARTNER-PARZER *et al.*, 2020). As metodologias de análise genética do gene *CYP21A2* são:

As triagens rápidas de variantes específicas, que são técnicas mais utilizadas como primeira estratégia de análise de grandes grupos. Essas triagens visam detectar as variantes pontuais mais comum, como p.P31L, IVS2-13A/C>G, p.G111fs, p.I173N, Cluster E6, p.V282L, p.L308fs, p.Q319X, p.R357W e p.P454S. Como exemplos, há a técnica de minisequenciamento multiplex após o PCR alelo específico para descarte do pseudogene (KRONE *et al.*, 2013; PRADO *et al.*, 2017).

Já para a análise do gene completo, ou quase completo, a metodologia considerada padrão ouro é o sequenciamento do tipo Sanger baseado em PCR alelo específico. Na maioria dos estudos utiliza-se essa técnica para uma análise complementar (CARVALHO *et al.*, 2016). Ademais, uma nova alternativa para a análise do gene completo e que tem começado a mostrar relativamente bons resultados para o gene *CYP21A2* é as tecnologias de NGS (BAUMGARTNER-PARZER *et al.*, 2020; KaRAOĞLan *et al.*, 2021).

Quanto à análise de grandes eventos genéticos e de número de cópias do gene o método mais utilizado para o gene *CYP21A2* tem sido o do kit de

amplificação de sondas dependentes de ligação multiplex (MLPA, sigla em inglês) da MRC-Holland. Mais recentemente, o kit CAH RealFast CNV Assay da Viennalab foi desenvolvido com o mesmo propósito do MLPA, detectando o número de cópias gênicas. Por fim, uma técnica mais tradicional, o PCR quantitativo, que é utilizado com sondas padronizadas em cada laboratório (BAUMGARTNER-PARZER *et al.*, 2020).

A análise genética completa do gene *CYP21A2* incluindo mutações pontuais e o número de cópias é altamente indicado devido às variantes espontâneas *de novo* (BAUMGARTNER-PARZER *et al.*, 2020). Complementarmente, a elucidação de variantes novas é importante para a obtenção de um diagnóstico molecular mais preciso e relacionado ao fenótipo do paciente.

1.7 Elucidação de novas variantes

Diferentemente de outras doenças monogênicas, inferir o efeito de uma nova mutação pelo fenótipo do portador da deficiência da *CYP21A2* é ainda mais complicado, pois pode ser facilmente mascarado. Isso se deve ao fato do gene *CYP21A2* ser complexo, podendo apresentar número de cópia(s) extra(s), deleção do gene, mutações *de novo* não rastreadas, múltiplas mutações no mesmo alelo, além dos possíveis modificadores do fenótipo descritos anteriormente no item 1.5.2 (TUSIÉ-LUNA & WHITE, 1995; MOURA-MASSARI *et al.*, 2016; NARASIMHAN & KATTAB, 2019).

1.7.1 Caracterização in silico

A primeira estratégia para a caracterização de uma nova variante é realizada através de análises computacionais que podem envolver análise evolutiva,

bioquímica, físico-química e estrutural. Entre as vantagens dessas técnicas estão o custo e a agilidade na obtenção de predições de efeito de uma ou várias variantes. No entanto, muitas vezes há variabilidade nos resultados entre os preditores e a eficiência entre eles difere para cada gene e, até o momento, não havia sido realizada uma análise comparativa entre a eficiência de preditores para o gene *CYP21A2*. A seguir serão abordadas as principais características estruturais da proteína *CYP21A2* analisadas em análises de danos estruturais de variantes.

A proteína *CYP21A2* apresenta a típica conformação estrutural das CYPs humanas, que remete a um prisma triangular com pelo menos 12 α -hélices e um domínio folha- β perto do N-terminal. A estrutura da proteína *CYP21A2* humana é composta de 13 α -hélices, sendo consideradas as subpartes da hélice B (B e B'), bem como da F (F e F'), uma única hélice. Além disso, a proteína *CYP21A2* tem nove folhas- β , sendo sete no domínio folha- β do N-terminal (β 1- β 7) e duas no C-terminal (β 8- β 9) (Figura 14) (PALLAN *et al.*, 2015).

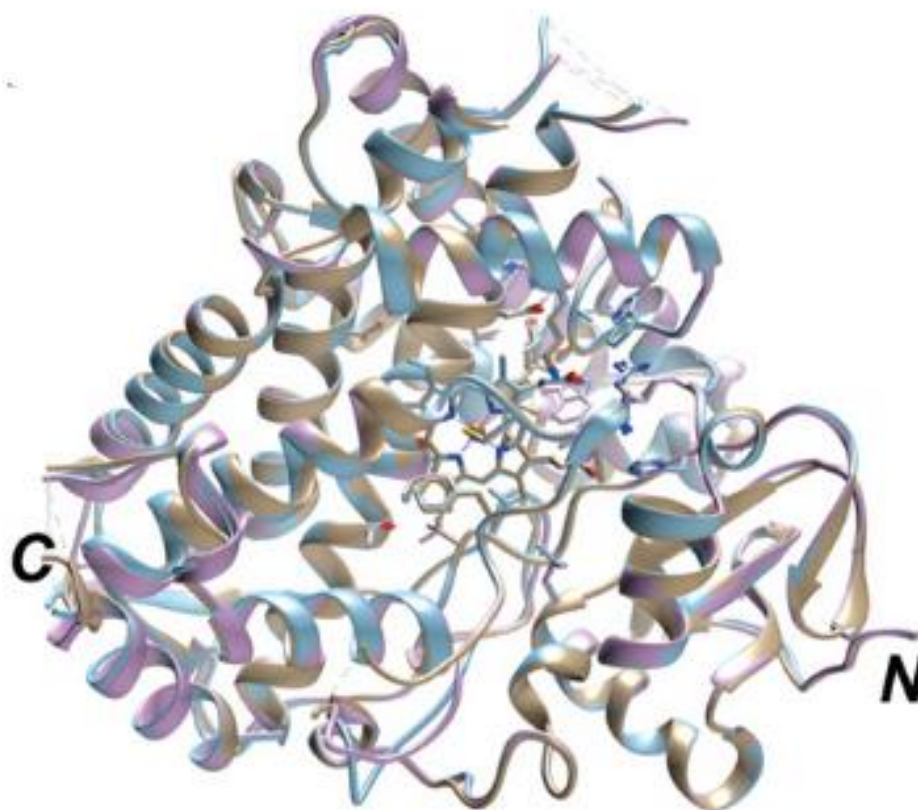


Figura 14. Estrutura em forma de prisma triangular típica da família de proteínas citocromo P450s. As estruturas são compostas por pelo menos 12 α -hélices e um domínio folha- β perto do domínio N-terminal. Fonte da figura PALLAN *et al.* (2015).

O grupo de pesquisa do PALLAN *et al.* (2015) obteve o primeiro cristal da estrutura da CYP21A2 humana, sendo este com o complexo do substrato progesterona. A estrutura está depositada no *Protein Data Bank* (PDB) sob a identificação (ID) 4Y8W (Figura 15A-B). Pouco tempo depois, o mesmo grupo de pesquisa obteve o cristal da CYP21A2 humana com o complexo do substrato 17OHP, PDB ID 5VBU (Figura 15C-D) (WANG *et al.*, 2017). O esteroide 17OHP

tem, virtualmente, exatamente o mesmo sítio ativo identificado para a progesterona, com apenas uma torção de ângulo entre os carbonos 17 e 20. A única diferença conformacional entre as duas estruturas é do resíduo 411 ao 418 na estrutura aleatória (*random coil*) que conecta as α -hélices L e M (Figura 15A e C) (WANG *et al.*, 2017).

Com a estrutura da CYP21A2 humana foi possível fazer uma análise estrutural mais real dos aminoácidos e dobramentos importantes para a conformação da enzima e, conseqüentemente, para a sua funcionalidade. A seguir, serão apresentados alguns desses pontos que foram descritos em dois artigos originais do grupo (PALLAN *et al.*, 2015 e WANG *et al.*, 2017). As estruturas das moléculas da progesterona e da 17OHP são apresentadas na Figura 16.

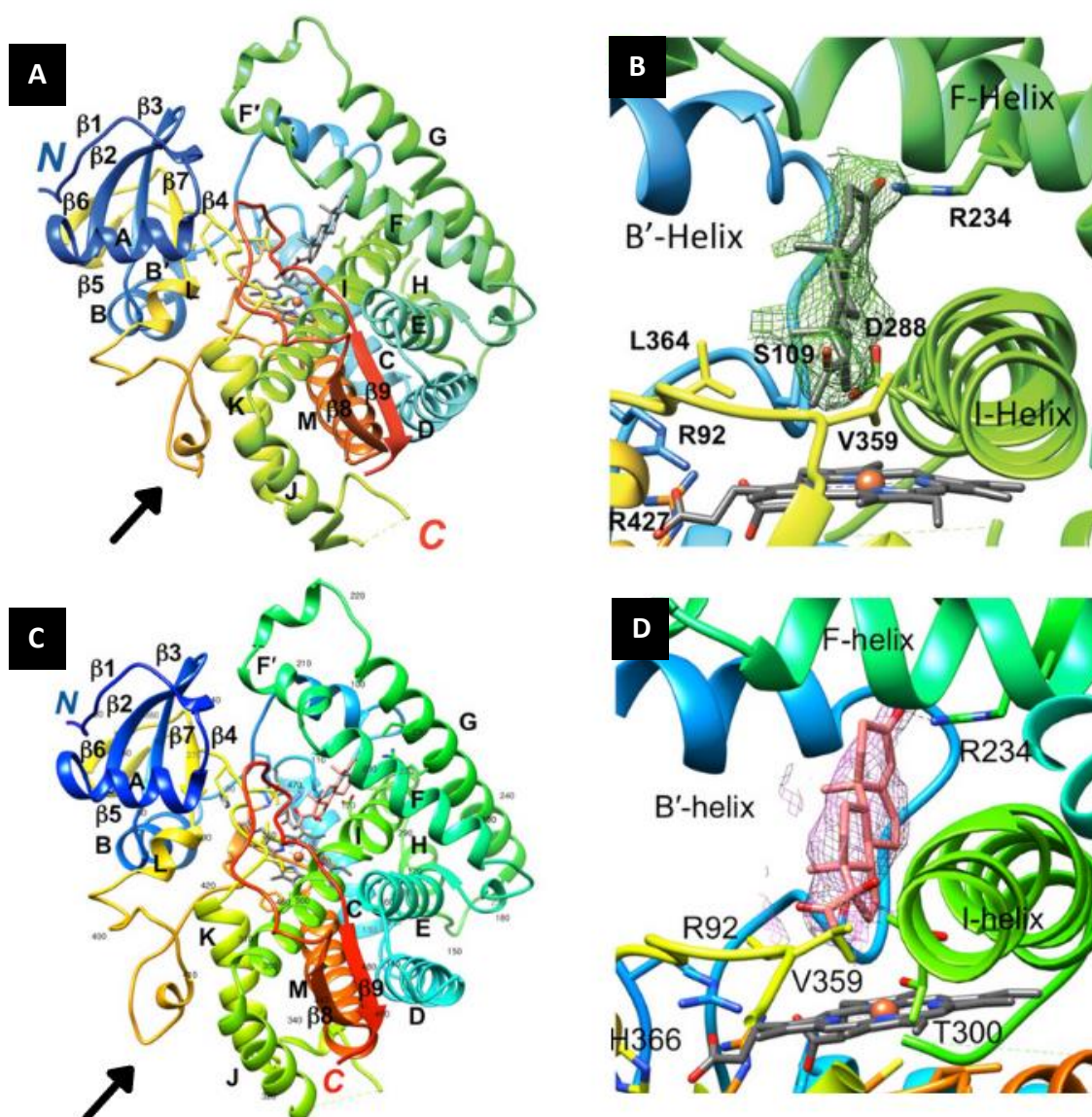


Figura 15. Estruturas tridimensionais da proteína 21-hidroxilase obtidas por cristalografia de raio X. (A) Estrutura obtida por PALLAN *et al.*, (2015) com o esteroide progesterona conectado (PDB ID 4Y8W). (B) Esteroide progesterona ligado no sítio catalítico. (C) Estrutura obtida por Wang *et al.*, (2017) com o esteroide 17-hidroxiprogesterona conectado (PDB ID 5VBU). (D) Esteroide 17-hidroxiprogesterona ligado ao sítio catalítico. Flechas pretas indicam alterações

encontradas na comparação entre as duas estruturas. Figura adaptada de PALLAN *et al.*, 2015 e WANG *et al.*, 2017.

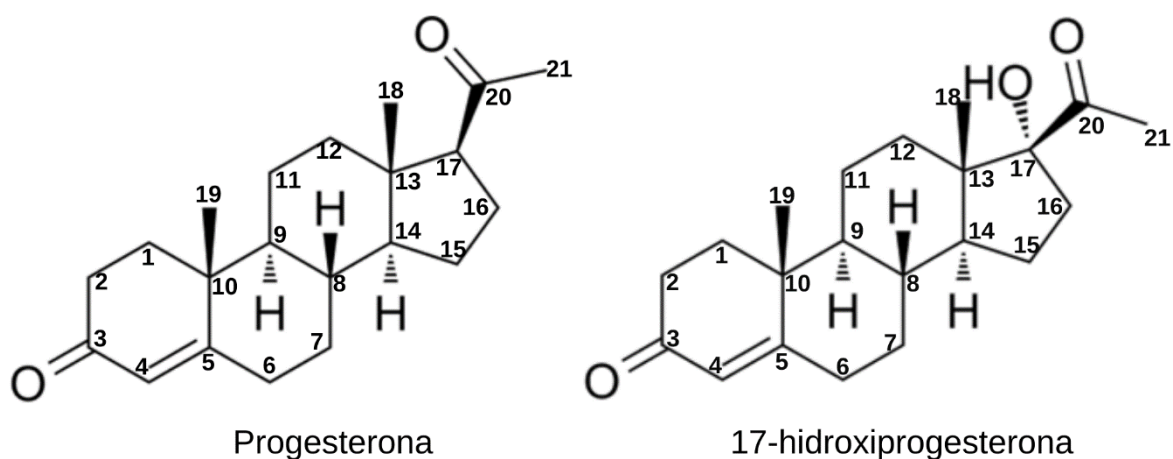


Figura 16. Estrutura química dos substratos da 21-hidroxilase em seus estados naturais.

- Estrutura do sítio catalítico:

O espaçamento entre o íon Fe^{3+} do grupamento heme da CYP21A2 e a metila ($-\text{CH}_3$) do carbono 21 da progesterona é entre $\sim 4,0 - 4,3 \text{ \AA}$ (PALLAN *et al.*, 2015; WANG *et al.*, 2017) (Figura 15B e D). No geral, a distância média entre moléculas conectadas é menor que 4 \AA , sendo essas distâncias obtidas consideradas grandes segundo PALLAN *et al.* (2015). Todavia, a adição da hidroxila reduziu essa distância para $\sim 3 \text{ \AA}$ entre o Fe^{3+} e o carbono 21, e para $\sim 2,3 \text{ \AA}$ entre o Fe^{3+} e a molécula de hidrogênio do carbono 21 (PALLAN *et al.*, 2015).

Já na extremidade oposta da molécula do substrato, o grupamento carbonila ($\text{C}=\text{O}$) (Figura 16) faz uma ligação do tipo ponte de hidrogênio com o grupamento amino ($-\text{NH}_2$) do resíduo p.R234, localizado na α -hélice G (distância $\sim 2,66 \text{ \AA}$ entre

O₃ --NH₂). Além disso, ao longo da molécula do substrato, uma de suas faces interage principalmente com os resíduos hidrofóbicos p.V287, p.I291 e p.G292 da α -hélice I (PALLAN *et al.*, 2015). O grupamento metil do carbono 19 do substrato (Figura 16) se dispõe próximo do resíduo **p.W202** da CYP21A2, localizado na α -hélice F, sendo sugerida uma possível interação entre eles do tipo ponte de hidrogênio (PALLAN *et al.*, 2015). Já o carbono 6 do substrato fica a 3,90 Å da porção carboxilato do resíduo p.D107, enquanto, o carbono 7 fica a 3,37 Å do Cy do resíduo p.S109. O resíduo p.D288 (α -hélice I) fica perto de dois carbonos do substrato (~3,34 Å do carbono 6 e ~3,34 Å do carbono 7) (Figura 15, Figura 16) (PALLAN *et al.*, 2015). Por fim, apesar de alguns *loops* da estrutura se localizarem perto do domínio do esteroide, não há contato visível com os resíduos dos *loops* entre as α -hélices B' e C, entre a α -hélice K e folha- β 4, e entre a folhas- β 8 e folha- β 9 (PALLAN *et al.*, 2015).

- Pontos críticos do enovelamento e estabilização da estrutura:

Tanto a desestabilização da proteína como o aumento significativo da estabilidade podem afetar a funcionalidade da proteína devido a um aumento da degradação proteica quanto a perda de sua dinâmica. Um estudo que analisou mutações na proteína CYP21A2 demonstrou um alto impacto da estabilidade estrutural proteica para a funcionalidade enzima, sendo que das 76 mutações que alteraram a estabilidade da proteína, 52 tinham impacto biológico (BRUQUE *et al.*, 2016).

As interações de enovelamento da proteína são mediadas principalmente pelas α -hélices B (B e B'), C, J e M, além dos resíduos terminais das α -hélice F e

H, posições 189 e 264, respectivamente. Os resíduos de *loops* que também foram observados participar dessas interações de enovelamento são os das posições **35-36**, 66-67, 213-223, 247-252, 380-382, 412-414 (PALLAN *et al.*, 2015).

Um dos módulos importantes para na condução de formação da α -hélice M é a Tríade-ERR (**p.E352**, p.R355 e Rp.409). Nessa tríade, o resíduo ácido glutâmico 352 (aminoácido carregado negativamente) tem uma ligação do tipo ponte salina com a arginina 409 (aminoácido carregado positivamente). Ensaios substituindo o ácido glutâmico 352 pela lisina resultaram na repulsão de três cadeias laterais básicas, devido ao rompimento da ponte salina com o R409 (PALLAN *et al.*, 2015). Sabe-se que a mutação p.R409H causa a total perda da funcionalidade da enzima CYP21A2, o que poderia também estar relacionado com a perda dessa ponte salina devido à cadeia lateral da histidina ser mais curta (HAIDER *et al.*, 2013).

Além do dobramento da proteína em si, problemas na incorporação do grupamento heme também são considerados como problemas no enovelamento da proteína. Testes com a proteína CYP21A2 mutada em 10 resíduos específicos (31, 65, 108, 172, 282, 292, 296, 303, 357 e 409) mostrou que praticamente todas as eles têm influenciam o conteúdo de heme (WANG *et al.*, 2017). O menor teor de heme obtido foi com as variantes p.P31Q, p.W303R e p.R357W, que apresentaram 10x menos conteúdo heme em comparação com a proteína nativa (WANG *et al.*, 2017). Já a mutação p.I173N, diminui a incorporação de heme em 7x e gera uma perda de atividade catalítica de 2% em relação a forma nativa. O p.I173 é situado na α -hélice E, em uma região hidrofóbica onde apresenta conexões com p.V140

(α -hélice D), p.L176 (α -hélice E) e p.L434 (α -hélice M), além de ter conexão com o aminoácido hidrofílico p.G438 (α -hélice M). A troca da isoleucina pela asparagina na posição 172 afeta a região hidrofóbica, porém, mantém as conexões com o p.G438 e p.V140 (WANG *et al.*, 2017).

O enovelamento da proteína CYP21A2 acomoda alguns tipos de alterações. No entanto, dependendo da posição, a adição de um único grupo metil numa cadeia lateral pode ser o suficiente para alterar a funcionalidade da enzima (WANG *et al.*, 2017). Como exemplo, a substituição de uma valina por uma leucina na posição 282. O aumento do volume pela adição de apenas uma metila é acomodado em um certo grau nessa posição, entretanto, é o suficiente para diminuir a atividade de hidroxilação da CYP21A2, a ~20% para progesterona e ~50% para 17OHP, ambas em relação a CYP21A2 selvagem. Adicionalmente, a estabilidade térmica decai, sendo está praticamente inexistente. A p.V282 é localizada na α -hélice I, onde é rodeada pelos resíduos hidrofóbicos p.L242, na α -hélice G, e p.M258, p.M261 e p.L262 na α -hélice H (WANG *et al.*, 2017).

- Interação oxidorreductase:

No córtex adrenal, a proteína CYP21A2 humana é reduzida pela enzima POR (Figura 17). A POR é a principal mediadora do transporte de elétrons provenientes do fosfato de dinucleotídeo de nicotinamida e adenina (NADPH) para CYPs ancoradas na membrana do retículo endoplasmático (PANDEY & FRÜCK, 2013). Essa enzima tem dois distintos domínios cofatores de flavina, o dinucleotídeo de flavina e adenina (FAD) e o mononucleotídeo de flavina (FMN) que são conectados por uma região de dobradura flexível. O NADPH liga-se ao

domínio FAD, onde faz a transferência íons híbridos equivalentes. Esses íons são então transferidos para o anel de isoaloxazina no domínio FMN. Em uma interação entre POR (agente redutor) e um dos seus parceiros redox, como a enzima CYP21A2, (agente oxidante) os elétrons são transferidos do domínio FMN da POR para o domínio heme da proteína a ser reduzida, que irá realizar a hidroxilação de moléculas (FLÜCK & PANDEY, 2013). No entanto, diferentemente de outras CYPs, como a P450c17, a enzima CYP21A2 é menos sensível a abundância da POR (MILLER & AUCHUS, 2011).

Os exatos resíduos envolvidos nessa ligação não são conhecidos. No entanto, através de análises de similaridade de sequência e carga dos aminoácidos com outras CYPs (CYP2B4 e CYP17) foram sugeridos possíveis candidatos: p.K118, p.R122, p.R133, p.R415, p.R427, p.R436 e p.R340 (ROBIONS *et al.*, 2006). No entanto, cabe ressaltar, que a modelagem molecular foi realizada tendo como base a estrutura das proteínas com os alvos conhecidos (CYP2B4 ou CYP17) e a sequência referência de aminoácidos da CYP21A2, podendo haver diferenças das estruturas secundárias preditas (Figura 17).

O citocromo b₅, é outra enzima que tem a mesma função catalítica da POR para proteínas CYPs. Entretanto, foi observado que esse cofator não tem o mesmo efeito na atividade da enzima CYP21A2 que para outras CYPs, como a P450c17. Uma inibição da reação de hidroxilação da CYP21A2 em uma escala de quatro vezes foi demonstrado com a adição do citocromo b₅, porém mais estudos ainda devem ser realizados para a compreensão desse mecanismo (WANG *et al.*, 2017).

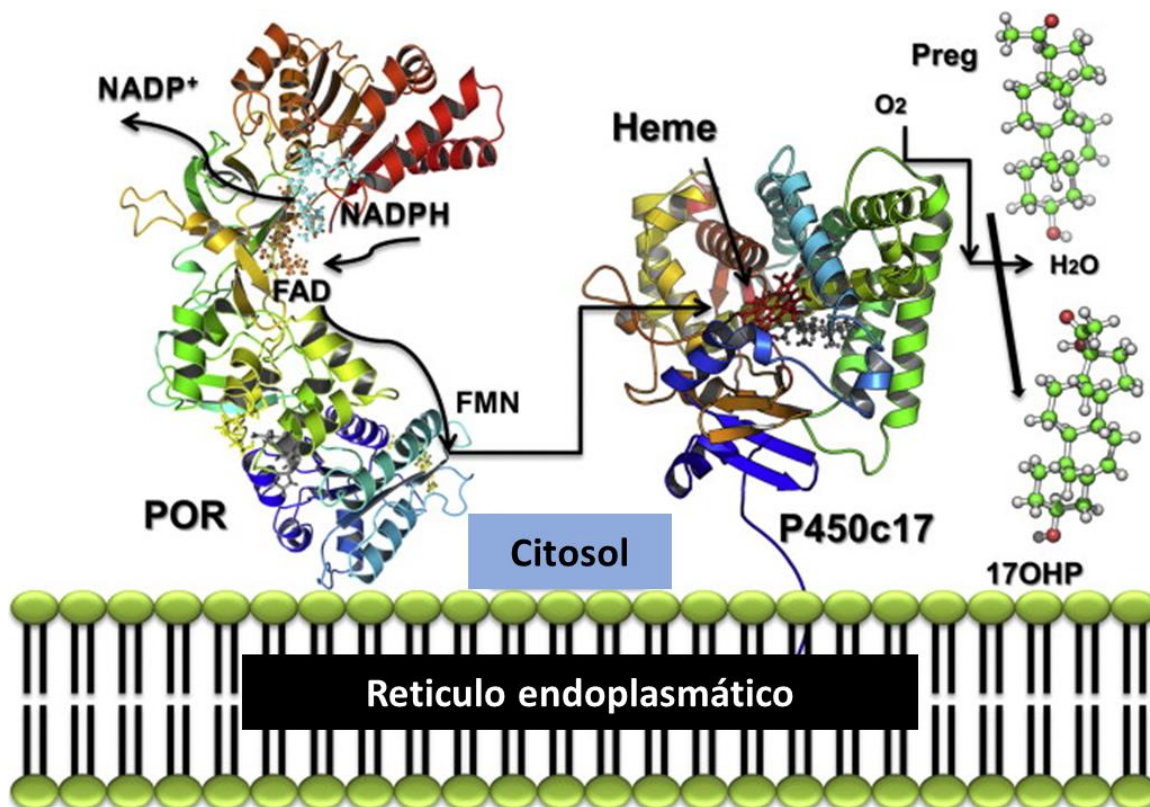


Figura 17. Esquema da reação de eletro-transferase mediada pela enzima P450 oxidoreductase (POR). A POR transfere elétrons para enzimas P450s, como a 21-hidroxiase (P450c21; CYP21A2), 17 α -hidroxilase (P450c17; 17 α -OH) e CYP19A1 (P450c19) catalisando as suas reações. A partir da ligação de uma molécula de NADPH na proteína POR, um par de elétrons é recebido pelo domínio dinucleotídico de flavina e adenina (FAD) da proteína, que por sua vez entrega esses elétrons para o domínio mononucleotídico de flavina (FMN). Quando a POR interage com uma P450s, é o domínio FMN que chega até o domínio heme da enzima, catalisando a reação desta. No exemplo do esquema apresentado, a enzima catalisada é a P450c17, que converte pregnenolona (Preg) em 17-hidroxiprogesterona (17OHP) é catalisada pela enzima POR. Figura adaptada de Pandey & Frück (2013).

1.7.2 Caracterização *in vitro*

A metodologia padrão ouro para a elucidação do dano de uma variante no gene *CYP21A2* é o teste de funcionalidade realizado *in vitro* com a enzima *CYP21A2* mutada com a variante de interesse. Esses ensaios podem ser realizados tanto com a proteína purificada (WANG *et al.*, 2017) quanto com a proteína expressa em cultura de células (RIEPE *et al.*, 2008; BARBARO *et al.*, 2015). Em ambos os ensaios, a atividade da *CYP21A2* é obtida pela taxa de conversão dos substratos nos produtos, relacionando à proteína *CYP21A2* mutada com a selvagem. A eficiência de hidroxilação da *CYP21A2* não é a mesma para ambos os substratos tanto em sistemas utilizando a proteína purificada quanto em células humanas em um sistema de cultura de células (WANG *et al.*, 2015; RIEPE *et al.*, 2008; KHAJURIA *et al.*, 2018).

Considerando a proteína *CYP21A2* purificada, foi demonstrado que a eficiência catalítica (k_{cat}/K_m) da progesterona é 5x maior em comparação com a obtida para a 17OHP, sendo a constante de afinidade do substrato (K_m) progesterona 7x maior e o tempo necessário para converter uma molécula de substrato (k_{cat}) 1.4x menor do que para a 17OHP (PALLAN *et al.*, 2015).

O ensaio com a proteína purificada utiliza um plasmídeo de expressão com o gene recombinante para a expressão em bactéria. A fim de se obter uma melhor ancoragem da proteína na membrana bacteriana, os nucleotídeos do gene humano *CYP21A2* codificadores dos aminoácidos referentes à região da proteína envolvida no ancoramento na membrana do retículo endoplasmático são substituídos. Ademais, uma cauda pode ser adicionada na extremidade C-terminal da proteína

para auxiliar sua purificação (ex., cauda de histidina). Esse tipo de ensaio é realizado em um tubo com a presença do cofator POR, da molécula de NADPH e do substrato da enzima. A vantagem desse método é que permite o máximo de isolamento da proteína CYP21A2 mutada para a sua avaliação. Por outro lado, as modificações da proteína podem afetar a análise de mutações perto dessas regiões.

Já o método em cultura celular é mais simples. Ele utiliza de um plasmídeo de expressão de célula eucariótica com o gene humano íntegro e o sistema basal de células humanas, não necessitando suplementações para a reação, além dos substratos a serem analisados. Além disso, deve-se ter cuidado com a linhagem celular escolhida, pois não pode expressar o gene *CYP21A2*. As mais utilizadas para esse fim são as células HEK293t, COS-1 e COS-9. A vantagem desse método é a obtenção de um sistema mais parecido com o *in vivo* e de uma forma mais ágil. No entanto, as desvantagens em relação ao anterior é o custo para a realização e a quantidade menor de enzima obtida, necessitando a utilização de técnicas mais sensíveis para a detecção do produto da reação, principalmente para as enzimas mutadas com atividade mínima ou residual.

Entre as vantagens em geral do ensaio funcional estão a obtenção da atividade da enzima CYP21A2 com a variante em estudo sem interferência de fatores externos do *in vivo* e a categorização entre os grupos de mutações da doença (nulo, A, B e C). Entretanto, ambas as técnicas requerem estruturas e custos normais de laboratório de biologia molecular e celular e demandam tempo

para a obtenção do resultado, não sendo viável para a análise de toda nova variante identificada.

2. Justificativa

A deficiência na enzima CYP21A2 é autossômica recessiva e um dos erros inatos do metabolismo mais comuns (NEW *et al.*, 2017). Essa deficiência enzimática afeta entre 1:10,000 a 20,000 n.v. no Brasil com suas formas clássicas, sendo essa incidência possivelmente 20 vezes maior para a forma NC (KOPACEK *et al.* 2017; HANNAH-SHMOUNI *et al.*, 2017a).

A perda da atividade da CYP21A2 impacta a síntese de glicocorticoides, podendo também afetar a síntese de mineralocorticoides e andrógenos. Seu vasto espectro fenotípico está relacionado principalmente ao grau de dano na atividade da enzima CYP21A2 causado pelos diversos tipos de mutações encontrados no complexo gene que a codifica, o gene *CYP21A2* (PIGNATELLI *et al.*, 2020). Além disso, a maior parte dos genótipos associados à deficiência da CYP21A2 são heterozigoto composto, podendo haver inclusive mais de uma mutação por alelo. (NEW *et al.*, 2013). Até o momento, mais que 380 mutações no gene *CYP21A2* foram reportadas no banco de dados *Human Mutation Database* (HMDB) e 122 variantes com a atividade da CYP21A2 caracterizada (SIMONETTI *et al.*, 2018).

Nos países desenvolvidos e em grande parte dos em desenvolvimento, como o Brasil, a deficiência da enzima CYP21A2 é triada em programas de rastreio de doenças genéticas que afetam gravemente recém-nascidos não tratados. As dosagens hormonais são uma importante ferramenta nessa etapa, entretanto, o uso do diagnóstico molecular para a detecção de variantes no gene *CYP21A2* tem sido uma valiosa estratégia na elucidação de casos complexos e acompanhamento da carga genética de cada população (SILVEIRA *et al.*, 2009; PIGNATELLI *et al.*,

2020). Entre as técnicas recomendadas para a análise genética, está o sequenciamento de todo o gene *CYP21A2*, o que tem permitido identificar novas variantes (BAUMGARTNER-PARZER *et al.*, 2020). Muitas são as variantes não caracterizadas, sendo a elucidação destas para o grupo populacional onde foi identificada de grande valia tanto para o futuro manejo de outros portadores quanto para o aconselhamento genético.

Capítulo II: Objetivos

Objetivo Geral

Caracterizar o efeito estrutural e funcional de variantes não elucidadas no gene *CYP21A2* previamente identificadas na população Brasileira ou Portuguesa e estabelecer um grupo de ferramentas de predição *online* com boa performance para a realização de análises de patogenicidade de variantes neste gene.

Objetivos específicos

- Identificar variantes no gene *CYP21A2* não caracterizadas na população Brasileira e Portuguesa.
- Analisar os danos das variantes na proteína *CYP21A2* através de análises *in silico* e seleção para testes *in vitro*.
- Caracterizar a funcionalidade das variantes selecionadas no modelo de cultura de células.
- Selecionar preditores *online* utilizados para a predição de dano de variantes *missense*.
- Estabelecer as ferramentas de predição *online* com melhor desempenho para com as variantes da proteína *CYP21A2*.

Capítulo III – Artigo 1

Characterization of mutations causing CYP21A2 deficiency in Brazilian and Portuguese populations

Publicado na revista:

International Journal of Molecular Science

Contribuição dos autores:

Conceituação, M.J.P., A.Z., M.L.R.R. e A.V.P.; metodologia, M.J.P., A.V.P, R.L.-B., S.S., B.V.M., T.R. e K.M.; análise formal, M.J.P. e A.V.P.; investigação, M.J.P.; recursos, A.V.P. e A.Z.; redação - preparação do rascunho original, M.J.P.; redação - revisão e edição, A.Z. e A.V.P.; visualização, M.J.P. e A.V.P.; supervisão, A.Z., M.L.R.R., A.V.P. e R.L.-B.; administração do projeto, A.V.P.; aquisição de financiamento, M.J.P., A.V.P., A.Z. e K.M..



Article

Characterization of Mutations Causing CYP21A2 Deficiency in Brazilian and Portuguese Populations

Mayara J. Prado ^{1,2,3,4}, Shripriya Singh ^{3,4}, Rodrigo Ligabue-Braun ⁵, Bruna V. Meneghetti ^{1,2}, Thaiane Rispoli ^{1,2}, Cristiane Kopacek ⁶, Karina Monteiro ^{1,2}, Arnaldo Zaha ^{1,2,†}, Maria L. R. Rossetti ^{1,7,†} and Amit V. Pandey ^{3,4,*,†}

¹ Graduate Program in Cell and Molecular Biology, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre CEP 91501-970, Brazil; mayjorgens@hotmail.com (M.J.P.);

bvalandromeneghetti@gmail.com (B.V.M.); thaianerispoli@gmail.com (T.R.);

karina.monteiro@ufrgs.br (K.M.); zaha.arnaldo@gmail.com (A.Z.); mrossett@terra.com.br (M.L.R.R.)

² Center for Biotechnology, Universidade Federal do Rio Grande do Sul (UFRGS),

Porto Alegre CEP 91501-970, Brazil

³ Department of Biomedical Research, University of Bern, 3010 Bern, Switzerland; shripriyasingh@gmail.com

⁴ Pediatric Endocrinology Unit, Department of Pediatrics, University Children's Hospital Bern, 3010 Bern, Switzerland

⁵ Department of Pharmacosciences, Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSA),

Porto Alegre CEP 90050-170, Brazil; ligabue.braun@gmail.com

⁶ Serviço de Referência em Triagem Neonatal, Hospital Materno Infantil Presidente Vargas,

Porto Alegre CEP 90035-074, Brazil; criskopacek@gmail.com

⁷ Graduate Program in Molecular Biology Applied to Health, Universidade Luterana do Brasil (ULBRA),

Canoas CEP 92425-020, Brazil

* Correspondence: amit@pandeylab.org

† These authors contributed equally to this work.

Citation: Prado, M.J.; Singh, S.; Ligabue-Braun, R.; Meneghetti, B.V.; Rispoli, T.; Kopacek, C.; Monteiro, K.; Zaha, A.; Rossetti, M.L.R.; Pandey, A.V. Characterization of Mutations Causing CYP21A2 Deficiency in Brazilian and Portuguese Populations. *Int. J. Mol. Sci.* **2022**, *23*, x. <https://doi.org/10.3390/xxxxx>

Academic Editors: Emil Alexov

Received: 07 December 2021

Accepted: 26 December 2021

Published: date

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Deficiency of 21-hydroxylase enzyme (CYP21A2) represents 90% of cases in congenital adrenal hyperplasia (CAH), an autosomal recessive disease caused by defects in cortisol biosynthesis. Computational prediction and functional studies are often the only way to classify variants to understand the links to disease-causing effects. Here we investigated the pathogenicity of uncharacterized variants in the CYP21A2 gene reported in Brazilian and Portuguese populations. Physicochemical alterations, residue conservation, and effect on protein structure were accessed by computational analysis. The enzymatic performance was obtained by functional assay with the wild-type and mutant CYP21A2 proteins expressed in HEK293 cells. Computational analysis showed that p.W202R, p.E352V, and p.R484L have severely impaired the protein structure, while p.P35L, p.L199P, and p.P433L have moderate effects. The p.W202R, p.E352V, p.P433L, and p.R484L variants showed residual 21OH activity consistent with the simple virilizing phenotype. The p.P35L and p.L199P variants showed partial 21OH efficiency associated with the non-classical phenotype. Additionally, p.W202R, p.E352V, and p.R484L also modified the protein expression level. We have determined how the selected CYP21A2 gene mutations affect the 21OH activity through structural and activity alteration contributing to the future diagnosis and management of CYP21A2 deficiency.

Keywords: 21-hydroxylase deficiency; congenital adrenal hyperplasia; CYP21A2; functional characterization

1. Introduction

Congenital adrenal hyperplasia (CAH) is an autosomal recessive disease caused by defects in steroid biosynthesis [1]. More than 90% of reported CAH cases are due to 21-hydroxylase (CYP21A2) deficiency (OMIM # 201910). The CYP21A2 is a member of the cytochrome P450 superfamily and has 495 amino acids forming thirteen α -helix (A-M) and nine β -sheets [2,3]. This protein is located in the endoplasmic reticulum of the adrenal cortex and has a role in both the glucocorticoid and mineralocorticoid biosynthesis by the

hydroxylations of 17-hydroxyprogesterone into 11-deoxycortisol and progesterone into 11-deoxycorticosterone, which are then converted into cortisol and aldosterone (Figure 1) [1]. Therefore, defects in CYP21A2 affect both the mineralocorticoid and glucocorticoid biosynthesis, besides the increase in sex steroids biosynthesis due to changes in the steroidogenesis pathway by elevated levels of sex steroid precursors [4,5].

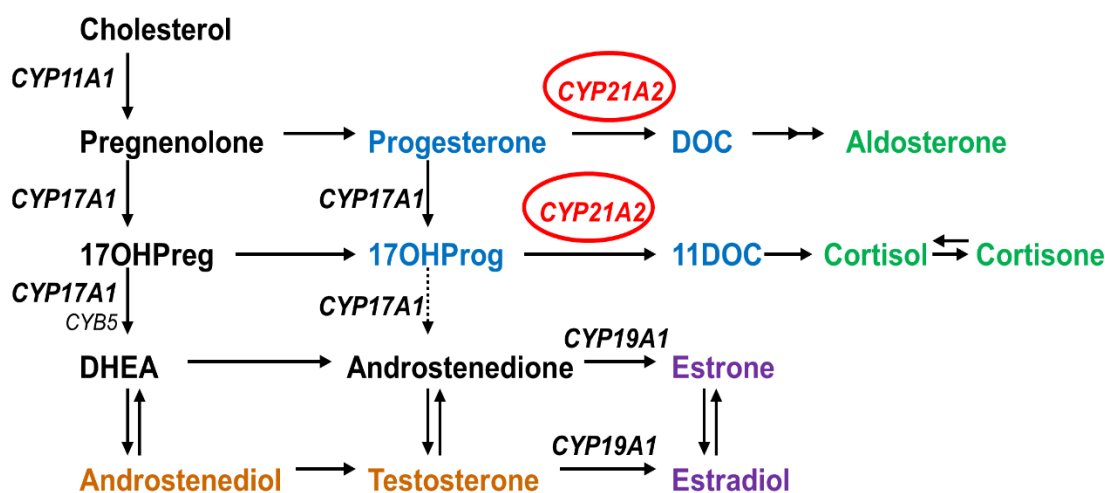


Figure 1. Role of CYP21A2 in human steroid biosynthesis. CYP21A2 is a heme containing cytochrome P450 protein that requires cytochrome P450 reductase as a redox partner for its metabolic reactions [6]. Together with CYP11A1 in mitochondria and CYP17A1 and CYP19A1 in the endoplasmic reticulum, CYP21A2 directs the biosynthesis of steroids.

Glucocorticoid and mineralocorticoid levels determine the CYP21A2 deficiency phenotype, which can be (1) salt wasting, when both types of steroids are not produced, causing the adrenal crisis by electrolytic deregulation with infant mortality risks and severe virilization by elevated sex steroids; (2) simple virilizing, when there is some residual synthesis of both glucocorticoid and mineralocorticoid that is enough to prevent the adrenal crisis; or (3) non-classical (mild form) when just the glucocorticoid synthesis is partially affected and is linked to hyperandrogenism and mild late-onset CAH. The first two are the classical forms of CAH, which have a worldwide incidence ranging from 1:14,000 to 1:18,000 live births [7]. The non-classic form has a frequency around 1:100 to 1:1000 [8]. The frequencies of CAH vary with ethnicity and geographic population groups. However, the description of SW vs. SV, SV vs. NC, etc. is rather arbitrary, hence caution must be employed when following [9]. The diagnosis of CYP21A2 deficiency is confirmed by steroid profile, mainly 17OHP in the first screening, which becomes elevated [1,7]. However, 17OHP can be altered by a deficiency in other enzymes in the steroidogenic pathway, in premature infants, and in unrelated diseases causing physiologic stress [7,8,10–15]. Therefore, the molecular diagnosis is essential for the confirmation of complex cases and follow-up management of asymptomatic CAH, avoiding unnecessary treatment, along with genetic counseling [11,16,17].

The *CYP21A2* gene is located on the short arm of chromosome 6 (6p21.3), situated 30 kb apart from its pseudogene (*CYP21A1P*). These genes share 98% sequence identity for the exons and 96% among the introns [18]. Besides that, 95% of pathogenic variants of *CYP21A2* originate in recombination events [3]. More than 1300 variants in the *CYP21A2* have been reported in the human gene mutation database (HGMD), and more than 200 of these are described to affect human health. With high variability between different ethnic groups and single nucleotide variants (SNVs), missense and nonsense account for half of the total variations [19,20].

CYP21A2 variants are classified according to the impact on the enzyme activity. Group Null consists of deletions or nonsense variants that critically affect the enzyme activity, resulting in a complete loss of function due to altered enzyme stability, steroid or heme binding, and membrane anchoring. The most common variants are thirty kilobase deletion (30-kb del), eight base pairs, Cluster E6 (p.I237N, p.V238E, p.M240K), p.Q319X, p.R357W, and p.L307fs [3]. Group A is composed of a variant in which the enzyme activity is minimal, around 0-1%. This group is represented by an intron variant IVS2-13A/C>G, which is created by an additional splice acceptor site causing retention of 19 intronic nucleotides of the intron 2 [21,22]. Homozygous or compound heterozygote variants with the null group are often associated with the salt-wasting form, but approximately 20% of cases have a simple virilizing phenotype [3]. Group B has a residual activity of 1–10%, which is enough to prevent adrenal crisis. The p.I173N variant is representative of this group and is associated with the simple virilizing form of 21-hydroxylase deficiency [23]. Finally, group C has an enzyme activity of about 20-60% and is associated with the mild form of CAH. The variants common in this group are p.V282L, p.P454S, and p.P31L, which show phenotype variability [24].

The biochemical correlation for CYP21A2 activity works well for the CAH diagnosis. However, external factors or a combination of genetic diseases can change the steroid levels, making genotype elucidation an important tool for patient management [11,17]. In general, there is a good genotype-phenotype correlation for CAH, and the elucidation of new variants found in each population is important to improve the correct treatment [25,26]. Initially, the bovine CYP21A2 structure was used as a template to elucidate the impact of variants damage from 2011 to 2015 (PDB ID 3QZ1), but the human CYP21A2 crystal structures have recently become available. Human CYP21A2 is deposited in the RSCB Protein Data Bank (PDB) under entries 4Y8W and 5VBU, with two different steroid ligands (progesterone in 4Y8W and 17-hydroxyprogesterone in 5VBU). The human structure has just one steroid-binding site, which is different from the bovine enzyme, which has two catalytic sites [2,27].

This study investigated the uncharacterized SNVs in the CYP21A2 gene through computational and functional analysis, establishing a correlation with the residual enzyme activity and a possible CAH phenotype. We selected six missense variants in the CYP21A2 gene (p.P35L, p.L199P, p.W202R, p.E352V, p.P433L, and p.R484L) that were previously reported in the Brazilian or Portuguese populations without previous functional characterization. To determine how these variants impair the CYP21A2 enzyme, we performed a detailed analysis based on the chemical changes due to amino acid alterations, residue conservation, and effect on protein structure. In silico predictions using human CYP21A2 structure (PDB ID 4y8w) were used to understand the structural damage, while the in vitro analysis using recombinant protein expression and enzyme kinetics was performed to determine the impact on protein function.

2. Results

2.1. Variant Collection and Description

We selected six missense variants that had patient genotype and phenotype available in our literature review after the initial screening described in methods. Four of these variants were found in Brazilian (p.P35L, p.L199P, p.E352V, and p.R484L) and two in Portuguese (p.W202R and p.P433L) populations. Genetic and clinical data describing the carriers of the selected SNVs are summarized in Table 1. These data were collected from the original papers.

Table 1. Genetic and clinical features of the subjects carrying the selected SNVs (bold) in the CYP21A2 gene. All data are from the original papers. Normal values for 17 α -hydroxyprogesterone (17OHP) are <30 nmol/L, serum sodium between 132–142 mmol/L and potassium 3.6–6.1 mmol/L. LGC, large gene conversion; 30-kb del, large deletion from CYP21A1P to CYP21A2 gene; nd, non-determined; DSD, ambiguous/atypical genitalia; AC, adrenal crisis; M, male; F, female. ^a Age at diagnosis, ^b non-diluted measurements (reference value 5.97 nmol/L), ^c after ACTH stimulation.

Genotypes		Phenotype	Clinical Data	Sex	17OHP (nmol/L)	Na+/K+ (mmol/L)	Age ^a	Reference
Allele 1	Allele 2							
P35L, H63L, 30-kb del	LGC	SW	vomiting, dehydration	M	49.3	116/9.2	2.8 m	[28] Coeli <i>et al.</i> (2010)
P35L, H63L, 30-kb del	Q319X	SW	vomiting, dehydration	M	>25 ^b	120/6.8	13 m	[28] Coeli <i>et al.</i> (2010)
P35L, H63L, 30-kb del	c.920_921insT	SW	DSD (Prader IV)	F	>6 ^b	119/5.3	15 d	[28] Coeli <i>et al.</i> (2010)
P35L, H63L, 30-kb del	R357W	SW	vomiting, dehydration, AC	M	>200 ^b	119/9.7	28 d	[28] Coeli <i>et al.</i> (2010)
L199P	normal	ND	DSD (Prader I)	F	408.5	nd		[29] Silveira <i>et al.</i> (2009)
W202R	LGC	SW	DSD, salt wasting	F	nd	nd	<2 m	[30] Santos-Silva <i>et al.</i> (2019)
E352V	E6 cluster	SV	nd	F	nd	nd	nd	[16] De Carvalho <i>et al.</i> (2016)
E352V	IVS-13A/C > G	SW	nd	M	nd	nd	nd	[16] De Carvalho <i>et al.</i> (2016)
E352V	IVS-13A/C > G	SW	nd	M	nd	nd	nd	[16] De Carvalho <i>et al.</i> (2016)
E352V	G425S	SV	nd	M	nd	nd	nd	[16] De Carvalho <i>et al.</i> (2016)
P433L	P454S	NC	nd	F	42.7 ^c	nd	17 y	[31] Carvalho <i>et al.</i> (2012)
R484L, IVS-13A/C > G	Del CYP21A2	SW	DSD (Prader III-IV)	F	1537.2	134/4.9	<4 m	[29] Silveira <i>et al.</i> (2009)
R484L, IVS-13A/C > G	Q319X	SW	undefined	M	1358.7	130/6.2	<9 m	[29] Silveira <i>et al.</i> (2009)

The variant p.P35L was identified in four Brazilian patients as compound heterozygous with other mutations (Table 1). One allele presents a mutation with symptoms of severe enzyme damage, while the other alleles carried three mutations: p.P35L, p.H63L, and 30-kb del. It was concluded that the p.P35L origin was not from the same genetic event of the p.H63L and 30-kb del, as it was not present in the pseudogene screening. Therefore, we analyzed this variant alone.

The variant p.L199P was found as heterozygous in a female patient with mild clitoromegaly atypical genitalia and elevated 17OHP levels detected in the Brazilian newborn screening program (Table 1). According to the authors, the child remained asymptomatic at 3.3 years old, and the clitoromegaly and the high 17OHP levels detected in the screening were likely due to the premature birth. The impact of this new variant has been unknown, as it was not possible to infer the consequence once it was identified in heterozygosis.

The variant p.W202R was found as compound heterozygous with large gene conversion in a Portuguese female (Table 1). This patient presented atypical genitalia and salt wasting, being diagnosed with classical CAH at < 2 months old.

The variant p.E352V was identified in a compound heterozygous state in four classical CAH patients from Brazil (Table 1), two of them with the same intronic mutation, IVS2-13A/C>G heterozygous with p.E52V. These two patients showed the classical CAH form with salt-wasting. One female presented p.R484L as compound heterozygous with Cluster E6. These patients presented the simple virilizing CAH form. The latter has the p.R484L and p.G425S, which resulted in the simple virilization form of CAH.

The variant p.P433L was identified in a Portuguese female who presented it as compound heterozygous with the mild mutation p.P454S (Table 1). This patient was diagnosed at 17 years old with a nonclassical form of CAH after ACTH stimulation, which indicated the 17OHP elevation characteristic of that form of CAH.

The variant p.R484L was found as compound heterozygous in two Brazilian patients (Table 1). Both present classical CAH form, with salt-wasting and high 17OHP levels detected during newborn screening. One allele in both cases has p.R484L together with a splice variant IVS2-13A/C > G. The second allele in one patient was a CYP21A2 deletion and in the other p.Q319X.

2.2. Computational Characterization Indicated the Structural Impact of the SNVs

We performed a screening with five predictive tools that have different approaches, PolyPhen-2, SNAP2, MutPred2, Meta-SNP, and PredictSNP. Results obtained for the six variants chosen in this work are shown in Table 2. The variants p.L199P, p.E352V, and p.R484L were predicted to damage the protein by all predictor tools, while p.P35L had damage predicted by four tools, p.W202R by three, and p.P433L by two.

Among the variants with known activity, the neutral variants (p.R103K, p.D183E, and p.S269T) agreed with all tools. However, among the three variants known to cause damage, we had different results. The two severe variants (p.I173N and p.R427H) were predicted correctly by all tools; however, the mild (p.V282L) variant was wrongly classified by all of them. The second screening about the protein stability of all variants presented a variation of ΔG , from -1.47 kcal/mol (p.R484L) to 1.23 kcal/mol (p.E352V) (Tables 2 and 3).

Table 2. Prediction of the possible impact on the structure and/or function of CYP21A2 by genetic variants. Seven variants were used to validate the method, with four polymorphisms (●) and three with a known negative impact (↓). ^a SNV nomenclature according to UniProt ID Q16874-1. Scores ≥ 0.5 by PolyPhen-2, MutPred2, and Meta-SNP indicate protein damage. SNAP2 scores >50 indicate a strong signal for effect, between 50 and -50 weak signal and <-50 strong signal for neutral effect.

SNV ^a	In Silico Tools										
	dbSNP rs#	PolyPhen-2	Score	SNAP2	Score	MutPred2	Score	Meta-SNP	Score	PredictSNP	Enzyme Activity % of Control
● R103L	rs6474	Benign	0.023	Neutral	-47	Neutral	0.191	Neutral	0.310	Neutral	120
● D184E	rs397515531	Benign	0.000	Neutral	-85	Neutral	0.260	Neutral	0.412	Neutral	100
● S269T	rs6472	Benign	0.016	Neutral	-93	Neutral	0.076	Neutral	0.340	Neutral	103
↓ I173N	rs6475	Damage	1.000	Effect	70	Pathogenic	0.838	Disease	0.811	Deleterious	1.1
↓ V282L	rs6471	Benign	0.273	Neutral	-76	Neutral	0.084	Neutral	0.344	Neutral	16.4
↓ R427H	rs151344504	Damage	1.000	Effect	82	Pathogenic	0.821	Disease	0.827	Deleterious	0.5
P35L	<u>rs200648381</u>	Damage	0.988	Effect	13	Pathogenic	0.676	Disease	0.624	Neutral	13
L199P		Damage	0.997	Effect	68	Pathogenic	0.875	Disease	0.711	Deleterious	10.3
W202R		Damage	1.000	Effect	45	Pathogenic	0.577	Neutral	0.451	Neutral	1.2
E352V		Damage	0.693	Effect	90	Pathogenic	0.919	Disease	0.937	Deleterious	1.1
P433L	rs751456004	Benign	0.402	Effect	15	Pathogenic	0.614	Neutral	0.372	Neutral	7.5
R484L		Damage	1.000	Effect	81	Pathogenic	0.690	Disease	0.748	Deleterious	3.4

Table 3. Summary of structural and functional data obtained for the six variants studied shows that all of them can be related to CYP21A2 deficiency. ^a SNV nomenclature according to UniProt ID Q16874-1. ^b Two hydrogen bonds with E355. H, hydrogen-bound.

SNV ^a	Structure/Protein Localization	Physicochemical Properties		$\Delta\Delta G$ (kcal/mol)	Enzyme Activity (Mean \pm SD)	The Molecular Mechanism (Hypothesis)	Mutation Group
		WT	MUT				
P35L	N-term coil/Region of the membrane protein orientation	Polar residue with a ring at the N-term	Hydrophobic; Short side chain; Folding interactions	-0.58	13.2 \pm 1.56%	Disturbs the protein orientation in the membrane	C
L199P	F-helix/Close to the active site; α -helix stabilization	Hydrophobic; Short side chain; Folding Interactions; H: I195 and S203	Polar residue with a ring at the N-term; H: S203	0.79	10.3 \pm 0.33%	Breaks the secondary structure close to the active site	C
W202R	Turn of the F-helix/close to the active site and steroid-bound residue (R234)	Hydrophobic; Large, rigid aromatic group; H: V198	Hydrophobic; Long, flexible, and posit. charged side chain; Ionics-bound; H: V198	1.11	1.2 \pm 0.34%	Positive charge addition causes disorder in heme binding	B
E352V	K-helix/ERR-triad associated with the heme group	Hydrophobic; Long, slightly flexible side chain; Strongly neg. charged; Ionics-bound; Fix metal iron; H: A348, R355 ^b , L356 and W406	Hydrophobic; Short side chain; Folding interactions; H: A348 and L356	1.23	1.1 \pm 0.11%	Disorder of the ERR-triad decreases heme-binding and stability	B
P433L	L-helix/Adjacent to essential residues for the heme bond	Polar residue with a ring at the N-term; H: A435	Hydrophobic; Short side chain; Folding interactions; H: A435	0.39	7.5 \pm 0.67%	Changes the natural orientation of heme-binding residues R427 and C429	B
R484L	C-term coil/Hydrophobic cluster with ionic connection	Hydrophobic; Long, flexible, and posit. charged side chain; Ionics-bound; H: Q482, A449, and M486	Hydrophobic; Short side chain; Folding interactions; H: A449	-1.47	3.4 \pm 0.8%	Loss of hydrophobic organization at the C-term region	B

2.3. Amino Acid Chemical Properties, Conservation, and Structural Damage

We analyzed the amino acid chemical properties to assess the characteristics of each exchange and the conservation of that residue across species to determine how it has been retained during the evolution of CYPs P450. We worked with 200 CYP 450 sequences homologous to human CYP21A2. The structural damage was characterized by Gibbs's free energy and gain/loss of hydrogen bonds.

Proline at amino acid position 35 is located in a coil on the protein surface (Figure 2 and Figure 3). This residue shows a high conservation score, with only two residues found on homologous sequences (proline and guanine) (Figure 4 and Table 4). The variant p.P35L has the amino acid properties changed from a polar and uncharged to a nonpolar and aliphatic residue, increasing the structural stability, $\Delta\Delta G$ -0.58 kcal/mol (Table 3). PolyPhen-2, SNAP, Meta-SNP, and MutPred2 predicted the effect of this mutation on the protein stability and functionality. Besides that, MutPred2 predicts two structural effects: gain of helix and alteration of transmembrane features. PredictSNP classified this variant as neutral (Table 2).

Leucine at position 199 is localized at α -helix F, close to the CYP21A2 catalytic site <4 Å from heme, where it forms two hydrogen-bonds, one with p.S203 (on the loop between α -helix F and F') and one with p.I195, on α -helix F (Figure 3 and Table S1). This residue shows mild conservation, being found in 10 different sequences at this position (Figure 4, Table 4). The variant p.L199P changes a nonpolar and aliphatic residue to a polar and uncharged residue, losing the hydrogen bond with p.I195 (Table S2). All predictor tools showed that this mutation causes damage to the protein, and MutPred2 predicted an alteration of coiled-coil and transmembrane regions; $\Delta\Delta G$ was 0.79 kcal/mol.

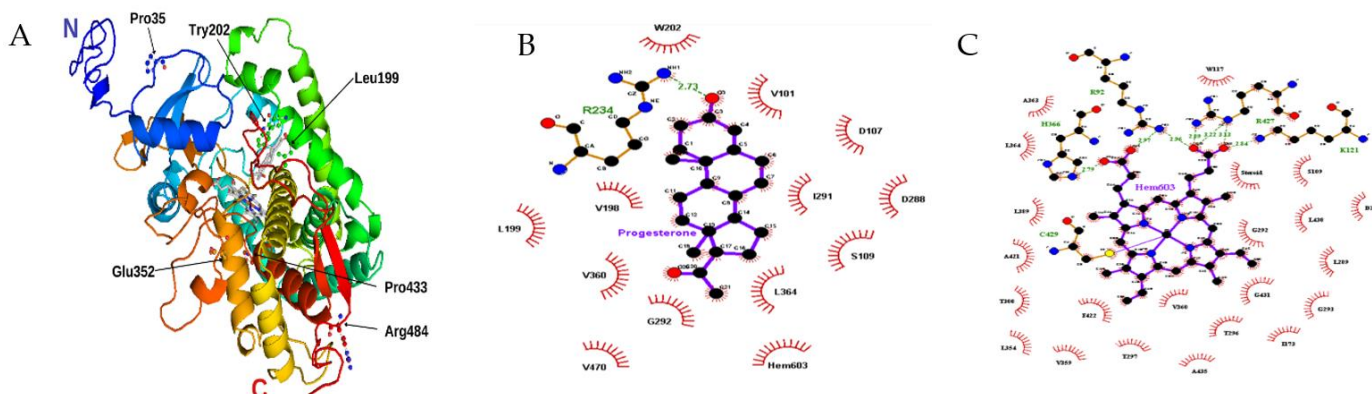


Figure 2. Structural model of CYP21A2. (A) We built the missing extremity based on PDB ID 4Y8W with I-TASSER and visualized by PyMOL. The amino acids mutated in our study are identified with black arrows. Progesterone and protoporphyrin containing Fe^{3+} are in the middle of the structure colored in grey. (B) Contact map with progesterone binding site in 21OH structure. The studied residue p.W202 is shown at the top. (C) Contact map with heme (protoporphyrin containing Fe) binding site in 21OH structure. The residues p.C429 and p.R427 are near the studied residue p.P433. Heme and progesterone contact maps were built with the structure of CYP21A2 (PDB # 4Y8W). Residue ligand contact maps were generated with LigPlot+ v.2.2.4 software using as maximum hydrogen -acceptor/-donor distance 2.7 Å and 3.35 Å (green line), respectively, while for non-bonded, the minimum contact distance was 2.9 Å and the maximum 3.9 Å.

Tryptophan at position 202 is also located <4 Å from the CYP21A2 catalytic site, on the α -helix F with a hydrogen bond with p.V198 (Figure 3). This residue presented a low conservation score, being highly variable between its homologous sequences (Figure 4 and Table 4). The variant p.W202R changes a hydrophobic residue to a hydrophilic and positively charged residue, decreasing the CYP21A2 structural stability with a $\Delta\Delta G$ of 1.11 kcal/mol, but without losing the hydrogen bond (Table S2). Two predictor tools classified that residue replacement as neutral (Meta-SNP and PredictSNP), while the three others

predicted damage to the protein. An altered coiled-coil region was predicted by MutPred2.

Glutamate at position 352 is located in α -helix K, presenting five hydrogen bonds, four on the same helix (one with p.A348 and p.L356 and two with p.R355) and one with a coil (p.W406) (Figure 3 and Table S1). This residue is highly conserved across species, and there is no other variation on CYP21A2 homologs (Figure 4 and Table 4). The variant p.E352V changes a negatively charged residue to a nonpolar aliphatic residue. The stability of the CYP21A2 structure was increased, and the changed residue lost three hydrogen bonds with W406 and p.R355 (Table S2). All the predictor tools showed that p.E352V results in protein damage (Table 2). The MutPred2 predicted alteration of an interface, loss of allosteric site at p.E352, and altered metal binding.

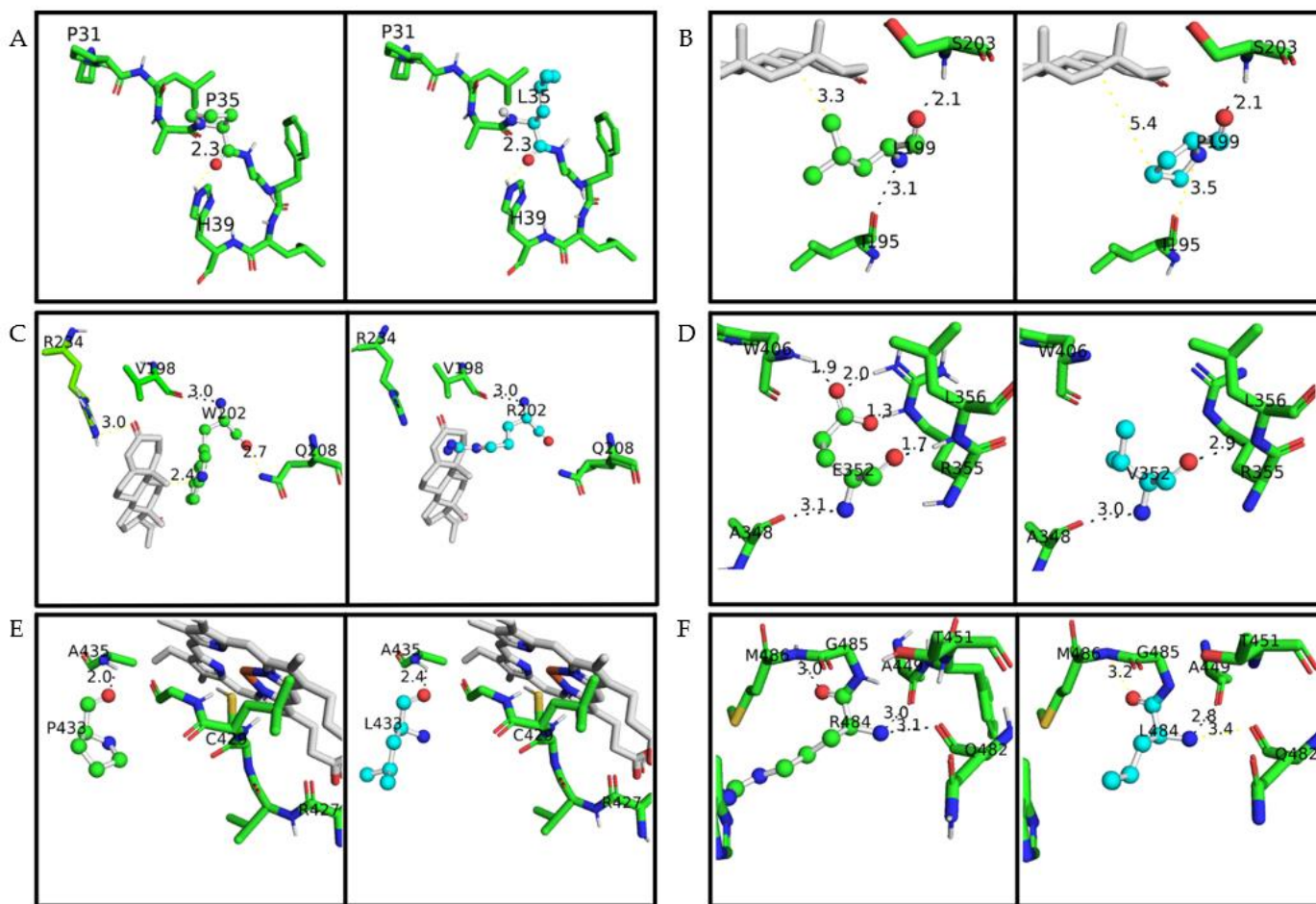


Figure 3. A closeup of the amino acid studied on the CYP21A2 structure. The protein structure is based on PDB ID 4Y8W. Wild-type amino acids are colored in green and mutated in cyan; in both of them, oxygen is red, nitrogen is blue, and sulfur is yellow. The distance measured between the main amino acids connections is with less than 4 Å. Hydrogen bonds are represented with black lines, while any other measurement between atoms is represented with yellow lines. Progesterone and heme are colored grey. (A) p.P35 and p.P35L, (B) p.L199 and p.L199P, (C) p.W202 and p.W202R, (D) p.E352 and p.E352V, (E) p.P433 and p.P433L, (F) p.R484 and p.R484L.

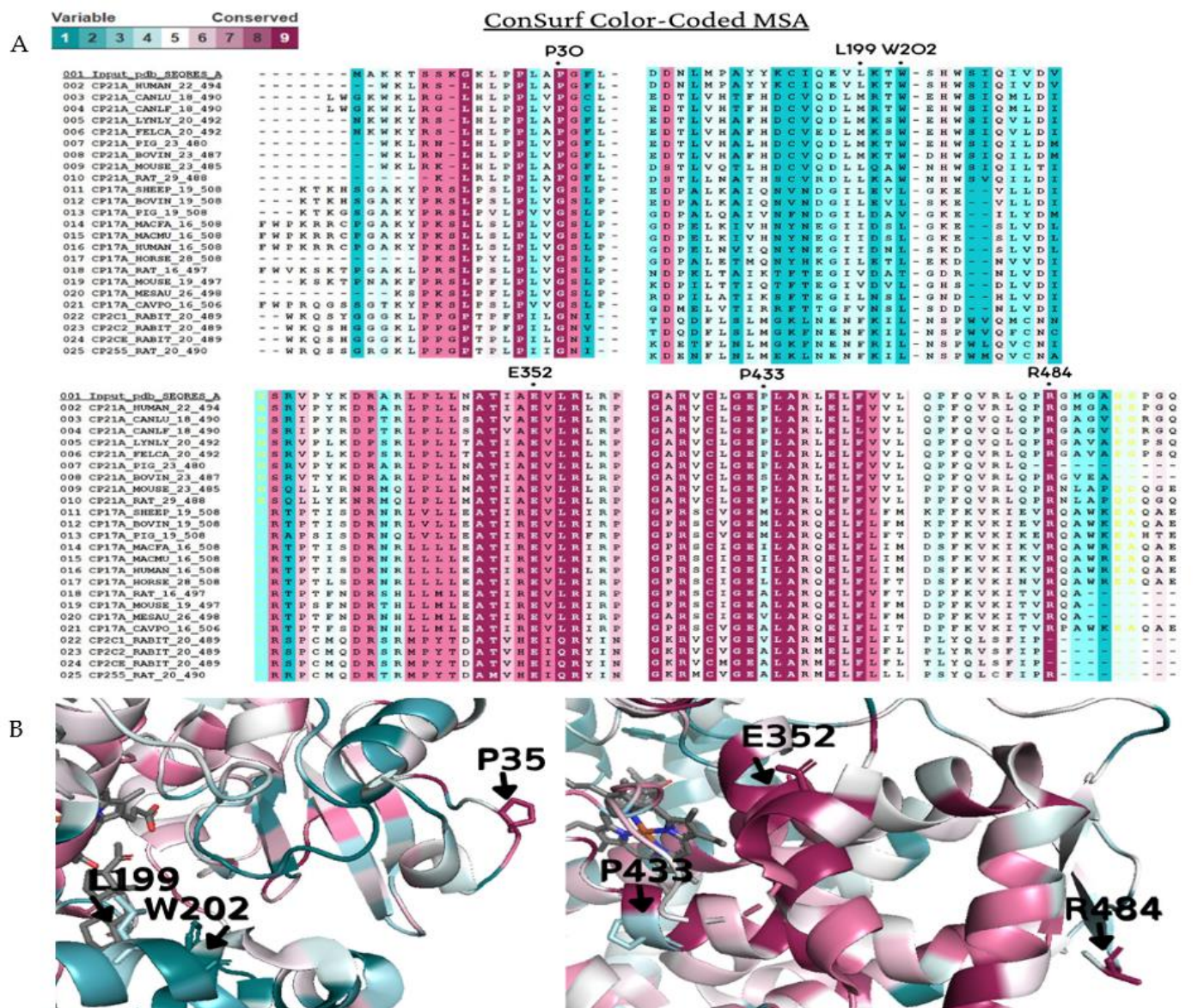


Figure 4. Amino acid conservation of the human CYP21A2. A CYP21A2 structure (PDB # 4Y8W) structure was used as a reference for ConSurf analysis. Multiple sequence alignment was built using CLUSTALW with 200-cytochrome P450 sequences homologous to the human protein 21-hydroxylase. The species considered for final alignment were from *Homo sapiens*, *Bos taurus*, *Canis lupus familiaris*, *Cavia porcellus*, *Capra hircus*, *Felis catus*, *Gorilla gorilla gorilla*, *Lynx lynx*, *Mesocricetus auratus*, *Macaca fascicularis*, *Macaca mulatta*, *Mus musculus*, *Oryctolagus cuniculus*, *Ovis aries*, *Rattus norvegicus* and *Sus scrofa*. (A) A representative part of the alignment from MSA. The specific positions of the six wild-type residues analyzed in this work are marked with a dot (•). (B) CYP21A2 structure with black marks on the six wild-types residues on the protein structure. Residues p.P35, p.E352, and p.R484 showed to be conserved, while p.L199, p.W202, and p.P433 are variable. Color-code: from pink (conserved) to cyan (variable).

Table 4. ConSurf amino acid conservation score. The score is from 1 (variable) to 9 (conserved). CI: Confidence interval of the score. ^a PDB ID 4Y8W. Details of any known variations in the amino acid positions studied are given in Table S3 (Supplementary Material).

SNV ^a	Conservation Score	CI	Residue Variety
P35	9	9 to 8	L, P, G
L199	3	4 to 2	T, N, L, V, F, I, A, S, G, M
W202	1	2 to 1	T, E, R, W, G, C, M, L, A, I, N, Q, D, V, F, S
E352	9	9 to 9	E
P433	3	4 to 2	R, K, G, M, L, V, H, A, S, I, T, P, E, Q

R484	9	9 to 9	R, T, L
-------------	---	--------	---------

Proline at position 433 is located at α -helix M, where it has a hydrogen bond with p.A435 in the same helix (Figure 3 and Table S1). This residue is close to the central heme group, but it is not a conserved residue, as it is found interchangeable as 14 different residues among the CYP21A2 homologous group (Figure 4 and Table 4). However, this indicates the evolution of this residue across species with different roles and different redox partners. The exchange of a polar and uncharged amino acid with a nonpolar and aliphatic showed $\Delta\Delta G$ 0.39 kcal/mol (Table 3). The SNAP2 predictor tool showed damage in the protein with the variant and MutPred2 a gain of helix and loss of catalytic site at p.E432; however, the other three predictors indicated a neutral effect (Table 2).

Arginine at position 484 is located on the protein surface in a coil that makes a cluster with nine other amino acids. This residue has three hydrogen bonds, one with p.M486 in the same coil; one with p.Q482, located at stand β 9; and one with p.A449, which makes the connection between α -helix M and β 8 (Figure 3 and Table S1). This residue showed the highest conservation score, with just two other amino acids found in the homology analysis (Figure 4 and Table 4). The variant p.R484L changes a positively charged polar residue to a nonpolar and uncharged residue, losing the two hydrogen bonds with p.M486 and Q482 (Table S2). This amino acid exchange showed damage by all predictor tools used, and stability increased with a $\Delta\Delta G$ of -1.47 kcal/mol (Tables 2 and 3). Loss of intrinsic disorder was predicted by MutPred2, as well.

2.4. Functional Testing for the Validation of In Silico Results

HEK293 cells were transfected with plasmids expressing CYP21A2 WT or variants p.P35L, p.L199P, p.W202R, p.E352V, p.P433L, p.R484L, p.I173N, and p.V282L as two controls with known activity (~2% and 18–60%, respectively) [4,23,24]. The CYP21A2 activity was quantified for both WT and variants. Using the TLC analysis, we could assess the conversion ratio of progesterone to 11-deoxycorticosterone and compare the activity of each variant with the WT enzyme (Figure 5A). Variants p.W202R and p.E352V showed only residual conversion (<2% enzyme activity), similar to p.I173N, which is associated with the classical form of CAH. Variants p.P35L and p.L199P showed partial activity, with the catalytic activities being 13.2% and 10.3% of the WT, similar to the p.V282L variant, which is associated with a nonclassical form of CAH. Variants p.P433L and p.R484L presented activity between the two controls, being 7.5% and 3.4% of the WT activity, respectively (Figure 5C).

To determine if the reduction in activity was associated with the decrease in protein expression, we quantified CYP21A2 protein in WT and variants using western blot assay (Figure 5B). Normalized by β -actin expression, we found the expression level was less than 50% of the WT for the p.P35L, p.V282L, p.E352V, and p.R484L variants, while p.W202R had 119% of WT. Therefore, we also calculated the specific activity of each variant compared to WT by dividing the activity obtained from the TLC results by the CYP21A2 expression level (Figure 5C). The specificity activity for p.W202R, p.E352V, and p.R484L present a significant difference ($p < 0.05$) when compared with the non-protein normalized activity. The variant p.W202R had a decrease in activity to <1%, while p.E352V and p.R484L had an increase in activity by twofold and threefold, respectively. Altogether, our results indicated that all variants tested significantly impacted the activity of CYP21A2. The variants p.W202R, p.E352V, and p.R484L also impacted the protein expression (Figure 5C).

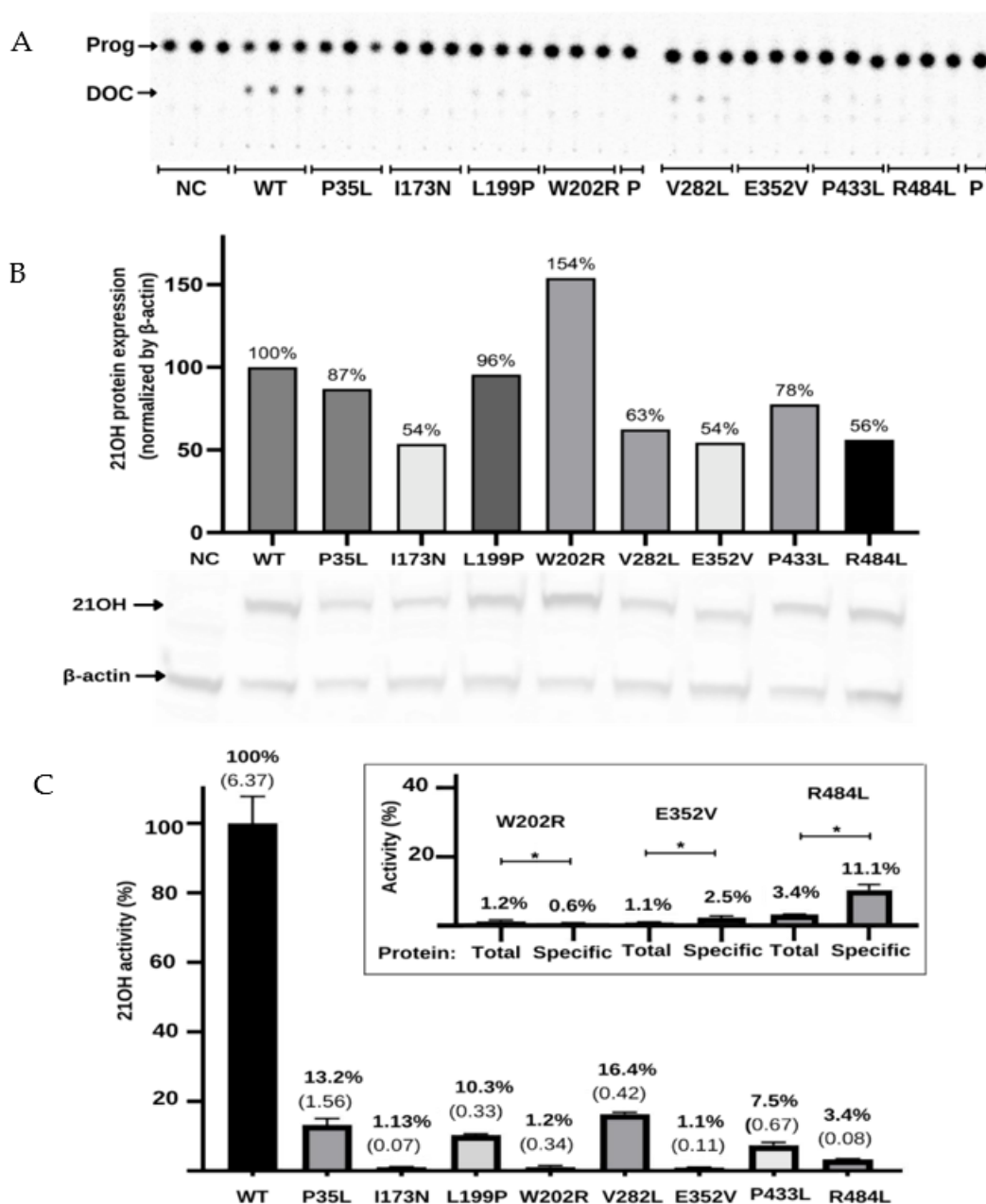


Figure 5. Functional characterization of the variants in the *CYP21A2* gene by relative steroid conversion of progesterone (Prog; P) to 11-deoxyprogesterone (DOC). The activity was obtained using 21OH wild-type (WT) and mutated expressed in the HEK293t cell where we measured the steroid conversion by the percentage of radioactivity in DOC to the whole sample. (A) Thin-layer chromatography (TLC) shows that all variants reduce the activity of 21-hydroxylase (21OH). (B) Semiquantitative 21OH expression level obtained by western blot with anti-flag antibody for the 21OH (53 KDa) and normalized by β -actin (42 KDa) expression with anti- β -actin antibody. Negative control (NC) presents the basal HEK293 cells. (C) The 21OH activity expressed accordingly with TLC spots densitometry and related to the WT. Bars represent the standard error from three samples. Inside the box is plotted the activity (total) and the specific activity (specific) of the three variants that showed statistically significant difference ($* p < 0.05$) between these values. The specific activity was obtained by dividing the enzyme activity by the 21OH protein estimated by western blot. All results were analyzed on GraphPad software. $* p$ value < 0.05

2.5. Kinetic Analysis of CYP21A2 Variants

The apparent kinetic constant revealed saturation for the 21OH WT with the Michaelis-Menten constant (K_m) of 1.57 μM for progesterone and an apparent maximal reaction velocity (V_{max}) of 0.360 $\text{nmol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ (Figure 6 and Table 4). The apparent saturation was a similar rate of the WT for p.P35L (K_m 2.06 μM) and p.P433L (K_m 1.91 μM), but it was six times higher for p.L199P (K_m 10.24 μM), indicating that p.L199P decreases the protein affinity for the substrate progesterone. The apparent V_{max} was lower than the WT for p.Pro35Leu (0.085 $\text{nmol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$), p.L199P (0.252 $\text{nmol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$), and p.P433L (0.055 $\text{nmol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$). The apparent catalytic efficiencies of the mutated proteins were also lower than those of WT, and p.P35L had catalytic efficiency of 18% compared to the WT, p.L199P of 11%, and p.P433L of 13% (Table 5, Figure 6).

Our results show that the p.L199P variant, which is located close to the catalytic site, affects the substrate binding. In contrast, p.P35L and p.P433L showed inhibition of the enzyme activity through a decrease in reaction velocity.

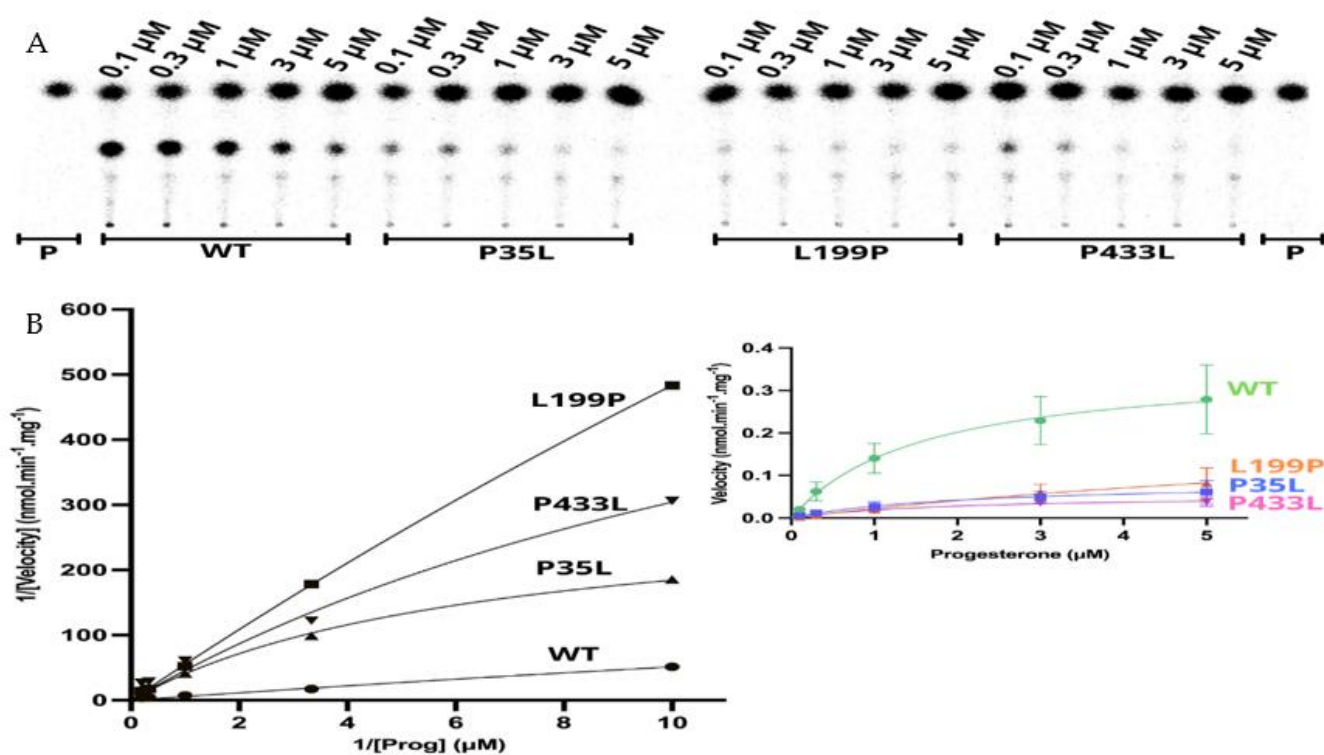


Figure 6. Kinetics assay shows a decrease in reaction velocity for three variants which have still presented activity in the functional assay. (A) Representative thin layer chromatography (TLC) from three replicates shows the CYP21A2 conversion of progesterone (Prog) (0.1, 0.3, 1, 3, and 5 μM) into 11-deoxycorticosterone (DOC). (B) The left-hand plot presents the linear plots of enzymatic activity of 21OH WT and mutations between $1/\text{Velocity}$ against $1/\text{Prog}$ (progesterone concentration) for the conversion of progesterone to DOC. Additionally, the right-hand plot presents the Michaelis Menten nonlinear curve with velocity reactions against the substrate progesterone used to obtain V_{max} and K_m values.

Table 5. Apparent kinetic constants and catalytic efficiency were calculated from three independent experiments.

	Wild-Type	P35L	L199P	P433L
V_{max} ($\text{nmol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$)	0.360	0.086	0.252	0.055
K_m (μM)	1.57	2.07	10.24	1.91
V_{max}/K_m	0.229	0.041	0.025	0.029

3. Discussion

We have presented a detailed investigation of the molecular and functional damage caused by six missense variants of the CYP21A2 gene. These variants were selected from a literature review of uncharacterized missense variants. We performed *in silico* analysis and functional assays to assess the effect of these CYP21A2 variants. We showed the analysis of two variants with changes in positions close to the catalytic site, p.L199P and p.W202R. Both showed a low conservation score among mammalian CYPs P450. However, looking specifically for the CYP21A2, residue 199 still had variability while 202 had high conservation. The p.W202R changed a heavy and hydrophobic residue to a hydrophilic and positively charged residue. This exchange at 202 abolished the enzyme activity (1.1% of WT), which is in agreement with the severe CAH (SW) form described for the patient [30]. In the other case, the leucine replacement by proline at position 199 broke the hydrogen bond with a residue on the same helix, destabilizing the conformation of helix F. Previous studies with helical substitution of leucine to proline at position 168 (helix E), 262 (helix H), and 322 (helix J) showed a break of the helix, as well, and association with the severe form of CAH [4,25,32]. In our study, this alteration decreased the affinity of the enzyme for the substrate by about sixfold, making the enzyme activity for the L199P variant of CYP21A2 drop to 10% of the WT.

We analyzed two CYP21A2 variants by looking at the protein surface: one at the N-terminal, residue 35, and the other on the C-terminal, residue 484. We showed that the proline at the 35 position has a high conservation score, even being on the protein surface without strong interaction with closer residues. This variant decreases the reaction velocity four times due to the faster enzyme saturation, V_{max} 0.08552 nmol.min⁻¹.mg⁻¹, compared with WT, V_{max} 0.3604 nmol.min⁻¹.mg⁻¹. The apparent enzyme activity was around 13 % of the WT. That result suggests an important and conserved role of P35 across species. Variant p.P35L was previously described in the Brazilian population in a compound heterozygous state with the other three variants, two being in the same allele and one in the other [28]. All the four patients described by [28] carry two other pathogenic mutations (30-kb del and p.H63L) at the same allele of p.P35L, which by themselves would transcribe a non-functional protein [3,33]. The p.H63L is derived from pseudogene and has known mild activity in the homozygous state; however, it has a synergistic effect when it is associated with another mild mutation and results in a severe phenotype [33]. The 30-kb del is a gene rearrangement event in the RP1-C4A-CYP21A1P-TNXA and RP2-C4B-CYP21A2-TNXB regions, which causes CYP21A2/CYP21A1P chimera [3]. In [28], p.P35L was not presented in any pseudogene analyzed, which indicated that these appear in the same allele from independent genetics events.

The residue studied at the carboxyl-terminus 484 has high conservation, while all the others around this sequence position have mild or low conservation scores. The p.R484 stabilizes other residues through hydrogen bonds with p.M486 in a coil, p.Q482 at stand β 9, and p.A449, a crucial residue between helix M and a sheet β 8 [27]. The variant p.R484L changes a polar and positively charged amino acid to a nonpolar and uncharged residue inside a polar cluster, causing an intrinsic disorder. This variant showed enzyme activity around 3% of the WT when normalized by the specific amount of CYP21A2. However, the activity was significantly lower without the protein normalization, suggesting that this variant generates a lower transcription level or higher protein degradation than the WT. The same effect was observed for p.W202R and p.E352V variants. The p.R484L was found in a compound heterozygous state with the other two variants, one in the same allele and another in the opposite allele [29]. Both of these have a known severe impact on enzyme activity [3,26].

Another residue with a high interaction number is the p.E352. This amino acid is highly conserved, and there is no residue variation at the same position in homologous sequences, showing an important and conserved role in the CYP21A2 protein. The p.E352 forms five hydrogen bonds, stabilizing its helix through four interactions (two with p.A348 and p.L356 and two with p.R355) and stabilizing a network with p.W406 in the L-M loop. The tryptophan at 406 is a backbone residue, while arginine at 355 has an extensive network with glutamine at p.E352 and arginine at 409 in the ERR-triad [25]. This

triad acts to stabilize the three-dimensional structure that allows covalent binding of the heme group [25,34]. The variant p.E352V causes the loss of three hydrogen bonds (p.W406, and p.R355), destabilizing several structural elements. We showed that the CYP21A2 p.E352V decreased the enzyme activity to only 1.1% of the WT, which would be related to the classical form of CAH. De Carvalho *et al.* [16] identified four patients with this variant in the compound heterozygous state. Two of them had a simple virilization form of CAH, while two had the salt-wasting form. In the first form, one patient presented Cluster E6, which abolishes the enzyme activity, and the other had p.G425S, which has less than 2% of enzyme activity [4]. The latter form was detected in two patients with the same intronic variant IVS2-13A/C > G, which mostly results in the salt-wasting form (in around 79% of the cases) and simple virilizing form (in 20% of the cases) [3].

Lastly, the residue p.P433 had a low conservation score looking at the mammalian CYPs P450. Nonetheless, the score was high among the CYP21 group, highlighting the importance and conservation of this residue in this specific group. Proline at 433 is on the first helix M turn, and it has one hydrogen bond with a residue in the same helix M, p.A435, helping with the helix stabilization. Besides that, this residue lies adjacent to the L–M loop, where heme-binding residue p.R427 is located. The p.P433L makes the structure more flexible, modifying the optimal conditions for the heme-binding, as previously described [25]. This variant makes the reaction become saturated six times faster than the WT, with a V_{max} of 0.055 nmol.min⁻¹.mg⁻¹, and decreases the enzyme activity to 7.5%. The p.P433L was identified in the compound heterozygous state with the mild mutation p.P454S in a patient diagnosed with the nonclassical form of CAH [31].

4. Materials and Methods

4.1. Search and Selection of Variants in CYP21A2

We searched Ensembl, HGMD, OMIM, and ClinVar databases for all CYP21A2 variants with an amino acid exchange, excluding nonsense and frameshift variants. From that first list, we excluded the variants with the enzyme damage that has already been characterized. The clinical phenotype associated with the variants was accessed from the original article when it was not available in the databases. Variants without patient information were excluded, as well as when it had been clinically associated with a nonpathogenic form of CAH or it had been found with either homozygous or heterozygous compounds with a pathogenic variant. A second screening was performed to select variants that have been found in either the Brazilian or Portuguese population.

In a second step, we analyzed the degree of amino acid conservation, chemical properties, and structural damage. The first two characteristics were analyzed with the software Polymorphism Phenotyping (PolyPhen-2) [35], SNAP2 [36], Meta-SNP [37], PredictSNP [38], and MutPred2 [39]. To validate the methodology applied, SNPs with the known enzyme activity were used as control. Positive controls were p.R427H (from the null group), p.I173N (from B group), and p.V282L (from C group). Negative controls were p.R103K, p.D183E, p.M240K, and p.S269T.

For the remaining SNPs, we performed a structural impact analysis through the shift of Gibbs free energy ($\Delta\Delta G$) value by STRUM [40]. Gain/loss of hydrogen bonds was calculated on WHAT-IF web (<https://swift.cmbi.umcn.nl/servers/html/index.html> accessed on 05/12/2021). The human CYP21A2 structure used here was deposited on PDB by Pallan P.S., Lei L., and Egli M. [2]. This structure has residues from 28 to 485 of CYP21A2, so we added the missing amino acid with I-TASSER [41] in the N- and C-terminal. The I-TASSER input was the protein sequence RefSeq NM_000500.9 and PDB ID 4y8w, chain A. The CYP21A2 structures with each variant were built with STRUM, and the Gibbs free energy was compared with the wild-type (WT). Structures built with the variants were aligned with the WT using PyMOL (Schrödinger, LLC), where atoms with distance variation were measured. We used the structure-naming scheme of Pallan, *et al.* (2015) [2].

4.2. Conservation Analysis

To access the significance of these variants on the evolutionary conservation, we performed an evolutionary analysis with homologous proteins by ConSurf [42]. The structure of human CYP21A2 deposited under PDB ID 4y8w (chain A) was used as a reference structure, and cytochrome P450 homologous sequences from *Homo sapiens*, *Bos taurus*, *Canis lupus familiaris*, *Cavia porcellus*, *Capra hircus*, *Felis catus*, *Gorilla gorilla gorilla*, *Lynx lynx*, *Mesocricetus auratus*, *Macaca fascicularis*, *Macaca mulatta*, *Mus musculus*, *Oryctolagus cuniculus*, *Ovis aries*, *Rattus norvegicus*, and *Sus scrofa* were used for conservation analysis. A multiple sequence alignment was built using CLUSTALW using the Bayesian method for the calculation of the conservation score. ConSurf scores range from conserved (magenta or nine) to the variable (cyan or one). Homologous sequences were collected from SWISS-PROT with BLAST algorithm (PSI-BLAST E-value 0.0001, four iterations) [43–45].

4.3. Construction of Plasmid and Site-Directed Mutagenesis

Mammalian expression plasmid pcDNA3.1+/C-(K)-DYK carrying the ORF sequence of WT CYP21A2 (NM_000500.7) with a C-terminal DYK (FLAG) tag was purchased from GenScript (Piscataway, NJ, USA). We used this plasmid as a template to create the variants through site-directed mutagenesis. Mutagenesis reactions were performed with a QuikChange II Site-Directed Mutagenesis kit (Agilent Technologies, Santa Clara, USA), following the manufacturer protocols. Oligonucleotides sequences designed with the Quik Change Primer Design program (Agilent Technologies, Santa Clara, CA, USA), are shown in Table 6. Correct nucleotide change was confirmed by direct sequencing on an ABI 3500 Genetic Analyzer (Applied Biosystems, Carlsbad, CA, USA).

Table 6. Oligonucleotides for site-direct mutagenesis. Fwd–forward, Rev–reverse.

SNP	Oligonucleotide Sequence (5'→3')
P35L_Fwd	GTGCAAGAAGCCCAGGGCAAGAGGCCGG
P35L_Rev	CCGCCTCTTGCCCTGGGCTTCTTGAC
I173N_Fwd	CTCACCTGCAGCATCAACTGTTACCTCACCTTC
I173N_Rev	GAAGGTGAGGTAACAGTTGATGCTGCAGGTGAG
L199P_Fwd	CCAGTGGCTCCAGGTTTTTGGCACCTCCTGGATACATTTG
L199P_Rev	CAAATGTATCCAGGAGGTGCCAAAAACCTGGAGCCACTGG
W202R_Fwd	GGACCACTGGCTCCTGGTTTTTAACACCTCCT
W202R_Rev	AGGAGGTGTTAAAAACCAGGAGCCACTGGTCC
V282L_Fwd	GCAGCCATGTGCAGGTGCCCTTCCAGG
V282L_Rev	CCTGGAAGGGCACCTGCACATGGCTGC
E352V_Fwd	GGCGCAGCACCCACGGCGATGGTG
E352V_Rev	CACCATCGCCGTGGTGCTGCGCC
P433L_Fwd	GGCGCGCCAGCAGCTCGCCCAGG
P433L_Rev	CCTGGGCGAGCTGCTGGCGCGCC
R484L_Fwd	CCCCATCCCCAGGGGCTGCAGC

4.4. Cell Transfection and Enzymatic Activity Assay

Transient transfection was performed with the CYP21A2 WT vector and the variants. HEK293 cells were seeded in a six-well plate (6×10^5 cells/well). After 24 h, growth media were replaced, and the cells were transfected using Lipofectamine 2000 (Thermo Fisher, Bedford, MA, USA) and 2.6 μg of the plasmid. Transfected cells were reseeded in a 24-well plate (1.5×10^5 cells/well) after 24 h of the transfection for uniformity of cell population across wells. At 48 h after transfection, the functional assay was started by changing to 500 μL of fresh media and adding 1 μM of unlabeled progesterone with 10,000 cpm of [^{14}C]-Progesterone as a tracer. Media and cells were collected 45 min after incubation at 37 °C. Steroids were extracted from media with ethyl acetate and isooctane (1:1 *vol/vol*) dried and dissolved in methylene chloride. Steroids were separated by thin-layer chromatography (TLC), exposed to a phosphor screen, and visualized with Typhoon PhosphorImager FLA-7000 (GE Healthcare Bio-Sciences AB, Uppsala, Sweden). Image intensity was measured and quantified with ImageQuant TL v8 (Cytiva, MA, USA).

The CYP21A2 enzyme activity was expressed as relative steroid conversion. Steroids were quantified as a percentage of radioactivity incorporated into 11-deoxyprogesterone to the total radioactivity measured in the whole sample and compared between the WT and variants. Cells were collected with trypsin and washed with 1x PBS to quantify the amount of protein. Results were analyzed from three technical replicates. To ensure a similar amount of CYP21A2 in each reaction, a western blot analysis was performed to normalize the enzyme activity with the relative CYP21A2 expression.

4.5. Western Blot

The amount of CYP21A2 expressed was measured from the total protein extraction. Cells were incubated for 1 h with the lysis buffer previously described [46] and centrifuged at $15,000 \times g$ for 20 min and 4 °C. The supernatant was collected for the measurement of total protein through a Pierce Coomassie Plus (Bradford) Assay Kit (Thermo Fisher, Hanover Park, IL, USA). Seven μg of total protein were loaded on an SDS-PAGE gel (GenScript, Piscataway, NJ, USA) and then transferred to a PVDF membrane, as previously described [46]. Two primary antibodies were used at the same time: a mouse monoclonal DKY-Tag antibody diluted 1:1000 (GenScript, Cat# A00187) and a mouse monoclonal anti- β -Actin antibody diluted 1:1500 (Sigma Aldrich, St. Louis, MO, USA, Cat# SAB3500350). The secondary antibody, IRDye 800CW-conjugated donkey-anti-mouse (LI-COR, Nebraska, USA-Cat# 926-32212) was diluted at 1:15,000. An Odyssey SA Infrared Imaging system (LI-COR Bioscience Inc. Lincoln, NE, USA) was used to detect the fluorescence signal.

4.6. Enzyme Kinetics Assay

To obtain the apparent reaction efficiency from the variants that showed enzyme activity, enzyme kinetic assays were performed. Five unlabeled progesterone concentrations were used (0.1, 0.3, 1, 3, and 5 μM) with 15,000 cpm of [^{14}C]-Progesterone as a tracer, under the same conditions that were used in the functional assay. Enzyme velocity was normalized with the relative CYP21A2 expression derived from western blots. Results were analyzed from three biological replicates with Michaelis-Menten kinetics using GraphPad Prism 9.2.0 (GraphPad, San Diego, CA, USA).

4.7. Statistics Analysis

Statistical significance was calculated with a one-sample student's *t*-test, for comparing samples in the same group, and a one-way ANOVA, for comparing two groups (one variant group with the WT group). The P-value was considered significant with $p < 0.05$. Prism 8 (GraphPad, CA, USA) and Excel (Microsoft, Redmond, WA, USA) were used to perform the calculation and statistical analysis.

5. Conclusions

Here we elucidated the structural and functional impact of six CYP21A2 variants (p.P35L, p.L199P, p.W202R, p.E352V, p.P433L, and p.R484L) with unknown profiles via in silico and in vitro enzyme analysis. We showed a good correlation between in silico and in vitro functional studies for structural and conserved variants. However, residues with specific roles have ambiguous results, e.g., for the catalytic site, as the homology has a high weight for most of the predictor tools. Especially in these cases, the functional assay to access the enzyme activity was decisive to assign the damage caused by that variant. These results also emphasize the relevance of using multiple algorithms together with functional assays, especially when it is not possible to establish correlations with the phenotype based solely on bioinformatics predictions.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1.

Author Contributions: Conceptualization, M.J.P. A.Z., M.L.R.R., and A.V.P.; methodology, M.J.P., A.V.P., R.L.-B., S.S., B.V.M., and T.R.; formal analysis, M.J.P. and A.V.P.; investigation, M.J.P.; resources, A.V.P. and A.Z.; writing—original draft preparation, M.J.P.; writing—review and editing, A.Z. C.K. and A.V.P.; visualization, M.J.P. and A.V.P.; supervision, A.V.P., A.Z., M.L.R.R., and R.L.-B.; project administration, A.V.P.; funding acquisition, M.J.P., A.V.P., A.Z., and K.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a Swiss Government Excellence Scholarship (ESKAS), grant numbers 2020.0209 and 2020.0176 and *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq) Scholarship.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data are provided in text or supplementary materials. Publicly available databases used are referred in text.

Acknowledgments: The authors are grateful to Christa E. Flück, Kay Sauter, and Maria Natalia Rojas Velazquez for help with the experimental setup, enzyme assays, and valuable discussions.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Miller, W.L.; Auchus, R.J. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocr. Rev.* **2011**, *32*, 81–151.
2. Pallan, P.S.; Wang, C.; Lei, L.; Yoshimoto, F.K.; Auchus, R.J.; Waterman, M.R.; Guengerich, F.P.; Egli, M. Human Cytochrome P450 21A2, the Major Steroid 21-Hydroxylase: Structure of the enzyme:progesterone substrate complex and rate-limiting c–h bond cleavage. *J. Biol. Chem.* **2015**, *290*, 13128–13143.
3. New, M.I.; Abraham, M.; Gonzalez, B.; Dunic, M.; Razzaghy-Azar, M.; Chitayat, D.; Sun, L.; Zaidi, M.; Wilson, R.C.; Yuen, T. Genotype-phenotype correlation in 1507 families with congenital adrenal hyperplasia owing to 21-hydroxylase deficiency. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 2611–2616.
4. Tardy, V.; Menassa, R.; Sulmont, V.; Lienhardt-Roussie, A.; Lecointre, C.; Brauner, R.; David, M.; Morel, Y. Phenotype-Genotype Correlations of 13 Rare CYP21A2 Mutations Detected in 46 Patients Affected with 21-Hydroxylase Deficiency and in One Carrier. *J. Clin. Endocrinol. Metab.* **2010**, *95*, 1288–1300.
5. Janner, M.; Pandey, A.V.; Mullis, P.E.; Flück, C.E. Clinical and biochemical description of a novel CYP21A2 gene mutation 962_963insA using a new 3D model for the P450c21 protein. *Eur. J. Endocrinol.* **2006**, *155*, 143–151.
6. Pandey, A.V.; Flück, C.E. NADPH P450 oxidoreductase: Structure, function, and pathology of diseases. *Pharmacol. Ther.* **2013**, *138*, 229–254.
7. Ali, S.R.; Bryce, J.; Haghpanahan, H.; Lewsey, J.D.; Tan, L.E.; Atapattu, N.; Birkebaek, N.H.; Blankenstein, O.; Neumann, U.; Balsamo, A.; *et al.* Real-World Estimates of Adrenal Insufficiency-Related Adverse Events in Children With Congenital Adrenal Hyperplasia. *J. Clin. Endocrinol. Metab.* **2021**, *106*, e192–e203.
8. Carmina, E.; Dewailly, D.; Escobar-Morreale, H.F.; Kelestimur, F.; Moran, C.; Oberfield, S.; Witchel, S.F.; Azziz, R. Non-classic congenital adrenal hyperplasia due to 21-hydroxylase deficiency revisited: An update with a special focus on adolescent and adult women. *Hum. Reprod. Update* **2017**, *23*, 580–599.

9. Krone, N.; Riepe, F.G.; Grötzinger, J.; Partsch, C.J.; Sippell, W.G. Functional characterization of two novel point mutations in the CYP21 gene causing simple virilizing forms of congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *J. Clin. Endocrinol. Metab.* **2005**, *90*, 445–454.
10. El-Maouche, D.; Arlt, W.; Merke, D.P. Congenital adrenal hyperplasia. *Lancet* **2017**, *390*, 2194–2210.
11. Held, P.K.; Bird, I.M.; Heather, N.L. Newborn Screening for Congenital Adrenal Hyperplasia: Review of Factors Affecting Screening Accuracy. *Int. J. Neonatal Screen.* **2020**, *6*, 67.
12. Flück, C.E.; Tajima, T.; Pandey, A.V.; Arlt, W.; Okuhara, K.; Verge, C.F.; Jabs, E.W.; Mendonca, B.B.; Fujieda, K.; Miller, W.L. Mutant P450 oxidoreductase causes disordered steroidogenesis with and without Antley-Bixler syndrome. *Nat. Genet.* **2004**, *36*, 228–230.
13. Flück, C.E.; Mallet, D.; Hofer, G.; Samara-Boustani, D.; Leger, J.; Polak, M.; Morel, Y.; Pandey, A.V. Deletion of P399_E401 in NADPH cytochrome P450 oxidoreductase results in partial mixed oxidase deficiency. *Biochem. Biophys. Res. Commun.* **2011**, *412*, 572–577.
14. Parween, S.; Roucher-Boulez, F.; Flück, C.E.; Lienhardt-Roussie, A.; Mallet, D.; Morel, Y.; Pandey, A.V. P450 Oxidoreductase Deficiency: Loss of Activity Caused by Protein Instability From a Novel L374H Mutation. *J. Clin. Endocrinol. Metab.* **2016**, *101*, 4789–4798.
15. Fernandez-Cancio, M.; Camats, N.; Fluck, C.E.; Zalewski, A.; Dick, B.; Frey, B.M.; Monne, R.; Toran, N.; Audi, L.; Pandey, A.V. Mechanism of the Dual Activities of Human CYP17A1 and Binding to Anti-Prostate Cancer Drug Abiraterone Revealed by a Novel V366M Mutation Causing 17,20 Lyase Deficiency. *Pharmaceuticals* **2018**, *11*, 37.
16. de Carvalho, D.F.; Miranda, M.C.; Gomes, L.G.; Madureira, G.; Marcondes, J.A.; Billerbeck, A.E.; Rodrigues, A.S.; Presti, P.F.; Kuperman, H.; Damiani, D.; *et al.* Molecular CYP21A2 diagnosis in 480 Brazilian patients with congenital adrenal hyperplasia before newborn screening introduction. *Eur. J. Endocrinol.* **2016**, *175*, 107–116.
17. Lidaka, L.; Bekere, L.; Lazdane, G.; Dzivite-Krisane, I.; Kivite-Urtane, A.; Gailite, L. Non-Classical Congenital Adrenal Hyperplasia-Causing Alleles in Adolescent Girls with PCOS and in Risk Group for PCOS Development. *Diagnostics* **2021**, *11*, 980.
18. Rodrigues, N.R.; Dunham, I.; Yu, C.Y.; Carroll, M.C.; Porter, R.R.; Campbell, R.D. Molecular characterization of the HLA-linked steroid 21-hydroxylase B gene from an individual with congenital adrenal hyperplasia. *EMBO J.* **1987**, *6*, 1653–1661.
19. Stenson, P.D.; Mort, M.; Ball, E.V.; Evans, K.; Hayden, M.; Heywood, S.; Hussain, M.; Phillips, A.D.; Cooper, D.N. The Human Gene Mutation Database: Towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum. Genet.* **2017**, *136*, 665–677.
20. Simonetti, L.; Bruque, C.D.; Fernández, C.S.; Benavides-Mori, B.; Delea, M.; Kolomenski, J.E.; Espeche, L.D.; Buzzalino, N.D.; Nadra, A.D.; Dain, L. CYP21A2 mutation update: Comprehensive analysis of databases and published genetic variants. *Hum. Mutat.* **2018**, *39*, 5–22.
21. Higashi, Y.; Tanae, A.; Inoue, H.; Hiromasa, T.; Fujii-Kuriyama, Y. Aberrant splicing and missense mutations cause steroid 21-hydroxylase [P-450(C21)] deficiency in humans: Possible gene conversion products. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 7486–7490.
22. Miller, W.L. Gene conversions, deletions, and polymorphisms in congenital adrenal hyperplasia. *Am. J. Hum. Genet.* **1988**, *42*, 4–7.
23. Tusie-Luna, M.T.; Traktman, P.; White, P.C. Determination of functional effects of mutations in the steroid 21-hydroxylase gene (CYP21) using recombinant vaccinia virus. *J. Biol. Chem.* **1990**, *265*, 20916–20922.
24. Barbaro, M.; Soardi, F.C.; Östberg, L.J.; Persson, B.; de Mello, M.P.; Wedell, A.; Lajic, S. In vitro functional studies of rare CYP21A2 mutations and establishment of an activity gradient for nonclassic mutations improve phenotype prediction in congenital adrenal hyperplasia. *Clin. Endocrinol.* **2015**, *82*, 37–44.
25. Haider, S.; Islam, B.; D'Atri, V.; Sgobba, M.; Poojari, C.; Sun, L.; Yuen, T.; Zaidi, M.; New, M.I. Structure-phenotype correlations of human CYP21A2 mutations in congenital adrenal hyperplasia. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 2605–2610.
26. Xu, C.; Jia, W.; Cheng, X.; Ying, H.; Chen, J.; Xu, J.; Guan, Q.; Zhou, X.; Zheng, D.; Li, G.; *et al.* Genotype–phenotype correlation study and mutational and hormonal analysis in a Chinese cohort with 21-hydroxylase deficiency. *Mol. Genet. Genom. Med.* **2019**, *7*, e671.
27. Wang, C.; Pallan, P.S.; Zhang, W.; Lei, L.; Yoshimoto, F.K.; Waterman, M.R.; Egli, M.; Guengerich, F.P. Functional analysis of human cytochrome P450 21A2 variants involved in congenital adrenal hyperplasia. *J. Biol. Chem.* **2017**, *292*, 10767–10778.
28. Coeli, F.B.; Soardi, F.C.; Bernardi, R.D.; De Araújo, M.; Paulino, L.C.; Lau, I.F.; Petroli, R.J.; De Lemos-Marini, S.H.; Baptista, M.T.; Guerra-Júnior, G.; *et al.* Novel deletion alleles carrying CYP21A1P/A2chimeric genes in Brazilian patients with 21-hydroxylase deficiency. *BMC Med. Genet.* **2010**, *11*, 104.
29. Silveira, E.L.; Elnecape, R.H.; dos Santos, E.P.; Moura, V.; Pinto, E.M.; van der Linden Nader, I.; Mendonca, B.B.; Bacheaga, T.A. Molecular analysis of CYP21A2 can optimize the follow-up of positive results in newborn screening for congenital adrenal hyperplasia. *Clin. Genet.* **2009**, *76*, 503–510.
30. Santos-Silva, R.; Cardoso, R.; Lopes, L.; Fonseca, M.; Espada, F.; Sampaio, L.; Brandão, C.; Antunes, A.; Bragança, G.; Coelho, R. CYP21A2 Gene Pathogenic Variants: A Multicenter Study on Genotype–Phenotype Correlation from a Portuguese Pediatric Cohort. *Horm. Res. Paediatr.* **2019**, *91*, 33–45.
31. Carvalho, B.; Pereira, M.; Marques, C.J.; Carvalho, D.; Leão, M.; Oliveira, J.P.; Barros, A.; Carvalho, F. Comprehensive genetic analysis and structural characterization of CYP21A2 mutations in CAH patients. *Exp. Clin. Endocrinol. Diabetes* **2012**, *120*, 535–539.

32. Loke, K.Y.; Lee, Y.S.; Lee, W.W.; Poh, L.K. Molecular analysis of CYP-21 mutations for congenital adrenal hyperplasia in Singapore. *Horm. Res.* **2001**, *55*, 179–184.
33. Menassa, R.T.V.; Despert, F.; Bouvattier-Morel, C.; Brossier, J.P.; Cartigny, M.; Morel, Y. p.H62L, a rare mutation of the CYP21 gene identified in two forms of 21-hydroxylase deficiency. *J. Clin. Endocrinol. Metab.* **2008**, *93*, 1901–1908.
34. Soardi, F.C.; Barbaro, M.; Lau, I.F.; Lemos-Marini, S.H.V.; Baptista, M.T.M.; Guerra-Junior, G.; Wedell, A.; Lajic, S.; De Mello, M.P. Inhibition of CYP21A2 Enzyme Activity Caused by Novel Missense Mutations Identified in Brazilian and Scandinavian Patients. *J. Clin. Endocrinol. Metab.* **2008**, *93*, 2416–2420.
35. Adzhubei, I.A.; Schmidt, S.; Peshkin, L.; Ramensky, V.E.; Gerasimova, A.; Bork, P.; Kondrashov, A.S.; Sunyaev, S.R. A method and server for predicting damaging missense mutations. *Nat. Methods* **2010**, *7*, 248–249.
36. Hecht, M.; Bromberg, Y.; Rost, B. Better prediction of functional effects for sequence variants. *BMC Genom.* **2015**, *16* (Suppl. 8), S1.
37. Capriotti, E.; Altman, R.B.; Bromberg, Y. Collective judgment predicts disease-associated single nucleotide variants. *BMC Genom.* **2013**, *14*, S2.
38. Bendl, J.; Stourac, J.; Salanda, O.; Pavelka, A.; Wieben, E.D.; Zendulka, J.; Brezovsky, J.; Damborsky, J. PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. *PLoS Comput. Biol.* **2014**, *10*, e1003440.
39. Pejaver, V.; Urresti, J.; Lugo-Martinez, J.; Pagel, K.A.; Lin, G.N.; Nam, H.-J.; Mort, M.; Cooper, D.N.; Sebat, J.; Iakoucheva, L.M.; *et al.* Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat. Commun.* **2020**, *11*, 5918.
40. Quan, L.; Lv, Q.; Zhang, Y. STRUM: Structure-based prediction of protein stability changes upon single-point mutation. *Bioinformatics* **2016**, *32*, 2936–2946.
41. Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein structure and function prediction. *Nat. Methods* **2015**, *12*, 7–8.
42. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* **2011**, *27*, 1164–1165.
43. Altschul, S. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402.
44. Altschul, S.F.; Wootton, J.C.; Gertz, E.M.; Agarwala, R.; Morgulis, A.; Schäffer, A.A.; Yu, Y.K. Protein database searches using compositionally adjusted substitution matrices. *FEBS J.* **2005**, *272*, 5101–5109.
45. Schäffer, A.A.; Aravind, L.; Madden, T.L.; Shavirin, S.; Spouge, J.L.; Wolf, Y.I.; Koonin, E.V.; Altschul, S.F. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* **2001**, *29*, 2994–3005.
46. Cohen, M.; Pignatti, E.; Dines, M.; Mory, A.; Ekhilevitch, N.; Kolodny, R.; Flück, C.E.; Tiosano, D. In Silico Structural and Biochemical Functional Analysis of a Novel CYP21A2 Pathogenic Variant. *Int. J. Mol. Sci.* **2020**, *21*, 585.

Supplemental Data

Characterization of mutations causing CYP21A2 deficiency in Brazilian and Portuguese populations.

Mayara J. Prado^{1,2,3,4*}, Shripriya Singh^{3,4}, Rodrigo Ligabue-Braun⁵, Bruna V. Meneghetti^{1,2}, Thaiane Rispoli^{1,2}, Cristiane Kopacek⁶, Karina Monteiro^{1,2}, Arnaldo Zaha^{1,2‡}, Maria L. R. Rossetti^{1,7‡}, Amit V. Pandey^{3,4*‡}

¹ Graduate Program in Cell and Molecular Biology, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre CEP 91501-970, Brazil.

² Center for Biotechnology, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre CEP 91501-970, RS, Brazil.

³ Department of Biomedical Research, University of Bern, Bern 3010, Switzerland.

⁴ Pediatric Endocrinology Unit, Department of Pediatrics, University Children's Hospital Bern, Bern 3010, Switzerland

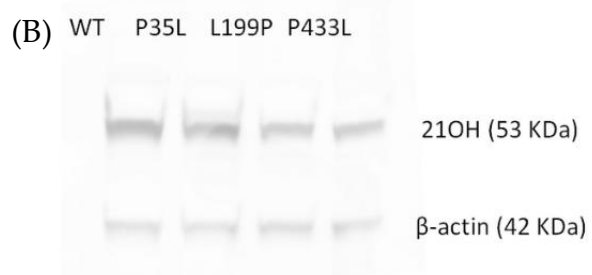
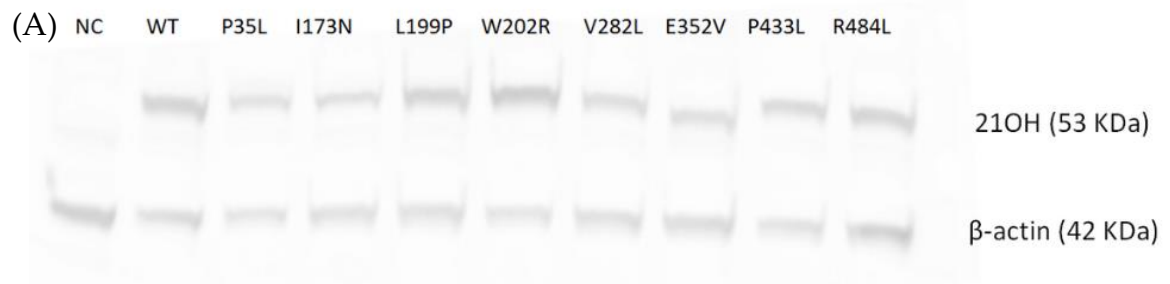
⁵ Departament of Pharmacosciences, Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), Porto Alegre CEP 90050-170, Brazil.

⁶ Serviço de Referência em Triagem Neonatal, Hospital Materno Infantil Presidente Vargas, Porto Alegre CEP 90035-074, Brazil.

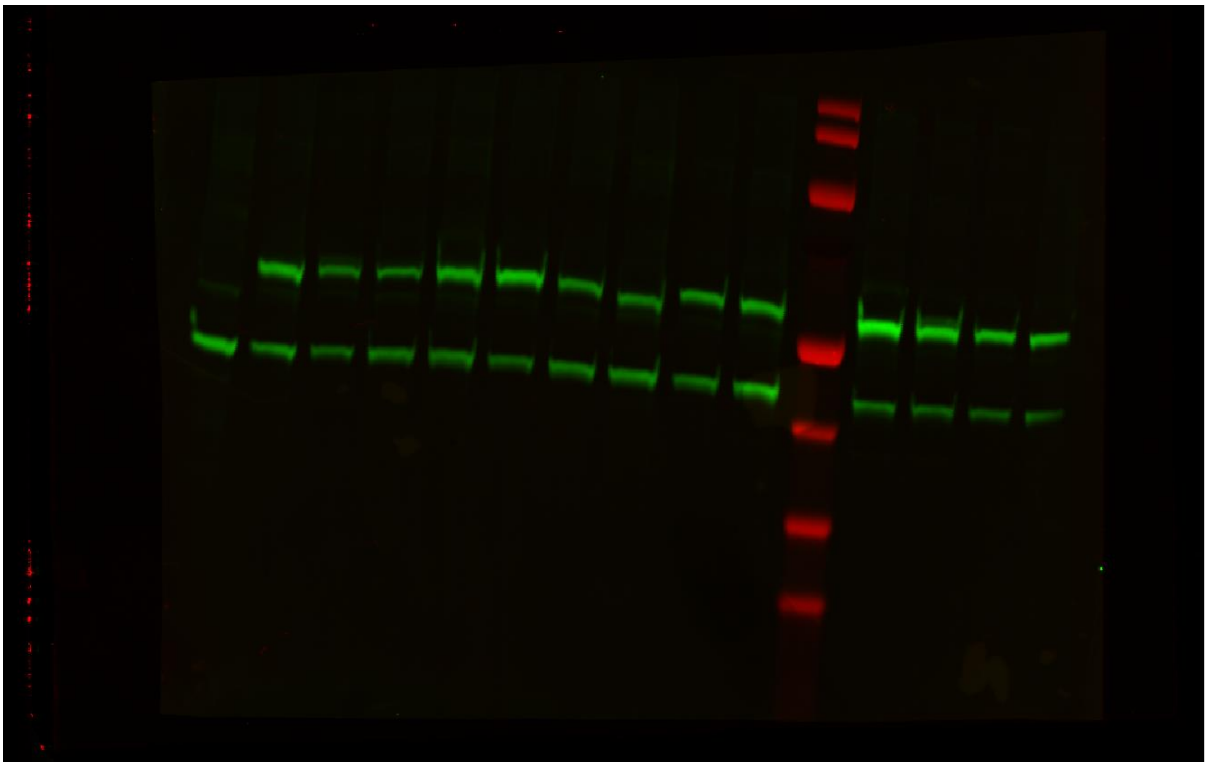
⁷ Graduate Program in Molecular Biology Applied to Health, Universidade Luterana do Brasil (ULBRA), Canoas CEP 92425-020, Brazil.

*Correspondence: amit@pandeylab.org (AVP); mayjorgens@hotmail.com (MJP)

‡These authors contributed equally to this work.



(C)



(D)

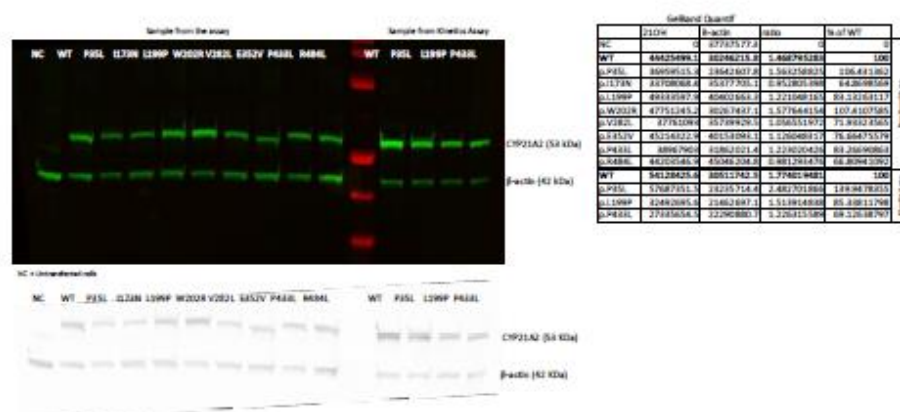


Figure S1. Specific 21-hydroxylase (21OH) protein expression by western blot. Cells from the functional and kinetic assay were collected and lysed. The total protein was measured through Pierce Coomassie Plus (Bradford) Assay Kit. Seven μg of total protein were loaded on an SDS-PAGE gel and then transferred to a PVDF membrane. Two primary antibodies were used, a mouse monoclonal anti-flag for the 21OH (53 KDa) and a mouse monoclonal anti- β -actin for the normalizing gene β -actin (42 KDa). One secondary antibody was used, IRDye 800CW-conjugated donkey-anti-mouse. The fluorescence signal was detected with Odyssey SA Infrared Imaging system. (A) 21OH protein from the functional assay with all variants. Western blot was performed for one sample of the technical triplicate. (B) 21OH protein from the kinetic assay. Western blot was performed for one sample of the biological triplicate. NC, negative control (basal HEK293 cells); WT, 21OH wild type. Original western blot used for analysis of proteins. D. Image analysis of western blot.

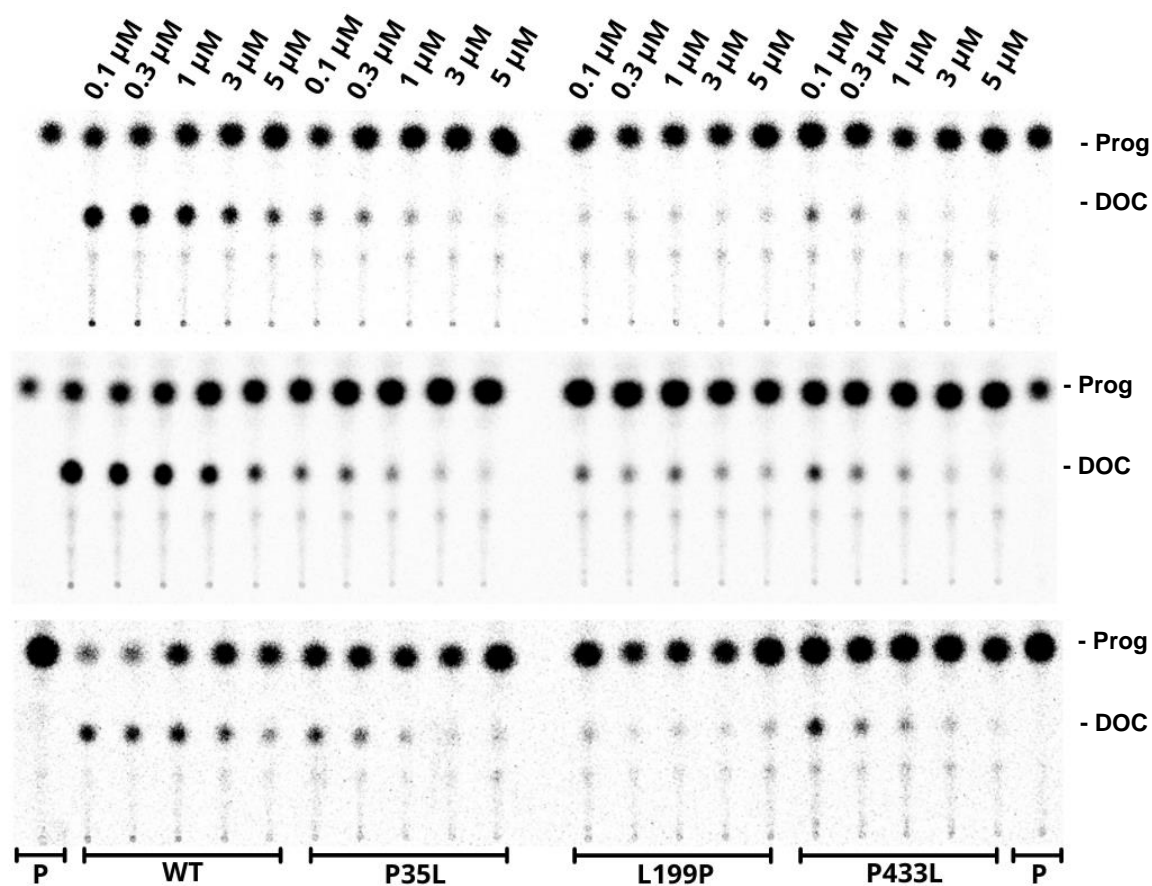


Figure S2. Thin layer chromatography (TLC) from the kinetics assays with P35L, L199P and P433L. The 21-hydroxylase activity was measured by the conversion of progesterone (Prog; P) (0.1, 0.3, 1, 3 and 5 μM) with 15,000 cpm of [^{14}C]-Progesterone into 11-deoxycorticosterone (DOC) radioactive. Three biological replicates were performed. WT, wild type.

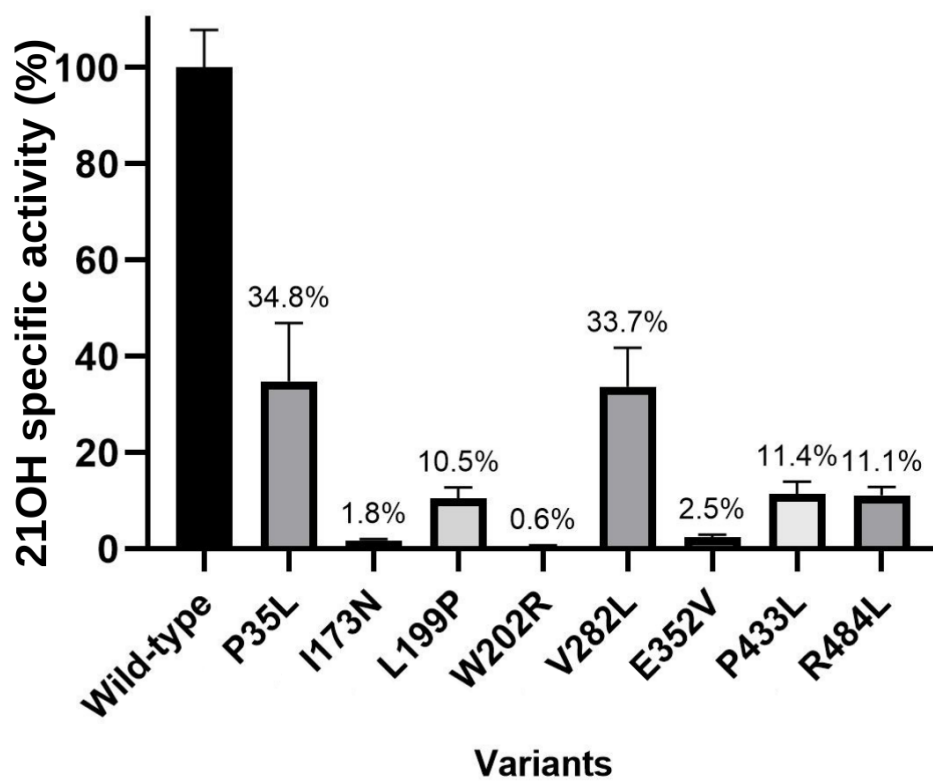


Figure S3. Plot of the 21-hydroxylase specific activity. The specific activities are expressed by the enzyme activity obtained from the TLC spot densitometry by the specific protein estimated through western blot. Results are from a technical triplicate, and all calculations were done on GraphPad software.

Table S1. Hydrogen-bonds of each wide-type residues studied on the CYP21A2 structure. The CYP21A2 structure was built on I-TASSER based on the x-ray crystallography structure deposited on PDB under the ID 4Y8W. Hydrogen bonds were calculated with the Optimal Hydrogen Bounding Network tool on WHAT-IF web (<https://swift.cmbi.umcn.nl/servers/html/index.html>).

Amino acid 1	Atom 1	Amino acid 2	Atom 2	Distance 1-2 (Å)
LEU 199	N	ILE 195	O	3.11
LEU 199	O	SER 203	N	3.00
TRP 202	N	VAL 198	O	3.00
GLU 352	N	ALA 348	O	2.99
GLU 352	OE2	ARG 355	NE	2.80
GLU 352	OE1	ARG 355	NH1	2.78
GLU 352	O	LEU 356	N	2.97
GLU 352	OE1	TRP 406	N	2.86
PRO 433	O	ALA 435	N	2.64
ARG 484	N	GLN 482	O	3.07
ARG 484	N	ALA 449	O	2.98
ARG 484	O	MET 486	N	2.96

Table S2. Hydrogen-bonds of the variants studied on the CYP21A2. The complete CYP21A2 protein structure based on the PDB ID 4Y8W was mutated on STRUM (one mutation per structure). Hydrogen bonds were calculated with the Optimal Hydrogen Bounding Network tool on WHAT-IF web (<https://swift.cmbi.umcn.nl/servers/html/index.html>).

Amino acid 1	Atom 1	Amino acid 2	Atom 2	Distance 1-2 (Å)
PRO 199	O	SER 203	N	2.79
ARG 202	N	VAL 198	O	3.02
VAL 352	N	ALA 348	O	2.97
VAL 352	O	LEU 356	N	2.91
LEU 433	O	ALA 435	N	3.16
LEU 484	N	ALA 449	O	2.85

Table S3. Table of missense variants known in NCBI database at the amino acid positions in CYP21A2 described in our study. Gene: CYP21A2. mRNA: [NM_000500.7](#) Protein: [NP_000491.4](#)

Chr. position	mRNA pos	dbSNP rs# cluster id	MAF	Function	dbSNP allele	Protein residue	Amino acid pos
32038526	211	rs200648381		missense	T	Leu [L]	35
				contig reference	C	Pro [P]	35
32039592	703	rs1396254543		missense	C	Ser [S]	199
				contig reference	T	Leu [L]	199
32039593	704	rs143240527	0.0036	missense	T	Phe [F]	199
32040071	912	rs752306014		missense	G	Gly [G]	269
				contig reference	A	Ser [S]	269
32040520	1161	rs771822460		missense	A	Lys [K]	352
				contig reference	G	Glu [E]	352
32040925	1386	rs1370167869		missense	T	Cys [C]	427
				contig reference	C	Arg [R]	427
32040944	1405	rs751456004		missense	T	Leu [L]	433
				contig reference	C	Pro [P]	433
32041096	1557	rs759736443		missense	T	Trp [W]	484
32041097	1558	rs200005406		missense	A	Gln [Q]	484
				missense	C	Pro [P]	484
				contig reference	G	Arg [R]	484

Capítulo IV – Artigo 2

***Meta-analysis of variant predictions in congenital adrenal hyperplasia
caused by mutations in CYP21A2***

Submetido à revista:

The Journal of Clinical Endocrinology & Metabolism

Contribuição dos autores:

Conceptualização, M.J.P, R.L-B, A.Z. e A.V.P; metodologia, M.J.P., A.V.P e R.L-B; análise formal, M.J.P; investigação, M.J.P .; recursos, A.V.P. e A.Z.; redação - preparação do rascunho original, M.J.P .; redação - revisão e edição, R.L-B., A.Z. e A.V.P.; visualização, M.J.P; supervisão, A.V.P., A.Z., M.L.R.R. e R.L-B.; aquisição de financiamento, M.J.P., A.V.P. e A.Z..

1 **Meta-analysis**

2 **Meta-analysis of variant predictions in congenital**
3 **adrenal hyperplasia caused by mutations in CYP21A2**

4 **Mayara J. Prado^{1,2,3,4}, Rodrigo Ligabue-Braun⁵, Arnaldo Zaha^{1,2}‡, Maria Lucia Rosa**

5 **Rossetti^{1,4}‡, Amit V. Pandey^{3,4} ‡**

6 ‡These authors contributed equally to this work.

7 ¹Graduate Program in Cell and Molecular Biology, Universidade Federal do Rio Grande do Sul (UFRGS),
8 Porto Alegre CEP 91501-970, Brazil. ² Center for Biotechnology, Universidade Federal do Rio Grande do
9 Sul (UFRGS), Porto Alegre CEP 91501-970, RS, Brazil. ³ Department of Biomedical Research, University of
10 Bern, Bern 3010, Switzerland. ⁴ Pediatric Endocrinology Unit, Department of Pediatrics, University
11 Children's Hospital Bern, Bern 3010, Switzerland. ⁵ Department of Pharmacosciences, Universidade
12 Federal de Ciências da Saúde de Porto Alegre (UFCSPA), Porto Alegre CEP 90050-170, Brazil. ⁶ Graduate
13 Program in Molecular Biology Applied to Health, Universidade Luterana do Brasil (ULBRA), Canoas CEP
14 92425-020, Brazil.

15
16 **ORCID numbers:** M.J.P (0000-0003-0647-4429), R.L-B. (0000-0002-2555-9754), A.Z.
17 (0000-0001-6336-474X), M.L.R.R (0000-0002-9672-9394), and A.V.P (0000-0001-8331-
18 5902).

19 **Supplementary materials:**

20 [https://www.biorxiv.org/content/10.1101/2021.12.21.473700v1.supplementary-](https://www.biorxiv.org/content/10.1101/2021.12.21.473700v1.supplementary-material)
21 [material](https://www.biorxiv.org/content/10.1101/2021.12.21.473700v1)

22 <https://www.biorxiv.org/content/10.1101/2021.12.21.473700v1>

23 **Short title:** Meta-analysis of variants predictors for CYP21A2.

24 **Corresponding author's contact information:**

25 Name: Mayara Jorgens Prado

26 Address: Federal University of Rio Grande do Sul (UFRGS), Biotechnology center 9500,

27 Bento Gonçalves avenue. Postcode 91501970 - Porto Alegre/Brazil

28 E-mail: mayjorgens@hotmail.com

29

30 **Name and address of the person to whom reprints should be addressed:** Same as the
31 corresponding author.

32

33 **Grants:** This research was funded by a Swiss Government Excellence Scholarship
34 (ESKAS), grant number 2020.0209, and a *Conselho Nacional de Desenvolvimento*
35 *Científico e Tecnológico* Scholarship (**CNPq**).

36

37 **Disclosure:** The authors declare no conflict of interest. The funders had no role in the
38 design of the study; in the collection, analyses, or interpretation of data; in the writing
39 of the manuscript, or in the decision to publish the results.

40 **Abstract:**

41 **Context:** CYP21A2 deficiency represents 95% of congenital adrenal hyperplasia cases
42 (CAH), a group of genetic disorders that affect steroid biosynthesis. The genetic and
43 functional analysis provides critical tools to elucidate complex CAH cases. One of the
44 most accessible tools to infer the pathogenicity of new variants is *in silico* predictions.

45 **Objective:** Analyze the performance of *in silico* prediction tools to categorize missense
46 single nucleotide variants (SNVs) of the CYP21A2.

47 **Methods:** SNVs of the CYP21A2 characterized *in vitro* by functional assays were selected
48 to assess the performance of online single and meta predictors. SNVs were tested
49 separately or in combination with the related phenotype (severe or mild CAH form). In
50 total, 103 SNVs of the CYP21A2 (90 pathogenic and 13 neutral) were used to test the
51 performance of 13 single-predictors and four meta-predictors.

52 **Results:** SNVs associated with the severe phenotypes were well categorized by all tools,
53 with an accuracy between 0.69 (PredictSNP2) and 0.97 (CADD), and Matthews'
54 correlation coefficient (MCC) between 0.49 (PredicSNP2) and 0.90 (CADD). However,
55 SNVs related to the mild phenotype had more variation, with the accuracy between 0.47
56 (S3Ds&GO and MAPP) and 0.88 (CADD), and MCC between 0.18 (MAPP) and 0.71
57 (CADD).

58 **Conclusion:** From our analysis, we identified four predictors of CYP21A2 pathogenicity
59 with good performance. These results can be used for future analysis to infer the impact
60 of uncharacterized SNVs' in CYP21A2.

61 **Keywords:** online prediction; CYP21A2; mutation analysis; pathogenicity prediction.

62 Introduction

63 One of the most common autosomal recessive genetic disorders is the impairment of
64 21-hydroxylase (CYP21A2). This enzyme deficiency represents about 95% of the
65 congenital adrenal hyperplasia (CAH) cases, a group of enzymatic disorders that affects
66 cortisol biosynthesis. The CYP21A2 enzyme is a member of the cytochrome P450 family
67 (CYPs) and catalyzes the conversion of 17-hydroxyprogesterone (17OHP) into 11-
68 deoxycortisol and progesterone into 11-deoxycorticosterone. Other enzymes
69 subsequently convert these steroids into cortisol and aldosterone, respectively ¹.

70 Clinically, the CYP21A2 deficiency in humans has a wide spectrum of phenotypes, from
71 severe to mild or asymptomatic ^{2,3}. The classical severe CAH has salt-wasting (SW) and
72 simple-virilizing (SV) forms. The classical SW form has no enzyme activity and is related
73 to severe virilization and electrolyte dysregulation. In contrast, the classical SV form has
74 enough residual enzyme activity to prevent adrenal crisis ². The mild CAH is the non-
75 classical (NC) form of CAH and has a partial CYP21A2 activity associated with
76 hyperandrogenism and mild late-onset CAH ³. Furthermore, there is a relatively good
77 genotype-phenotype correlation for CYP21A2 deficiency, which allows the variants
78 categorization according to the residual enzyme activity (obtained from *in vitro* studies)
79 and their expected phenotype ³. The classical CAH has less than 10% of wild-type (WT)
80 enzyme activity in 95% of the cases, while the NC form has an activity between 10-78%
81 of the WT in 90% of the cases ⁴.

82 The *CYP21A2* gene is a tandemly arranged module (RCCX: *RP-C4-CYP21-TNX*) and shows
83 96-98% of sequence identity with its pseudogene, *CYP21A1P* ⁵. These features make the
84 *CYP21A2* gene complex, with many different types of mutations - from single nucleotide

85 variants (SNVs) to genetic rearrangements - and contribute to the fact that most carriers
86 have compound heterozygous mutations ³. However, only ten mutations described in
87 the general population are sampled by CYP21A2 deficiency screening programs. The
88 whole gene sequence analysis by Sanger sequencing is an alternative method for
89 exceptional cases due to the cost and time-consuming nature of such studies ^{6,7}.

90 So far, with the whole *CYP21A2* gene sequencing, genetic studies have reported more
91 than 1,300 variants for the *CYP21A2* gene. Out of the 230 variants reported as affecting
92 human health, 153 are missense variants ⁴. The advancement of next-generation
93 sequencing (NGS) to analyze a large number of genes has facilitated the detection of
94 rare single nucleotide variants (SNVs) and single nucleotide polymorphisms (SNPs). A
95 few years ago, this technology was not applied to screen the *CYP21A2* gene due to its
96 high sequence identity with *CYP21A2P* which hampers the proper analysis of this
97 genomic region ⁵. However, recently some groups have found alternative ways to
98 perform NGS for the *CYP21A2* gene through a combination with other methods, such as
99 Multiplex ligation-dependent probe amplification ^{8,9}. These genetic analysis strategies
100 of the *CYP21A2* gene with the NGS technology represent a promising tool for the future,
101 opening the windows to identify new variants while improving the diagnosis of CYP21A2
102 deficiency, and establishing a more reliable estimate of mutation frequencies.

103 The gold standard method for the characterization of new variants is *in vitro* functional
104 assay. However, this approach takes too much time, and it is not viable for all-new
105 variants detected. One of the most accessible tools to predict the pathogenicity variants
106 is the *in silico* analysis, which usually has free access, a friendly interface, and provides
107 quick results. Many online predictors available have different features and approaches,
108 from single characteristic analysis to meta-predictors with different compositions and

109 algorithms. Some studies have shown the general performances of these tools against a
110 whole database with few predictors ^{10,11}. However, studies with variants on protein-
111 specific analysis showed that general analysis results cannot be extrapolated for all
112 proteins as each protein has unique characteristics, which is a key limitation of predictor
113 programs ¹²⁻¹⁴. Therefore, it is essential to be careful when choosing the prediction tools
114 and to consider their variable accuracies for each gene ¹¹.

115 The present study investigates the performance of online predictor tools to classify
116 missense SNVs of the CYP21A2. Missense SNV is the most common group of variants in
117 the human genome, one at every kilobase. In the *CYP21A2* gene, this type of SNV
118 represents about 60% of the variants on The Human Gene Mutation Database (HGMD,
119 RRID:SCR_001888)⁶ and 65% of the affecting human health ⁴. Additionally, missense SNV
120 is one of the hardest variant types for interpretation ^{4,15,16}. In total, we tested 17
121 predictors with multiple algorithms approaches and datasets. Thirteen of these were
122 based on single features: CADD (RRID:SCR_018393)¹⁷, ConSurf (RRID:SCR_002320) ¹⁸,
123 DANN ¹⁹, FATHMM ²⁰, MAPP (RRID:SCR_014375) ²¹, MutPred2 (RRID:SCR_010778) ²²,
124 PANTHER-PSEP (RRID: SCR_005145) ²³, PhD-SNP⁸ (RRID: SCR_010782) ²⁴, PolyPhen-2
125 (RRID:SCR_013189) ²⁵, PROVEAN (RRID: SCR_002182) ²⁶, SIFT (RRID:SCR_012813) ²⁷,
126 SNAP2 (RRID:SCR_002127) ²⁸, and SNPs&GO (RRID:SCR_005788)²⁹. Four meta-
127 predictors: PredictSNP ³⁰, PredictSNP2 ³¹, Meta-SNP ³², and S3Ds&GO ³³) (Figure 1). We
128 excluded nonsense and frameshift variants from our analysis since they have specific
129 settings in many predictors and a high agreement ratio between tools.

130

131 **Materials and Methods**

132 *SNVs selection and categorization*

133 To select CYP21A2 missense single nucleotide variants (SNVs) with clinical significance,
134 we used the list of variants reported to affect human health, as reviewed by ⁴.
135 Complementarily, we searched for SNVs reported by dbSNP, Ensembl, and GeneCards
136 applying the following filter when present: “missense”, clinical significance
137 “pathogenetic” or “benign”, “without conflicting interpretation”, and “human or homo
138 sapiens”. We excluded nonsense and frameshift mutations. In addition, we performed
139 a cross-check of the databases with original articles or reviews to remove variants
140 without the enzyme activity data available.

141 To standardize the effect of the SNVs selected, we categorized them into three groups
142 according to the CYP21A2 activity for at least one of the steroids: classical (CL),
143 nonclassical (NC), and neutral. The CL group has SNVs with the CYP21A2 activity level <
144 10 % relative to WT, the NC group has SNVs with the activity level between > 10% and <
145 78 % relative to WT, and the neutral group has SNV with the enzyme activity > 78 %
146 relative to WT. The CAH group is composed of all SNVs from the CL and NC. The enzyme
147 activity's mean and standard deviation (SD) were calculated by steroid and mutation
148 groups.

149 *Predictor tools selection*

150 To choose predictors with different features, we reviewed the literature for software
151 applied to *in silico* analysis of SNP or SNVs. Predictors used in more than two studies by
152 different research groups or significant performance in a large study were selected. In
153 addition, we filtered for tools with free access and online availability, thus not requiring

154 local powerful computational resources. Characteristics of each predictor chosen are
155 shown in Tables S1 and S2.

156 *Data treatment*

157 The default setting for missense mutation was used on all predictors. However, when
158 there was no set instruction for that purpose highlighted on the program, we followed
159 the developers' recommendation from the tutorial or original paper. In addition, three
160 scores were extracted indirectly from meta-predictors: MAPP (v.28.6.2005) and SIFT
161 (v.4.0.4) score were obtained from PredictSNP, and DANN (v.1.2) score from
162 PredictSNP2. For statistical purposes, we standardized two variables for the outputs of
163 all the predictors: "**damage**" for SNVs with the potential to affect CYP21A2 and "**neutral**"
164 for SNVs with no or very low potential to affect the enzyme. The following outputs were
165 standardized as "damage": score > 0.5 to Meta-SNP, SNP&GO, S3D&GO, MutPred2,
166 FATHMM-MKL (weighted) and PhD-SNP⁶; "deleterious" message to PredictSNP,
167 PredictSNP2, MAPP, and SIFT; score > 0.9 to DANN; score > -2.5 to PROVEN; score > 10
168 to CADD (GRCh38-v1.5-6); score > 0.45 PolyPhen-2 (HumVar); score < 0 to ConSurf; score
169 > 0 to SNAP2; and score > 450 millions of years to PANTHER. Otherwise, we classified
170 the outputs as "neutral".

171 *Analytical parameters*

172 We analyzed the performance of each predictor in two ways. Firstly, to assess the
173 performance to discriminate the effect of CYP21A2 SNVs, we compared SNVs of the CAH
174 group with the neutral group. Secondly, to get the number of hit and miss per group, we
175 analyzed CL and NC groups separated against the neural group. We used Microsoft Excel

176 for the data organization and, together with IBM SPSS Statistic software v.2.1, we
 177 performed the statistical analysis.

178 *Statistical methods*

179 We considered true positive (TP) result for correct "damage" prediction, true negative
 180 (TN) for correct "neutral" prediction, false positive (FP) for incorrect "neutral"
 181 prediction, and false-negative (FN) for incorrect "damage" prediction. We calculated the
 182 positive predictive values (PPV) to access the ratio of TP results for all positive results
 183 (Eqn. 1), and negative predictive values (NPV) to the ratio of TN for all negative results
 184 (Eqn. 2).

185

$$186 \quad PPV = \frac{TP}{TP+FP} \quad (1)$$

$$187 \quad NPV = \frac{TN}{TN+FN} \quad (2)$$

188

189 The proportion of correct SNVs identified as harmful was assessed with the sensibility
 190 (Se) equation (Eqn. 3), while the correct neutral identification was assessed with the
 191 specificity (Sp) (Eqn. 4). Besides that, we obtained the accuracy (Ac) by the ratio of true
 192 results (TP and TN) (Eqn. 5). The accuracy was classified as excellent ($0.9 < Ac < 1.0$),
 193 good ($0.8 < Ac < 0.9$), fair ($0.7 < Ac < 0.8$), and not good ($0.6 < Ac < 0.7$).

$$194 \quad Se = \frac{TP}{TP+FN} \quad (3)$$

$$195 \quad Sp = \frac{TN}{TN+FP} \quad (4)$$

$$196 \quad Ac = \frac{TP+TN}{TP+FP+TN+FN} \quad (5)$$

197 Finally, we applied the Matthews' correlation coefficient (MCC) to measure the two-
 198 class quality (harmful and neutral). This method is suitable for imbalanced data and has
 199 been used to evaluate *in silico* prediction approaches. MCC score ranges from 1 (perfect
 200 prediction) to -1 (total disagreement between the result predicted and the observed),
 201 being 0 no better than random prediction (Eqn. 6) (Chicco *et al.*, 2021).

$$202 \text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (6)$$

203

204 **Results**

205 *Data of the selected SNVs*

206 From variants in the *CYP21A2* gene reported on the literature and databanks, we
 207 selected missense SNVs with clinical significance (criteria described in section 4.1). We
 208 obtained 96 valid SNVs out of 299 missense variants in the list [4], 85 out of 614 missense
 209 variants in dbSNP, 66 out of 459 missense variants in Ensembl, 45 out of 71 missense
 210 variants in GeneCards, 47 out of 83 missense variants in ClinVar, 26 in OMIM, and 81 in
 211 UniProt databases.

212 By removing SNVs that were duplicated and with no functional characterization, we
 213 obtained 103 SNVs, 51 classified as classical, 39 as non-classical, and 13 as neutral. The
 214 SNVs selected with the respective enzyme activity are described in Table S3. All studies
 215 presented the *CYP21A2* activity measured by the hydroxylation of 17OHP, while 86 also
 216 measured progesterone hydroxylation. Mutations of the CL group have the mean
 217 enzyme activity for the 17OHP hydroxylation of 1.5 ± 2 (SD)% and the progesterone
 218 hydroxylation of $1.3 \pm 1.6\%$. While mutations of the NC group have 17OHP hydroxylation

219 activity of $42.9 \pm 23\%$ and progesterone hydroxylation activity of $37.4 \pm 21\%$. Finally,
220 mutations in the neutral group have 17OHP hydroxylation activity of $100.17 \pm 11\%$ and
221 progesterone hydroxylation activity of $94.1 \pm 8.25\%$.

222 *General analysis of mutation groups*

223 We obtained 22 SNVs in the *CYP21A2* gene with the correct prediction for all tested
224 predictors, although the exact number was incorrectly predicted for at least half of them
225 (Table S3). There was no SNV wrongly predicted by all predictors. We compared the hit
226 and miss by the 17 predictors for all SNV affecting *CYP21A2* activity, the CAH group, and
227 for all the non-pathogenic SNV of the neutral group. We showed that 22% (22 of 90)
228 SNVs from the CAH group obtained the correct score by all predictors, while 23% (21 of
229 90) were wrongly predicted by at least nine tools. The neutral group got 2 of its 13 SNVs
230 (15.4%) rightly mispredicted by all and one by nine predictors. Moreover, we divided the
231 SNV of the CAH group into the CL and NC groups. We got 37% (out of 51 SNVs) from the
232 CL and 2.6% (of 39 SNVs) from the NC groups SNVs correctly predicted by all tools. While
233 5.9% and 46% from CL and NC groups, respectively, were wrongly predicted by nine
234 tools.

235 *Performance of predictors to identify SNVs detrimental to CYP21A2*

236 We analyzed the performance of 17 predictors to identify the 90 SNVs that affect
237 *CYP21A2* functionality against the 13 SNVs with neutral effect (Table 1). All the 17
238 predictors obtained a good PPV rate (> 0.90). However, only CADD (0.73) and DANN
239 (0.56) showed negative predictive values (NPVs) higher than 0.5. PANTHER-PSEP has no
240 result for the NPV, as it could not identify benign variants.

241 We obtained both sensibility and specificity higher than 0.8 for three predictors, CADD
242 (se=0.96 and sp=0.85), ConSurf (se=0.88 and sp=0.90), and PolyPhen-2 (se=0.87 and
243 sp=0.85). Moreover, five predictors obtained the accuracy between excellent and good,
244 CADD (0.94), PANTHER-PSEP (0.91), DANN (0.89), ConSurf (0.88), and PolyPhen-2 (0.86)
245 (Table).

246 The MCC test showed positive values for that of almost all predictors (except by
247 PANTHER-PSEP), being five of them with the MCC > 0.5 (Table). The best performance,
248 value closer to +1, was obtained by CADD (0.75), followed by ConSurf (0.58), PolyPhen-
249 2 (0.57), DANN (0.56), and PROVEN (0.51). Figure 3 shows a Venn diagram of the four
250 predictors with better accuracy and MCC values.

251 *Performance of predictors to identify SNVs affecting the specific CAH groups*

252 We analyzed the performance of the selected predictors to identify 51 SNVs of the CL
253 group and 39 SNVs of the NC group against 13 SNVs with a neutral effect (Table).
254 Seventeen predictors obtained an excellent PPV rate (> 0.90) for the SNV CL group, 15
255 for the SNVs NC group. Four tools obtained excellent-good (> 0.8) NPV values for the CL
256 group: CADD (1.0), DANN (0.9), PolyPhen-2 (0.85), and ConSurf (0.82). However, for the
257 NC group, only CADD (0.73), DANN (0.6), and PolyPhen-2 (0.52) showed NPV > 0.5.

258 Taking the sensibility and specificity balance, we obtained 12 tools with excellent-good
259 values (>0.8) for the CL group (Table). However, for the NC group, only CADD (se=0.9
260 and sp=0.85) obtained both sensibility and specificity with excellent-good values (Table).
261 The accuracy was excellent for 7 predictors on CL group, CADD (0.97), ConSurf (0.95),
262 PolyPhen-2 and PROVEN (0.94), DANN and MutPred2 (0.92), and Meta-SNP (0.91); and

263 good for four tools on NC group, CADD (0.88), DANN and PANTHER-PSEP (0.81), and
264 ConSurf (0.8).

265 Finally, the MCC test with positive values was obtained by almost all predictors (except
266 for PANTHER-PSEP). The MCC was higher than 0.5 for 16 predictors in the CL group and
267 five in the NC group (Table). For the CL and NC group, the same predictor, CADD, got the
268 MCC value closer to +1, being MCC=0.9 for the CL group and MCC=0.71 for the NC group.

269 **Discussion**

270 Genetics analysis is an essential approach for elucidating complex CYP21A2 deficiency
271 cases, mainly to confirm asymptomatic carriers and unfollow false-positive cases^{7,16}.
272 Therefore, fast and accessible tools to infer variants' pathogenicity are essential to
273 quickly deduce the harm of unknown variants. *In silico* prediction is one of the most
274 accessible tools to infer the pathogenicity of SNVs.

275 Here, for the first time, we analyzed the performance of *in silico* prediction tools to
276 discriminate pathogenic and neutral variants of the CYP21A2. We focus on the
277 performance of 13 single predictors and four meta predictors chosen accordingly with
278 the popularity and performance of free access programs. All these programs were able
279 to identify pathogenic variants. Nonetheless, only PANTHER-PSEP could not distinguish
280 neutral variants, which is unacceptable for testing variants of the CYP21A. Moreover, all
281 tools showed better performance with variants of the CL group than the NC group, as
282 expected, since the CL group gathers the most harmful variants.

283 Our databank for performance tests comprises all missense variants of the CYP21A2 that
284 are functionally characterized. With this strategy, we could get a more realistic result on
285 the prediction evaluation. However, the number of variants was imbalanced between

286 the two categories, 90 pathogenic and 13 neutral. Therefore, the primary statistics data
287 considered for the performance evaluation were the accuracy and MCC, which consider
288 all values (TP, TN, FP, and FN)^{34,35}. As sensibility and specificity are calculated with half
289 of the information, they cannot represent all the performance by themselves, so we
290 considered the sensibility-specificity balance. Additionally, we also calculated PPV and
291 NPV but, as both are more sensitive to data balance, they were not considered for the
292 program performance³⁴.

293 The main feature assessed by most single predictors tested is the evolutionary data since
294 residue conservation over time can indicate critical residues for the protein function.
295 Four of the tools tested use only this feature for the prediction calculation, FATHMM²⁰,
296 PhD-SNPg²⁴, PROVEAN²⁶, and SIFT²⁷. A similar performance was obtained between
297 these four tools, with a fair accuracy ranging from 0.73 (FATHMM) to 0.79 (SIFT), and
298 MCC from 0.42 (PhD-SNPg) to 0.51 (PROVEAN). Additionally, SIFT had the most excellent
299 sensibility-specificity balance between them, similar to the performance shown by²⁷. In
300¹⁴, with *HSD17B3*, *NR5A1*, *AR*, and *LHCGR* genes, SIFT and PROVEAN also had the same
301 performance, with an accuracy of 0.74-0.75, and MCC of 0.5. In¹², SIFT and PROVEN
302 showed the best results between nine programs tested for *GJB2*, *GJB6*, and *GJB3* genes,
303 with an accuracy of 0.89, while FATHMM produced a large number of erroneous
304 predictions with an accuracy of 0.33. FATHMM also had poor performance in another
305 study, with an accuracy of 0.56 and MCC of 0.04¹⁴.

306 Changes in the secondary and tertiary structure by missense mutations are likely to
307 affect the protein activity¹⁸. Therefore, it is no surprise that the second most evaluated
308 feature is the structural information, being present on ConSurf¹⁸, MutPred2²²,
309 PolyPhen2¹⁰, and SNAP2²⁸. In addition, ConSurf includes phylogenetics relationships,

310 MutPred2 functional proprieties, and SNAP2 uses a matrix of effect probabilities with a
311 neural network method ^{18,22,28}. Meanwhile, PolyPhen2 has two trained datasets as
312 options, HumVar and HumDir. The first trained dataset is suggested for diagnostics of
313 Mendelian diseases, which requires variants with a drastic difference effect ²⁵.
314 Moreover, PolyPhen2 has a low dependency on the sequence alignment employed ¹⁰.
315 PolyPhen2 showed good prediction performance (ac=0.88, MCC=0.64), with the same
316 sensibility but better sensibility-specificity balance than reported in ¹⁰. ConSurf obtained
317 similar performance with the setting used in or test (ac=0.88, MCC=0.58). Since the
318 ConSurf setting is chosen for the alignment sequences, databank, and algorithms, the
319 performance is not comparable with other studies, as it was reported by ConSurf's
320 developers ¹⁸. On the other hand, MutPred2 and SNAP2 both had sensibility-specificity
321 imbalanced of 1.4 and 1.5-folds, respectively, and almost the same fair performance.
322 However, these values were relatively better than reported for MutPred2 by ²² with
323 ClinVar and UniProt database and SNAP2 by ²⁸ with a databank with more than 9,500
324 variants from human genes.

325 Talking about meta predictors, which work with many databank sources and combine
326 outputs from other predictors to generate their own, we expected to obtain one of the
327 best performances. However, counterintuitively, our study showed an intermediate
328 performance compared with the single predictors tested, and the number of tools
329 combined was not related to the prediction improvement. Meta-SNP and PredictSNP
330 performance were better than PredictSNP2 and S3Ds&GO. Furthermore, comparing
331 with the developer tests, we obtained for Meta-SNP ³² and PredictSNP ³⁰ a similar
332 accuracy, while the performances for PredictSNP2 ³¹ and S3Ds&GO ³³ were lower. Meta-
333 SNP and PredictSNP share three single predictors, PhD-SNP, SNAP, and SIFT.

334 Therefore, for the CYP21A2 variants tested, the best performance to categorize
335 missense variants pathogenicity was CADD, with an overall accuracy of 0.94 (CL 0.97; NC
336 0.88) and MCC of 0.75 (CL 0.9, NC 0.71). The specificity (0.85) and sensibility (0.9) also
337 got a good balance. Interestingly, the accuracy and specificity obtained for CYP21A2
338 were even higher than reported by the software developers ¹⁷ with the ClinVar
339 database, which was 0.85 and 0.57, respectively. ConSurf, DANN, and PolyPhen-2
340 showed a similar performance, holding the second-best results accordingly with the
341 accuracy (CAH group 0.86-0.89; CL 0.92-0.95; NC 0.77-0.81,) and MCC (CAH group 0.56-
342 0.58; CL 0.75-0.83; NC 0.51-0.56) values. The sensibility and specificity for ConSurf and
343 PolyPhen2 were well balanced, while DANN had 1.3-folds less specificity than
344 sensibility. The original article of DANN ¹⁹ presents only the area under the curve (AUC)
345 ROC, which was 0.95 using the ClinVar database for the performance test. PolyPhen2
346 showed better sensibility-specificity balance for the CYP21A2 variants than the values
347 presented by ¹⁰, testing the tool with gene-specific mutations (*BRCA1*, *MSH2*, *MLH1*, and
348 *TP52*). We obtained a sensibility similar to ¹⁰, but the specificity was lower, 0.85 and
349 0.60, respectively.

350 The individual error of the top four predictors was six on CADD, 12 on ConSurf, 11 on
351 DANN, and 14 on PolyPhen2. However, computing their prediction together, we would
352 have four false results from 103 missense SNV in the CPY21A2, one neutral (p.L13M),
353 and three pathogenic from the NC phenotype group (p.P106L, p.R225W, and p.M474I).
354 The variant p.P106L was correctly categorized by SNAP2 and PANTHER-PSEP. In turn,
355 PROVEAN²⁶ could type the other three variants correctly, even with a lower sensibility
356 value than the other four tools, mainly for the pathogenic SNVs of the NC group (0.49).
357 Nonetheless, we obtained an intermediated performance with PROVEAN (ac=0.77;

358 MCC=0.51), which could be because it uses the neighbourhood sequences as input,
359 which can be a trick for enzymes since they have some residues with high conservation
360 making critical connections between variable residues. Comparing with the developer'
361 test ²⁶ with the UniProt database (se=0.78; sp=0.79), we had imbalanced sensibility-
362 specificity imbalanced, with similar overall sensibility (0.73) and higher specificity (1.0).

363 In conclusion, we could identify CADD, ConSurf, DANN, and PolyPhen2 as good programs
364 for missense variants prediction of the CYP21A2. Moreover, CADD had the best
365 performance also for identifying mild mutations from the NC group, followed by ConSurf
366 and DANN. These results may be applicable in the future analysis of new or
367 uncharacterized missense variants of the CYP21A2.

368

369 **Author Contributions:**

370 Conceptualization, M.J.P, R.L-B, A.Z., and A.V.P; methodology, M.J.P., A.V.P, and R.L-B;
371 formal analysis, M.J.P; investigation, M.J.P.; resources, A.V.P. and A.Z.; writing—
372 original draft preparation, M.J.P.; writing—review and editing, R.L-B., A.Z, L.M.R., and
373 A.V.P.; visualization, M.J.P; supervision, A.V.P., A.Z., M.L.R.R., and R.L-B.; funding
374 acquisition, M.J.P., A.V.P., and A.Z. All authors have read and agreed to the published
375 version of the manuscript.

376

377

378 **Supplementary Materials:**

379 Supplementary materials can be accessed at
380 <https://www.biorxiv.org/content/10.1101/2021.12.21.473700v1.supplementary->
381 material
382 <https://www.biorxiv.org/content/10.1101/2021.12.21.473700v1>

383 **References**

- 384 1. Miller WL, Auchus RJ. The molecular biology, biochemistry, and physiology of
385 human steroidogenesis and its disorders. *Endocrine Reviews*. 2011;32(1):81-151.
386 doi:10.1210/er.2010-0013
- 387 2. Witchel SF. Congenital Adrenal Hyperplasia. *Journal of Pediatric and Adolescent*
388 *Gynecology*. 2017;30(5):520-534. doi:10.1016/j.jpag.2017.04.001
- 389 3. New MI, Abraham M, Gonzalez B, *et al.* Genotype-phenotype correlation in 1,507
390 families with congenital adrenal hyperplasia owing to 21-hydroxylase deficiency.
391 *Proceedings of the National Academy of Sciences of the United States of America*.
392 2013;110(7):2611-2616. doi:10.1073/pnas.1300057110
- 393 4. Simonetti L, Bruque CD, Fernández CS, *et al.* CYP21A2 mutation update:
394 Comprehensive analysis of databases and published genetic variants. *Human*
395 *Mutation*. 2018;39(1):5-22. doi:10.1002/humu.23351
- 396 5. Rodrigues NR, Dunham I, Yu Y, Carroll MC, Porter RR, Campbell RD. *Molecular*
397 *Characterization of the HLA-Linked Steroid 21-Hydroxylase B Gene from an*
398 *Individual with Congenital Adrenal Hyperplasia*. Vol 6.; 1987.
- 399 6. Stenson PD, Mort M, Ball E v., *et al.* The Human Gene Mutation Database: towards
400 a comprehensive repository of inherited mutation data for medical research,
401 genetic diagnosis and next-generation sequencing studies. *Human Genetics*.
402 2017;136(6):665-677. doi:10.1007/s00439-017-1779-6
- 403 7. Baumgartner-Parzer S, Witsch-Baumgartner M, Hoepfner W. EMQN best practice
404 guidelines for molecular genetic testing and reporting of 21-hydroxylase
405 deficiency. *European Journal of Human Genetics*. 2020;28(10):1341-1367.
406 doi:10.1038/s41431-020-0653-5
- 407 8. Lee C yu, Yen HY, Zhong AW, Gao H. Resolving misalignment interference for NGS-
408 based clinical diagnostics. *Human Genetics*. 2021;140(3):477-492.
409 doi:10.1007/s00439-020-02216-5
- 410 9. Gangodkar P, Khadilkar V, Raghupathy P, *et al.* Clinical application of a novel next
411 generation sequencing assay for CYP21A2 gene in 310 cases of 21- hydroxylase

- 412 congenital adrenal hyperplasia from India. *Endocrine*. 2021;71(1):189-198.
413 doi:10.1007/s12020-020-02494-z
- 414 10. Hicks S, Wheeler DA, Plon SE, Kimmel M. Prediction of missense mutation
415 functionality depends on both the algorithm and sequence alignment employed.
416 *Human Mutation*. 2011;32(6):661-668. doi:10.1002/humu.21490
- 417 11. Tang B, Li B, Gao LD, *et al.* Optimization of in silico tools for predicting genetic
418 variants: individualizing for genes with molecular sub-regional stratification.
419 *Briefings in Bioinformatics*. 2020;21(5):1776-1786. doi:10.1093/bib/bbz115
- 420 12. Pshennikova VG, Barashkov NA, Romanov GP, *et al.* Comparison of Predictive in
421 Silico Tools on Missense Variants in GJB2, GJB6, and GJB3 Genes Associated with
422 Autosomal Recessive Deafness 1A (DFNB1A). *Scientific World Journal*. 2019;2019.
423 doi:10.1155/2019/5198931
- 424 13. Hart SN, Polley EC, Shimelis H, Yadav S, Couch FJ. Prediction of the functional
425 impact of missense variants in BRCA1 and BRCA2 with BRCA-ML. *npj Breast
426 Cancer*. 2020;6(1):13. doi:10.1038/s41523-020-0159-x
- 427 14. Montenegro LR, Lerário AM, Nishi MY, Jorge AAL, Mendonca BB. Performance of
428 mutation pathogenicity prediction tools on missense variants associated with
429 46,xy differences of sex development. *Clinics*. 2021;76:1-5.
430 doi:10.6061/clinics/2021/e2052
- 431 15. Khan S, Vihinen M. Spectrum of disease-causing mutations in protein secondary
432 structures. *BMC Structural Biology*. 2007;7. doi:10.1186/1472-6807-7-56
- 433 16. Pignatelli D, Carvalho BL, Palmeiro A, Barros A, Guerreiro SG, Maçut D. The
434 complexities in genotyping of congenital adrenal hyperplasia: 21-hydroxylase
435 deficiency. *Frontiers in Endocrinology*. 2019;10(JULY).
436 doi:10.3389/fendo.2019.00432
- 437 17. van der Velde KJ, de Boer EN, van Diemen CC, *et al.* GAVIN: Gene-Aware Variant
438 INterpretation for medical sequencing. *Genome Biology*. 2017;18(1).
439 doi:10.1186/s13059-016-1141-7
- 440 18. Ashkenazy H, Abadi S, Martz E, *et al.* ConSurf 2016: an improved methodology to
441 estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids
442 Research*. 2016;44(W1):W344-W350. doi:10.1093/nar/gkw408
- 443 19. Quang D, Chen Y, Xie X. DANN: A deep learning approach for annotating the
444 pathogenicity of genetic variants. *Bioinformatics*. 2015;31(5):761-763.
445 doi:10.1093/bioinformatics/btu703
- 446 20. Shihab HA, Gough J, Cooper DN, *et al.* Predicting the Functional, Molecular, and
447 Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov
448 Models. *Human Mutation*. 2013;34(1):57-65. doi:10.1002/humu.22225

- 449 21. Stone EA, Sidow A. Physicochemical constraint violation by missense substitutions
450 mediates impairment of protein function and disease severity. *Genome Research*.
451 2005;15(7):978-986. doi:10.1101/gr.3804205
- 452 22. Pejaver V, Urresti J, Lugo-Martinez J, *et al.* Inferring the molecular and phenotypic
453 impact of amino acid variants with MutPred2. *Nature Communications*.
454 2020;11(1). doi:10.1038/s41467-020-19669-x
- 455 23. Tang H, Thomas PD. PANTHER-PSEP: Predicting disease-causing genetic variants
456 using position-specific evolutionary preservation. *Bioinformatics*.
457 2016;32(14):2230-2232. doi:10.1093/bioinformatics/btw222
- 458 24. Capriotti E, Fariselli P. PhD-SNPg: A webserver and lightweight tool for scoring
459 single nucleotide variants. *Nucleic Acids Research*. 2017;45(W1):W247-W252.
460 doi:10.1093/nar/gkx369
- 461 25. Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human
462 Missense Mutations Using PolyPhen-2. *Current Protocols in Human Genetics*.
463 2013;76(1). doi:10.1002/0471142905.hg0720s76
- 464 26. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the Functional Effect of
465 Amino Acid Substitutions and Indels. *PLoS ONE*. 2012;7(10).
466 doi:10.1371/journal.pone.0046688
- 467 27. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for
468 genomes. *Nature Protocols*. 2016;11(1):1-9. doi:10.1038/nprot.2015.123
- 469 28. Hecht M, Bromberg Y, Rost B. Better prediction of functional effects for sequence
470 variants. *BMC Genomics*. 2015;16(S8):S1. doi:10.1186/1471-2164-16-S8-S1
- 471 29. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations
472 improve the predictive score of human disease-related mutations in proteins.
473 *Human Mutation*. 2009;30(8):1237-1244. doi:10.1002/humu.21047
- 474 30. Bendl J, Stourac J, Salanda O, *et al.* PredictSNP: Robust and Accurate Consensus
475 Classifier for Prediction of Disease-Related Mutations. *PLoS Computational*
476 *Biology*. 2014;10(1). doi:10.1371/journal.pcbi.1003440
- 477 31. Bendl J, Musil M, Štourač J, Zendulka J, Damborský J, Brezovský J. PredictSNP2: A
478 Unified Platform for Accurately Evaluating SNP Effects by Exploiting the Different
479 Characteristics of Variants in Distinct Genomic Regions. *PLoS Computational*
480 *Biology*. 2016;12(5). doi:10.1371/journal.pcbi.1004962
- 481 32. Capriotti E, Altman RB, Bromberg Y. Collective judgment predicts disease-
482 associated single nucleotide variants. *BMC genomics*. 2013;14 Suppl 3.
483 doi:10.1186/1471-2164-14-s3-s2
- 484 33. Capriotti E, Altman RB. Improving the prediction of disease-related variants using
485 protein three-dimensional structure. *BMC Bioinformatics*. 2011;12(SUPPL. 4).
486 doi:10.1186/1471-2105-12-S4-S3

- 487 34. Vihinen M. How to evaluate performance of prediction methods? Measures and
488 their interpretation in variation effect analysis. *BMC genomics*. 2012;13 Suppl 4.
489 doi:10.1186/1471-2164-13-S4-S2
- 490 35. Chicco D, Tötsch N, Jurman G. The matthews correlation coefficient (Mcc) is more
491 reliable than balanced accuracy, bookmaker informedness, and markedness in
492 two-class confusion matrix evaluation. *BioData Mining*. 2021;14:1-22.
493 doi:10.1186/s13040-021-00244-z
- 494
- 495

496 **TABLES LEGENDS**

497 Table 1. Performance of 17 programs to predict the effect SNVs in the *CYP21A2*. We
498 performed the analysis with 103 functionally characterized variants, 90 damaging the
499 protein functionality, and 13 neutral—Color scores from blue (good result) to yellow
500 (not good). ^a meta-predictor. nr, no result; PPV, positive predictive value; NPV, negative
501 predictive value; Se, sensibility; Sp, specificity; Ac, accuracy; MCC, Matthews'
502 correlation coefficient test.

503

504 Table 2. Performance of 17 programs for the specific *CYP21A2* groups. We predict the
505 effect of SNVs in the *CYP21A2* dividing them by the two levels of protein damage: the
506 severe (classical mutation, CL group) and mild (non-classical mutation, NC group). We
507 performed the analysis with 103 SNVs of known effect, 51 being CL, 39 NC, and 13
508 neutral. Color score from blue (good result) to yellow (not good). ^a meta-predictor. nr,
509 no result; PPV, positive predictive value; NPV, negative predictive value; Se, sensibility;
510 Sp, specificity; Ac, accuracy; MCC, Matthews' correlation coefficient test.

511 **FIGURES LEGENDS**

512 Figure 4. Composition of the four meta-predictors tested. The PredictSNP algorithm comprises
513 outputs of six single-predictors, the PredictSNP2 of six, the Meta-SNP of 4, and the S3Ds&GO of
514 three predictors.

515

516 Figure 5. The frequency of hit and miss was obtained for each SNV by mutation groups. Each
517 SNV (vertical list) was analyzed by seventeen predictors (horizontal measurement) performed
518 with the default setting for missense mutation. Please, refer to Table S4-S6 for future details.
519 TP, true positive; TN, true negative; FN, false negative; FP, false positive.

520

521 Figure 6. Venn diagram indicating overlaps of the hit for four predictors with good-excellent
522 performance for SNVs on CYP21A2. Total, indicate all 103 SNVs tested. (Image generated with
523 <https://bioinformatics.psb.ugent.be/webtools/Venn/>).

524 **Table 1.** Performance of 17 programs to predict the effect SNVs in the *CYP21A2*. We
 525 performed the analysis with 103 functionally characterized variants, 90 damaging the
 526 protein functionality, and 13 neutral—Color scores from blue (good result) to yellow
 527 (not good).

Predictors	TP	FN	TN	FP	PPV	NPV	Se	Sp	Ac	MCC
PredictSNP ^a	58	32	13	0	1.00	0.29	0.64	1.00	0.69	0.43
PredictSNP2 ^a	44	46	13	0	1.00	0.22	0.49	1.00	0.55	0.33
Meta-SNP ^a	60	30	13	0	1.00	0.30	0.67	1.00	0.71	0.45
S3Ds&GO ^a	58	32	8	0	1.00	0.20	0.64	1.00	0.67	0.36
CADD	86	4	11	2	0.98	0.73	0.96	0.85	0.94	0.75
ConSurf	79	11	9	1	0.99	0.45	0.88	0.90	0.88	0.58
DANN	83	7	9	4	0.95	0.56	0.92	0.69	0.89	0.56
FATHMM	62	28	13	0	1.00	0.32	0.69	1.00	0.73	0.47
MAPP	54	36	10	2	0.96	0.22	0.60	0.83	0.63	0.28
MutPred2	62	28	13	0	1.00	0.32	0.69	1.00	0.73	0.47
PANTHER-PSEP	90	0	0	9	0.91	nr	1.00	0.00	0.91	nr
PhD-SNPg	62	28	12	1	0.98	0.30	0.69	0.92	0.72	0.42
PolyPhen2-HumVar	79	11	12	1	0.99	0.52	0.88	0.92	0.88	0.64
PROVEAN	66	24	13	0	1.00	0.35	0.73	1.00	0.77	0.51
SIFT	70	20	11	2	0.97	0.35	0.78	0.85	0.79	0.45
SNP2	59	31	13	0	1.00	0.30	0.66	1.00	0.70	0.44
SNPs&GO	54	36	13	0	1.00	0.27	0.60	1.00	0.65	0.40

528 ^a meta-predictor. nr, no result; PPV, positive predictive value; NPV, negative predictive
 529 value; Se, sensibility; Sp, specificity; Ac, accuracy; MCC, Matthews' correlation
 530 coefficient test.

531

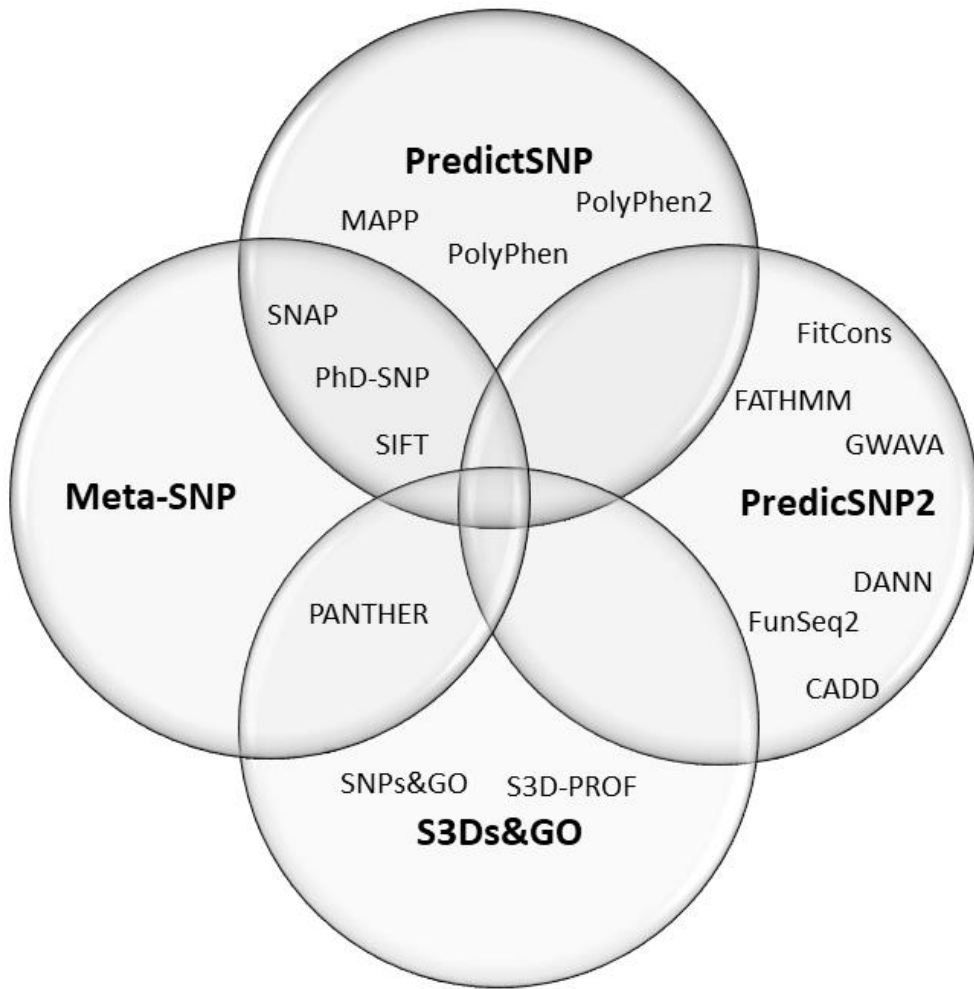
532 **Table 2.** Performance of 17 programs for the specific CYP21A2 groups. We predict the
 533 effect of SNVs in the CYP21A2 dividing them by the two levels of protein damage: the
 534 severe (classical mutation, CL group) and mild (non-classical mutation, NC group). We
 535 performed the analysis with 103 SNVs of known effect, 51 being CL, 39 NC, and 13
 536 neutral. Color score from blue (good result) to yellow (not good).

Specifics groups	PPV		NPV		Se		Sp		Ac		MCC	
	Cl	NC	Cl	NC	Cl	NC	Cl	NC	Cl	NC	Cl	NC
PredictSNP ^a	1.00	1.00	0.65	0.34	0.86	0.36	1.00	1.00	0.89	0.52	0.75	0.35
PredictSNP2 ^a	1.00	1.00	0.39	0.33	0.61	0.33	1.00	1.00	0.69	0.50	0.49	0.33
Meta-SNP ^a	1.00	1.00	0.68	0.35	0.88	0.38	1.00	1.00	0.91	0.54	0.78	0.37
S3Ds&GO ^a	1.00	1.00	0.53	0.24	0.86	0.36	1.00	1.00	0.88	0.47	0.68	0.29
CADD	0.96	0.95	1.00	0.73	1.00	0.90	0.85	0.85	0.97	0.88	0.90	0.71
ConSurf	0.98	0.97	0.82	0.50	0.96	0.77	0.90	0.90	0.95	0.80	0.83	0.56
DANN	0.93	0.89	0.90	0.60	0.98	0.85	0.69	0.69	0.92	0.81	0.75	0.51
FATHMM	1.00	1.00	0.52	0.45	0.76	0.59	1.00	1.00	0.81	0.69	0.63	0.51
MAPP	0.95	0.88	0.48	0.29	0.78	0.36	0.83	0.83	0.79	0.47	0.51	0.18
MutPred2	1.00	1.00	0.72	0.36	0.90	0.41	1.00	1.00	0.92	0.56	0.81	0.38
PANTHER-PSEP	0.85	0.81	nr	nr	1.00	1.00	0.00	0.00	0.85	0.81	nr	nr
PhD-SNPg	0.98	0.95	0.57	0.39	0.82	0.51	0.92	0.92	0.84	0.62	0.64	0.38
PolyPhen2-HumVar	0.96	0.94	0.85	0.52	0.96	0.74	0.85	0.85	0.94	0.77	0.81	0.52
PROVEAN	1.00	1.00	0.76	0.39	0.92	0.49	1.00	1.00	0.94	0.62	0.84	0.44
SIFT	0.96	0.92	0.69	0.42	0.90	0.62	0.85	0.85	0.89	0.67	0.70	0.40
SNP2	1.00	1.00	0.59	0.37	0.82	0.44	1.00	1.00	0.86	0.58	0.70	0.40
SNPs&GO	1.00	1.00	0.59	0.33	0.82	0.31	1.00	1.00	0.86	0.48	0.70	0.32

537 ^a meta-predictor. nr, no result; PPV, positive predictive value; NPV, negative predictive
 538 value; Se, sensibility; Sp, specificity; Ac, accuracy; MCC, Matthews' correlation
 539 coefficient test.

540

Figure 1.



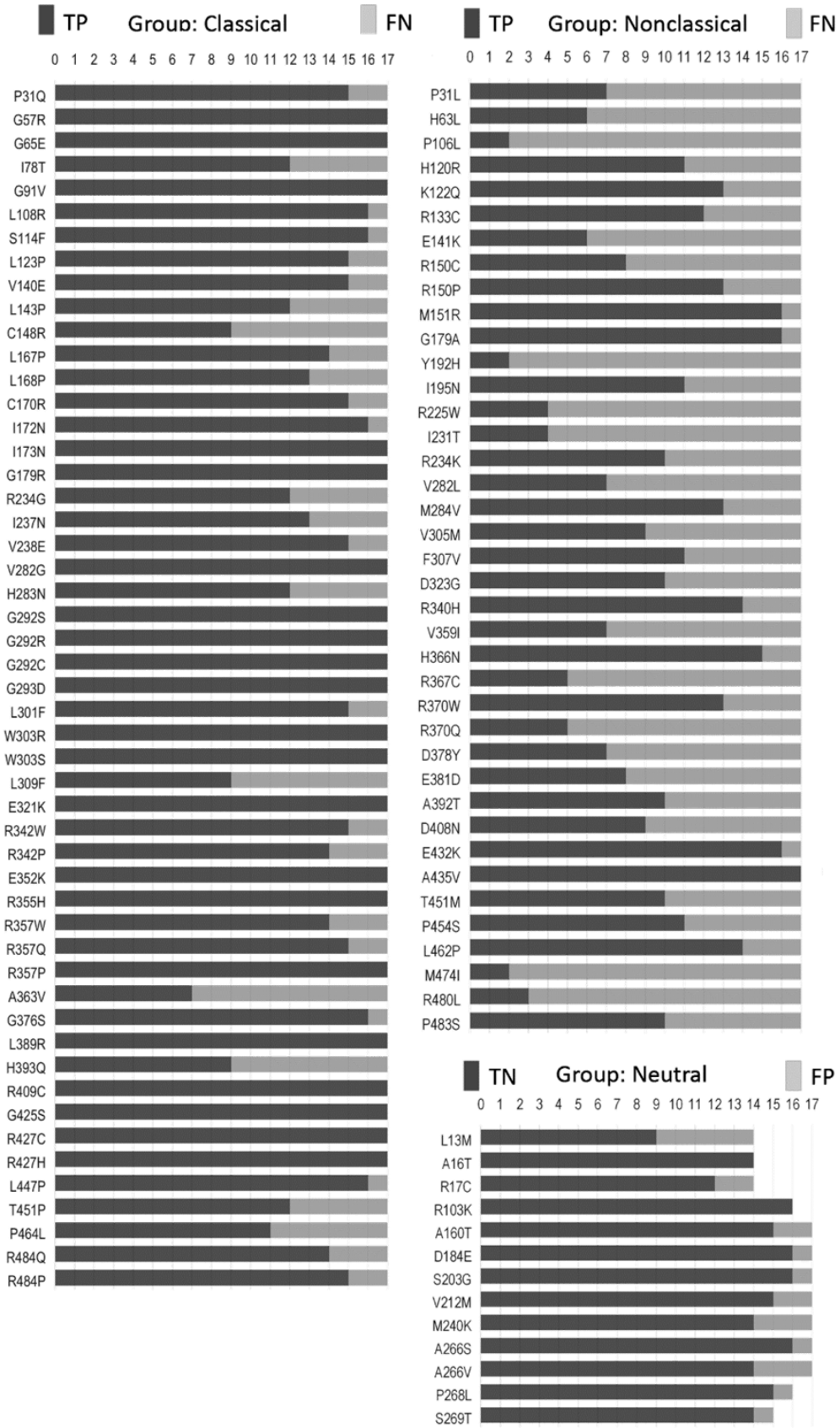
541

542

543

Figure 2.

544



545

Meta analysis of variant predictions in congenital adrenal hyperplasia caused by mutations in CYP21A2

Mayara J. Prado^{1,2,3,4}, Rodrigo Ligabue-Braun⁵, Arnaldo Zaha^{1,2}‡, Maria Lucia Rosa Rossetti^{1,4}‡,

Amit V. Pandey^{3,4} ‡

‡These authors contributed equally to this work.

¹Graduate Program in Cell and Molecular Biology, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre CEP 91501-970, Brazil. ²Center for Biotechnology, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre CEP 91501-970, RS, Brazil. ³Department of Biomedical Research, University of Bern, Bern 3010, Switzerland. ⁴Pediatric Endocrinology Unit, Department of Pediatrics, University Children's Hospital Bern, Bern 3010, Switzerland. ⁵Department of Pharmacosciences, Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), Porto Alegre CEP 90050-170, Brazil. ⁶Graduate Program in Molecular Biology Applied to Health, Universidade Luterana do Brasil (ULBRA), Canoas CEP 92425-020, Brazil.

ORCID numbers: M.J.P (0000-0003-0647-4429), R.L-B. (0000-0002-2555-9754), A.Z. (0000-0001-6336-474X), M.L.R.R (0000-0002-9672-9394), and A.V.P (0000-0001-8331-5902).

<https://boris.unibe.ch/id/eprint/162936>

DOI: <http://dx.doi.org/10.48350/162936>

Table S1. Single predictors selected for performance analysis with *CYP21A2* variants.

Single Predictors	Description	Website	Ref.
CADD	Integrative annotation built based on diverse genomic feature derived from surrounding sequence context, gene model annotation , evolutionary constraint, epigenetic measurements, and functional predictions.	https://cadd.gs.washington.edu/	[1]
ConSurf	Algorithm uses phylogenetic relationships among homologous sequences and the specific dynamics of the analyzed sequence with evolutionary models to estimate the evolutionary rates of the amino acid of the macromolecules and to map them onto the structure and/or sequence.	https://consurf.tau.ac.il/	[2]
DANN	Deep neural network which takes non-linear relationships among features based on diverse genomic derived from surrounding sequence context, gene model annotation , evolutionary constraint, epigenetic measurements, and functional predictions.	https://cbcl.ics.uci.edu/public_data/DANN/	[3]
FATHMM	Evolutionary conservation algorithm which uses homologous sequences with species-specific weighting to predict the protein's tolerance to missense variants.	http://fathmm.biocompute.org.uk/	[4]
MAPP	A statistical framework predictor which uses protein physicochemical characteristics of each amino acid position on the evolutionary variation.	http://mendel.stanford.edu/sidowlab/downloads/MAPP/index.html	[5]
MutPred2	Machine learning-based to predict amino acid substitution through evolutionary , structural , and functional proprieties.	http://mutpred.mutdb.org/	[6]
PANTHER-PSEP	Predict using evolutionary preservation data, measuring though the length of time estimation that a site has been preserved.	http://www.pantherdb.org/tools/csnpscoreForm.jsp	[7]
PhD-SNP ^g	Machine learning algorithm for predicting SNVs in both non-coding and coding regions through evolutionary data .	https://snps.biofold.org/phd-snp/	[8]
PolyPhen-2	It uses human protein evolutionary and structural data to predict amino acid substitution effect on the protein stability and functionality.	http://genetics.bwh.harvard.edu/pph2/	[9]
PROVEAN	Predict the functional effect through amino acid exchange evolutionary data and quality of the neighborhood sequence alignment rather than the target position.	http://provean.jcvi.org/genome_submit_2.php?species=human	[10]
SIFT	Predicts through sequence homology algorithm assuming evolutionary conserved regions tend to be less tolerant.	https://sift.bii.a-star.edu.sg/www/SIFT4G_vc_f_submit.html	[11]
SNAP2	A neural network method based on machine learning to predict the variant effect in the molecular function through evolutionary and structural protein data with an amino acid substitution matrix of effect probabilities.	https://roslab.org/services/snap2web/	[12]
SNPs&GO	Predict using evolutionary data, profile and gene ontology (biological process, cellular component and molecular function). When protein function is not available, it run PANTHER and PhD-SNP.	https://snps.biofold.org/snps-and-go/snps-and-go.html	[13]

Table S2. Performance of predictor tools.

Predictor	PPV	NPV	Se	Sp	Ac	MCC	AUC-ROC	Dataset	Ref.
CADD			93.6	57.1	0.85			ClinVar (2015)	[1]
DANN							0.95	ClinVar (2014)	[3]
FATHMM (weighted)	0.85	0.8	0.78	0.87	0.82	0.65		SwissVar (2012)	[4]
MAPP					0.626-0.767			Experimental studies	[5]
Meta-SNP	0.79	0.8	0.8	0.79	0.79	0.59	0.86	SwissVar (2009-2012)	[14]
MutPred2	96		42.3	95.6			84.9	ClinVar32 (2015) and UniProt80 (2015)	[6]
PANTHER-PSEP							0.721	Derived from SwissVar	[7]
PhD-SNPg	0.85	0.85	0.94	0.67	0.85	0.65	0.91	NewClinvar (2016)	[8]
PolyPhen-2*			0.85	0.6015			0.79	Mutations on the genes BRCA1, MSH2, MLH1 and TP53	[15]
PredictSNP					0.642	0.281	0.7	Protein Mutant Database (07Mar26)	[16]
PredictSNP2					0.773	0.55	0.804	Mendelian diseases (multiple databases)	[17]
PROVEAN			0.78	0.79				UniProt human protein	[10]
SNPs&GO	0.83	0.8	0.78	0.85	0.82	0.63		Derived from Swiss-Prot (2008)	[13]
SIFT 4G			0.8	0.735	0.7732	0.53		UniRef90 (2011)	[11]
SNAP2					0.688	0.24		Data set consisting of 9,657 variants from 678 human proteins	[12]
SNP&GO3d	0.84	0.86	0.87	0.83	0.85	0.7	0.92	Derived from Swiss-Prot (2009)	[18]

*Data from an article recommended on the original developer article. PPV, positive predictive value; NPV, negative predictive value; Se, sensibility; Sp, specificity; Ac, accuracy; MCC, Matthews' correlation coefficient test.

Table S3. List of the 103 single nucleotide variants (SNVs) on *CYP21A2* gene selected to test the performance of predictor tools. SNVs are grouped into classical (enzyme activity < 10%), non-classical (between 10 and 78 %) and neutral (> 78 %) groups. The enzyme activity levels of both 21-hydroxylase substrates - 17-hydroxyprogesterone and progesterone - were obtained from the original paper of the functional characterization. The phenotype was obtained from either the same paper or the original description of the new SNV. ^a Shows the percentage of enzyme activity measured for the conversion of both 21-hydroxylase substrates, considering as 100 % the 21-hydroxylase wild type activity. 17OHP: 17-hydroxyprogesterone. SW: salt wasting. SV: simple-virilizing. NC: non-classical. ND: non-determinate.

Group	NP_000491.4	CYP21A2 Activity <i>in vitro</i>				Phenotype	Publication
		17OHP ^a	SD (±)	Progesterone	SD (±)		
CL	p.P31Q	0.2	0.2	0	0	SW	[19]
	p.G57R	0.7	ND	1.4	ND	SV	[20]
	p.G65E	0	ND	0	ND	SW	[21]
	p.I78T	3	2	5	3	SV	[22]
	p.G91V	0	ND	0	ND	SW	[23]
	p.L108R	0.4	ND	0.3	ND	SW	[20]
	p.S114F	4	1	4	2	SV	[24]
	p.L123P	1.42	2.13	-1.86	5.19	SW	[25]
	p.V140E	0.7	1.3	0.5	0.6	SW	[26]
	p.L143P	0.4		0.4		SW	[20]
	p.C148R	4.3	0.9	3.6ny	1.8	SV-NC	[26]
	p.L167P	0.3	0.06	0.4	0.6	SW	[27]
	p.L168P	0.7	ND	0.4	ND	SW	[28]
	p.C170R	0.1	0.02	0	2	SW	[29]
	p.I172N	0.7	0.3	0.6	0.03	SV	[30]
	p.I173N	4.3	1.7	4.4	1.8	SV	[28]
	p.G179R	0.4	0.5	0	0.6	SW	[29]
	p.R234G	8	2	2	1	SV-NC	[31]
	p.I237N	1	0.7	2.4	1.4	SV	[32]
	p.V238E	0	0	0.1	0.3	SW	[32]
	p.V282G	3.9	1.7	3.9	2	SV	[33]
	p.H283N	1.6	6	2.7	5	SW	[34]
	p.G292C	0	ND	0	ND	SW	[23]
	p.G292R	0.5	0.7	0.7	0.2	SW	[26]
	p.G292S	0.8	0.4	0.8	0.4	SW	[35]
	p.G293D	0.5	0.2	0.7	0.4	SW	[28]
	p.L301F	9.5	6.4	4.4	2.5	SV	[33]
	p.W303S	3	0.3	3	0.5	SV-NC	[36]
	p.W303R	0.1	0.2	0	0.5	SW	[29]
	p.L309F	0.2	0.3	0.1	0.3	SW	[26]
	p.E321K	4.6	1.8	4.5	2.6	SV	[28]
	p.R342P	0.7	0.3	0.7	0.2	SV	[30]
	p.R342W	5	0.4	4	3	SV-NC	[31]
	p.E352K	1.1	0.5	1.2	0.3	SV	[37]
	p.R355H	0	ND	0	ND	SW	[23]
	p.R357P	0.15	0.3	0.15	0.3	SW	[38]
	p.R357Q	0.65	0.44	1.1	0.94	SV	[38]
	p.R357W	0	ND	0	ND	SW	[39]
	p.A363V	0	ND	0	ND	SW	[21]
	p.G376S	1.6	0.8	0.7	0.7	SW	[40]
	p.L389R	1.1	0.6	ND	ND	SW	[41]

Continuation (Table S3)

CL	p.H393Q	2.5	0.6	2.2	0.6	SW	[42]
	p.R409C	1.3	0.5			SW	[20]
	p.G425S	1.6	0.4	2	0.6	SV	[28]
	p.R427C	0	0.5	0	0.6	SW	[29]
	p.R427H	0.5	0.6	0.4	0.2	SW-SV	[30]
	p.L447P	0.5	0.6	0	0.1	SW-SV	[30]
	p.T451P	0.9	ND	0.9	ND	SW	[24]
	p.P464L	2.6	0.8	3	0.5	SV	[43]
	p.R484P	1	0.07	2.2	0.9	SV	[35]
	p.R484Q	1.1	0.7	3.8	1.9	SV	[27]
	Mean	1.52		1.32			
	SD	2.00		1.61			
NC	p.P31L	13	0.2	2	0.6	NC	[31]
	p.H63L	44.5	ND	20.7	ND	NC	[20]
	p.P106L	62	9	64	12	NC	[44]
	p.H120R	31.6	8	32.5	7	NC	[45]
	p.K122Q	14	5	19.5	4	NC	[46]
	p.R133C	35.4	7.4	15.5	2.7	NC	[47]
	p.E141K	11.3	2.4	ND	ND	SW	[41]
	p.R150C	35.8	14.6	47.3	12.9	NC	[47]
	p.R150P	23.4	1.7	16.9	2	NC	[48]
	p.M151R	17.66	1.87	4.57	1.96	NC	[25]
	p.G179A	19	ND	ND	ND	NC	[23]
	p.Y192H	37.1	7	25.8	9	NC	[34]
	p.I195N	33.2	9	46.7	10	NC	[45]
	p.R225W	51.9	9	45.6	8	NC	[49]
	p.I231T	63.1	22.3	70.6	17	NC	[28]
	p.R234K	15	ND	8.1	ND	SV-NC	[28]
	p.V282L	18	3	18	5	NC	[31]
	p.M284V	16.2	9.3	19	6.8	NC	[47]
	p.V305M	46	18	26	10	NC	[50]
	p.F307V	63.23	5.5	64.17	7.98	SV-NC	[51]
	p.D323G	18	1.2	27	4.7	NC	[36]
	p.R340H	67.1	2.4	45.8	3.7	NC	[52]
	p.V359I	72	7	34	3	NC	[53]
	p.H366N	46.13	4.8	57.77	3.69	NC	[51]
	p.R367C	37	7	28	4	NC	[31]
	p.R370Q	82	6	63	4	NC	[53]
	p.R370W	45.8	1.8	48.5	17.1	NC	[28]
	p.D378Y	81	6	58	4	NC	[53]
	p.E381D	30	ND	ND	ND	SW	[54]
	p.A392T	38.7	9.5	22.9	4.7	NC	[55]
	p.D408N	72.7	7	73.6	10	NC	[49]
	p.E432K	26.2	3.8	24.2	7.4	NC	[47]
	p.A435V	14	2	12	6	SV	[22]
	p.T451M	78	6	43	5	NC	[24]
	p.P454S	38	ND	22.4	3	NC	[31]
	p.L462P	55	8	40	2	NC	[53]
p.M474I	85	7	66	12	NC	[31]	
p.R480L	75.5	15.7	79.6	12	NC-Normal	[55]	
p.P483S	61	6	54	2	NC	[31]	
	Mean	42.94		37.41			
	SD	22.59		20.98			

Continuation (Table S3)

Neutral	p.L13M	99	1	100	1	Normal	[24]
	p.A16T	100	0	96	6	Normal- very mildNC	[24]
	p.R17C	95	3	81	3	Normal- very mildNC	[24]
	p.R103K	119.7	22.5	ND	ND	Normal	[41]
	p.A160T	126.6	29.9	ND	ND	Normal	[41]
	p.D184E	100	ND	100	ND	Normal	[56]
	p.S203G	85	2	81	3	Very mild NC	[24]
	p.V212M	99.5	32.4	ND	ND	Normal	[41]
	p.M240K	95.4	24.7	97.7	7.7	Normal	[32]
	p.A266S	90	9	104	15	Normal	[31]
	p.A266V	92	1.4	100	4.3	Normal	[36]
	p.P268L	97	1	87	7	Normal	[24]
	p.S269T	103	15	ND	ND	Normal	[57]
	Average	100.17		94.08			
	SD	10.92		8.25			

Table S4. Result of 17 predictors for 51 classical single nucleotide variants (SNVs) on the CYP21A2 gene. The classical group has an enzyme activity of < 10% of the wild-type activity. The genomic SNV nomenclature is based on the human chromatin remodeling 38 (Chr38). Del: deleterious; N: Neutral; Pby: Probably; Psb: Possible; B: Benign; Dse: Disease; Efc: Effect; P-Del: Proxy-deleterious; P-N: Proxy-neutral; Dmg: Damaging; T: Tolerated; Ptg: Pathogenic; Csv: Conserved; V: Variable; NR: no result.

Chr38	SNP	Meta-SNP	Predict SNP	Predict SNP2	S3Ds&GO	CADD	ConSurf	DANN	FATHMM	MAPP	MutPred2	PANTHER	PhD-SNPg	PolyPhen2	PROVEAN	SIFT	SNAP2	SNPs&GO
g.32038514C>A	p.P31Q	Dse	Del	N	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	N;
g.32038591G>A	p.G57R	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32038616G>A	p.G65E	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32038752T>C	p.I78T	N	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	N	N	N	N
g.32038791G>T	p.G91V	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32039124T>G	p.L108R	Dse	Del	N	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32039142C>T	p.S114F	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	B	Pby	Del	Del	Efc	Dse
g.32039169T>C	p.L123P	Dse	Del	N	Dse	P-Del	Csv	Del	T	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32039220T>A	p.V140E	Dse	Del	N	Dse	P-Del	Csv	Del	T	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32039229T>C	p.L143P	Dse	Del	N	Dse	P-Del	Csv	Del	T	Del	Del	Pby	Ptg	Pby	N	N	N	Dse
g.32039243T>C	p.C148R	N	N	N	N	P-Del	Csv	N	Dmg	Del	Del	Pby	Ptg	Psb	Del	N	N	N
g.32039408T>C	p.L167P	Dse	Del	N	Dse	P-Del	V	Del	T	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32039411T>C	p.L168P	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	B	Pby	N	N	N	Dse
g.32039416T>C	p.C170R	Dse	Del	N	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	B	Pby	Del	Del	Efc	Dse
g.32039423T>A	p.I172N	Dse	Del	N	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32039426T>A	p.I173N	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse

Continuation (Table S4)

Chr38	SNP	Meta-SNP	Predict SNP	Predict SNP2	S3Ds&GO	CADD	ConSurf	DANN	FATHMM	MAPP	MutPred2	PANTHER	PhD-SNPg	PolyPhen2	PROVEAN	SIFT	SNAP2	SNPs&GO
g.32039443G>A	p.G179R	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32039797A>G	p.R234G	Dse	N	N	Dse	P-Del	Csv	Del	T	N	Del	Pby	B	Pby	Del	Del	Efc	Dse
g.32039807T>A	p.I237N	Dse	Del	N	N	P-Del	V	Del	T	Del	Del	Pby	Ptg	Psb	Del	Del	Efc	Dse
g.32039810T>A	p.V238E	Dse	Del	N	Dse	P-Del	Csv	Del	T	Del	Del	Pby	Ptg	Psb	Del	Del	Efc	Dse
g.32040111T>G	p.V282G	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040113C>A	p.H283N	Dse	N	N	Dse	P-Del	Csv	Del	Dmg	N	N	Pby	Ptg	Pby	Del	Del	N	Dse
g.32040140G>A	p.G292S	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040140G>C	p.G292R	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040140G>T	p.G292C	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040144G>A	p.G293D	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040167C>T	p.L301F	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	N	Del	Pby	B	Pby	Del	Del	Efc	Dse
g.32040173T>C	p.W303R	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040174G>C	p.W303S	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040191C>T	p.L309F	N	N	Del	N	P-Del	Csv	Del	Dmg	N	N	Pby	B	Pby	Del	Del	N	N
g.32040427G>A	p.E321K	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040490C>T	p.R342W	Dse	Del	N	Dse	P-Del	Csv	Del	Dmg	N	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040491G>C	p.R342P	Dse	Del	N	Dse	P-Del	Csv	Del	T	Del	Del	Pby	B	Pby	Del	Del	Efc	Dse
g.32040520G>A	p.E352K	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040530G>A	p.R355H	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Psb	Del	Del	Efc	Dse

Continuation (Table S4)

Chr38	SNP	Meta-SNP	Predict SNP	Predict SNP2	S3Ds&GO	CADD	ConSurf	DANN	FATHMM	MAPP	MutPred2	PANTHER	PhD-SNPg	PolyPhen2	PROVEAN	SIFT	SNAP2	SNPs&GO
g.32040535C>T	p.R357W	Dse	Del	N	Dse	P-Del	Csv	Del	T	N	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040536G>A	p.R357Q	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	N	N	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040536G>C	p.R357P	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040554C>T	p.A363V	N	N	N	Dse	P-Del	Csv	Del	T	N	N	Pby	Ptg	Pby	N	N	N	N
g.32040675G>A	p.G376S	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	N	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040715T>G	p.L389R	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040728C>G	p.H393Q	N	N	N	Dse	P-Del	Csv	Del	T	N	Del	Pby	B	B	Del	Del	Efc	N
g.32040871C>T	p.R409C	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Psb	Del	Del	Efc	Dse
g.32040919G>A	p.G425S	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040925C>T	p.R427C	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040926G>A	p.R427H	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040986T>C	p.L447P	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	B	Pby	Del	Del	Efc	Dse
g.32040997A>C	p.T451P	Dse	Del	N	N	P-Del	Csv	Del	T	Del	Del	Pby	Ptg	Psb	Del	Del	N	N
g.32041037C>T	p.P464L	N	N	Del	N	P-Del	Csv	Del	Dmg	Del	N	Pby	Ptg	Pby	Del	Del	N	N
g.32041097G>A	p.R484Q	Dse	Del	Del	N	P-Del	Csv	Del	Dmg	N	Del	Pby	Ptg	Pby	Del	Del	Efc	N
g.32041097G>C	p.R484P	Dse	Del	N	N	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse

Table S5. Result of 17 predictors for 39 non-classical single nucleotide variants (SNVs) on the CYP21A2. The non-classical group has an enzyme activity between >10% and < 78% of the wild-type activity. The genomic SNV nomenclature is based on the human chromatin remodeling 38 (Chr38). Del: deleterious; N: Neutral; Pby: Probably; Psb: Possible; B: Benign; Dse: Disease; Efc: Effect; P-Del: Proxy-deleterious; P-N: Proxy-neutral; Dmg: Damaging; T: Tolerated; Ptg: Pathogenic; Csv: Conserved; V: Variable; NR: no result.

Chr38	SNP	Meta-SNP	Predict SNP	Predict SNP2	S3Ds&GO	CADD	ConSurf	DANN	FATHMM	MAPP	MutPred2	PANTHER	PhD-SNPg	PolyPhen2	PROVEAN	SIFT	SNP2	SNPs&GO
g.32038514C>T	p.P31L	N	N	N	N	P-Del	Csv	N	Dmg	Del	Del	Pby	Ptg	B	N	N	N	N
g.32038610A>T	p.H63L	Dse	N	N	N	P-Del	V	N	T	Del	Del	Pby	B	B	N	N	Efc	N
g.32039118C>T	p.P106L	N	N	N	N	P-N	V	N	T	N	N	Pby	B	B	N	N	Efc	N
g.32039160A>G	p.H120R	N	N	N	N	P-Del	Csv	Del	Dmg	Del	N	Pby	Ptg	Pby	Del	Del	Efc	N
g.32039165A>C	p.K122Q	Dse	Del	N	N	P-Del	Csv	Del	Dmg	Del	N	Pby	Ptg	Pby	Del	Del	Efc	N
g.32039198C>T	p.R133C	Dse	Del	N	Dse	P-Del	V	Del	T	N	N	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32039222G>A	p.E141K	N	N	N	N	P-Del	V	Del	T	N	Del	Pby	Ptg	Psb	N	N	N	N
g.32039356C>T	p.R150C	N	N	Del	Dse	P-Del	Csv	Del	Dmg	N	N	Pby	B	Pby	N	N	N	N
g.32039357G>C	p.R150P	Dse	Del	N	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	N	N	N	Dse
g.32039360T>G	p.M151R	Dse	Del	N	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Psb	Del	Del	Efc	Dse
g.32039444G>C	p.G179A	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	N	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32039570T>C	p.Y192H	N	N	N	N	P-N	Csv	N	T	N	N	Pby	B	B	N	N	N	N
g.32039580T>A	p.I195N	N	N	N	Dse	P-Del	Csv	Del	T	N	Del	Pby	Ptg	Pby	N	Del	Efc	Dse
g.32039770C>T	p.R225W	Dse	N	N	N	P-N	V	N	T	N	Del	Pby	B	B	Del	N	N	N
g.32039789T>C	p.I231T	N	N	N	N	P-Del	V	Del	T	N	N	Pby	B	B	N	Del	N	N
g.32039798G>A	p.R234K	N	N	N	Dse	P-Del	Csv	Del	Dmg	N	N	Pby	Ptg	Pby	N	Del	Efc	N

Continuation (Table S5)

Chr38	SNP	Meta-SNP	Predict SNP	Predict SNP2	S3Ds&GO	CADD	ConSurf	DANN	FATHMM	MAPP	MutPred2	PANTHER	PhD-SNPg	PolyPhen2	PROVEAN	SIFT	SNP2	SNPs&GO
g.32040110G>T	p.V282L	N	N	N	N	P-Del	Csv	Del	Dmg	N	N	Pby	Ptg	Psb	N	N	N	N
g.32040116A>G	p.M284V	N	N	Del	Dse	P-Del	Csv	Del	Dmg	N	N	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040179G>A	p.V305M	Dse	N	Del	N	P-Del	Csv	Del	Dmg	N	N	Pby	B	Pby	N	Del	N	N
g.32040185T>G	p.F307V	N	N	N	Dse	P-Del	Csv	Del	Dmg	N	Del	Pby	B	Pby	Del	Del	N	Dse
g.32040434A>G	p.D323G	Dse	Del	N	N	P-Del	V	Del	T	Del	Del	Pby	B	Pby	Del	Del	N	N
g.32040485G>A	p.R340H	Dse	Del	Del	N	P-Del	Csv	Del	Dmg	N	Del	Pby	B	Pby	Del	Del	Efc	Dse
g.32040541G>A	p.V359I	N	N	N	N	P-Del	Csv	Del	Dmg	N	N	Pby	B	Psb	N	Del	N	N
g.32040562C>A	p.H366N	Dse	Del	N	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	N
g.32040565C>T	p.R367C	N	N	N	N	P-Del	Csv	Del	T	N	Del	Pby	B	B	N	N	N	N
g.32040574C>T	p.R370W	Dse	Del	N	Dse	P-Del	Csv	Del	T	Del	N	Pby	B	Pby	Del	Del	Efc	Dse
g.32040575G>A	p.R370Q	N	N	N	N	P-Del	Csv	Del	T	N	N	Pby	B	Psb	N	N	N	N
g.32040681G>T	p.D378Y	N	N	N	N	P-Del	Csv	Del	T	N	Del	Pby	B	Psb	Del	N	N	N
g.32040692G>C	p.E381D	N	N	Del	N	P-Del	Csv	Del	Dmg	N	N	Pby	B	B	N	Del	Efc	N
g.32040723G>A	p.A392T	N	N	Del	Dse	P-Del	Csv	Del	Dmg	Del	N	Pby	B	Pby	N	N	N	Dse
g.32040771G>A	p.D408N	N	N	Del	N	P-Del	Csv	Del	Dmg	N	N	Pby	Ptg	Pby	N	Del	N	N
g.32040940G>A	p.E432K	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	N	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040950C>T	p.A435V	Dse	Del	Del	Dse	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	Efc	Dse
g.32040998C>T	p.T451M	N	Del	N	N	P-Del	Csv	Del	T	Del	N	Pby	Ptg	Psb	Del	Del	N	N
g.32041006C>T	p.P454S	N	Del	Del	N	P-Del	Csv	Del	Dmg	N	N	Pby	Ptg	Pby	Del	Del	N	N

Continuation (Table S5)

Chr38	SNP	Meta-SNP	Predict SNP	Predict SNP2	S3Ds&GO	CADD	ConSurf	DANN	FATHMM	MAPP	MutPred2	PANTHER	PhD-SNPg	PolyPhen2	PROVEAN	SIFT	SNP2	SNPs&GO
g.32041031T>C	p.L462P	Dse	Del	Del	N	P-Del	Csv	Del	Dmg	Del	Del	Pby	Ptg	Pby	Del	Del	N	N
g.32041068G>T	p.M474I	N	N	N	N	P-N	V	N	T	N	N	Pby	B	B	Del	N	N	N
g.32041085G>T	p.R480L	N	N	N	N	P-Del	V	Del	T	N	N	Pby	B	B	N	N	N	N
g.32041093C>T	p.P483S	N	N	Del	N	P-Del	Csv	Del	Dmg	N	N	Pby	Ptg	Pby	N	Del	Efc	N

Table S6. Result of 17 predictors for 13 neutral single nucleotide variants (SNVs) on the CYP21A2 gene. The neutral group has the enzyme activity known as > 78% of the wild-type activity. The genomic SNV nomenclature is based on the human chromatin remodeling 38 (Chr38). Del: deleterious; N: Neutral; Pby: Probably; Psb: Possible; Dse: Disease; Efc: Effect; P-Del: Proxy-deleterious; P-N: Proxy-neutral; Dmg: Damaging; T: Tolerated; Ptg: Pathogenic; B: Benign; Csv: Conserved; V: Variable; NR: no result.

Chr38	SNP	Meta-SNP	Predict SNP	Predict SNP2	S3Ds&GO	CADD	ConSurf	DANN	FATHMM	MAPP	MutPred2	PANTHER	PhD-SNPg	PolyPhen2	PROVEAN	SIFT	SNP2	SNPs&GO
g.32038459C>A	p.L13M	N	N	N	NR	P-Del	NR	Del	T	Del	N	NR	B	Pby	N	Del	N	N
g.32038468G>A	p.A16T	N	N	N	NR	P-N	NR	N	T	N	N	NR	B	B	N	N	N	N
g.32038471C>T	p.R17C	N	N	N	NR	P-N	NR	Del	T	N	N	NR	B	B	N	Del	N	N
g.32039109G>A	p.R103K	N	N	N	N	P-N	V	N	T	N	N	NR	B	B	N	N	N	N
g.32039386G>A	p.A160T	N	N	N	N	P-N	Csv	N	T	N	N	Pby	B	B	N	N	N	N
g.32039548C>G	p.D184E	N	N	N	N	P-N	V	N	T	N	N	Pby	B	B	N	N	N	N
g.32039603A>G	p.S203G	N	N	N	N	P-N	V	N	T	N	N	Pby	B	B	N	N	N	N
g.32039630G>A	p.V212M	N	N	N	N	P-N	V	Del	T	N	N	Pby	B	B	N	N	N	N
g.32039816T>A	p.M240K	N	N	N	N	P-Del	V	N	T	N	N	Pby	Ptg	B	N	N	N	N
g.32040062G>T	p.A266S	N	N	N	N	P-N	V	N	T	N	N	Pby	B	B	N	N	N	N
g.32040063C>T	p.A266V	N	N	N	N	P-N	V	Del	T	Del	N	Pby	B	B	N	N	N	N
g.32040069C>T	p.P268L	N	N	N	NR	P-N	V	N	T	N	N	Pby	B	B	N	N	N	N
g.32040072G>C	p.S269T	N	N	N	NR	P-N	V	N	T	NR	N	Pby	B	B	N	N	N	N

Reference

1. van der Velde, K.J.; de Boer, E.N.; van Diemen, C.C.; Sikkema-Raddatz, B.; Abbott, K.M.; Knopperts, A.; Franke, L.; Sijmons, R.H.; de Koning, T.J.; Wijmenga, C.; *et al.* GAVIN: Gene-Aware Variant INterpretation for Medical Sequencing. *Genome Biology* **2017**, *18*, doi:10.1186/s13059-016-1141-7.
2. Ashkenazy, H.; Abadi, S.; Martz, E.; Chay, O.; Mayrose, I.; Pupko, T.; Ben-Tal, N. ConSurf 2016: An Improved Methodology to Estimate and Visualize Evolutionary Conservation in Macromolecules. *Nucleic Acids Research* **2016**, *44*, W344–W350, doi:10.1093/nar/gkw408.
3. Quang, D.; Chen, Y.; Xie, X. DANN: A Deep Learning Approach for Annotating the Pathogenicity of Genetic Variants. *Bioinformatics* **2015**, *31*, 761–763, doi:10.1093/bioinformatics/btu703.
4. Shihab, H.A.; Gough, J.; Cooper, D.N.; Stenson, P.D.; Barker, G.L.A.; Edwards, K.J.; Day, I.N.M.; Gaunt, T.R. Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions Using Hidden Markov Models. *Human Mutation* **2013**, *34*, 57–65, doi:10.1002/humu.22225.
5. Stone, E.A.; Sidow, A. Physicochemical Constraint Violation by Missense Substitutions Mediates Impairment of Protein Function and Disease Severity. *Genome Research* **2005**, *15*, 978–986, doi:10.1101/gr.3804205.
6. Pejaver, V.; Urresti, J.; Lugo-Martinez, J.; Pagel, K.A.; Lin, G.N.; Nam, H.J.; Mort, M.; Cooper, D.N.; Sebat, J.; Lakoucheva, L.M.; *et al.* Inferring the Molecular and Phenotypic Impact of Amino Acid Variants with MutPred2. *Nature Communications* **2020**, *11*, doi:10.1038/s41467-020-19669-x.
7. Tang, H.; Thomas, P.D. PANTHER-PSEP: Predicting Disease-Causing Genetic Variants Using Position-Specific Evolutionary Preservation. *Bioinformatics* **2016**, *32*, 2230–2232, doi:10.1093/bioinformatics/btw222.
8. Capriotti, E.; Fariselli, P. PhD-SNPg: A Webserver and Lightweight Tool for Scoring Single Nucleotide Variants. *Nucleic Acids Research* **2017**, *45*, W247–W252, doi:10.1093/nar/gkx369.
9. Adzhubei, I.; Jordan, D.M.; Sunyaev, S.R. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Current Protocols in Human Genetics* **2013**, *76*, doi:10.1002/0471142905.hg0720s76.
10. Choi, Y.; Sims, G.E.; Murphy, S.; Miller, J.R.; Chan, A.P. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE* **2012**, *7*, doi:10.1371/journal.pone.0046688.
11. Vaser, R.; Adusumalli, S.; Leng, S.N.; Sikic, M.; Ng, P.C. SIFT Missense Predictions for Genomes. *Nature Protocols* **2016**, *11*, 1–9, doi:10.1038/nprot.2015.123.
12. Hecht, M.; Bromberg, Y.; Rost, B. Better Prediction of Functional Effects for Sequence Variants. *BMC Genomics* **2015**, *16*, doi:10.1186/1471-2164-16-S8-S1.

13. Calabrese, R.; Capriotti, E.; Fariselli, P.; Martelli, P.L.; Casadio, R. Functional Annotations Improve the Predictive Score of Human Disease-Related Mutations in Proteins. *Human Mutation* **2009**, *30*, 1237–1244, doi:10.1002/humu.21047.
14. Capriotti, E.; Altman, R.B.; Bromberg, Y. Collective Judgment Predicts Disease-Associated Single Nucleotide Variants. *BMC genomics* **2013**, *14 Suppl 3*, doi:10.1186/1471-2164-14-s3-s2.
15. Hicks, S.; Wheeler, D.A.; Plon, S.E.; Kimmel, M. Prediction of Missense Mutation Functionality Depends on Both the Algorithm and Sequence Alignment Employed. *Human Mutation* **2011**, *32*, 661–668, doi:10.1002/humu.21490.
16. Bendl, J.; Stourac, J.; Salanda, O.; Pavelka, A.; Wieben, E.D.; Zendulka, J.; Brezovsky, J.; Damborsky, J. PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. *PLoS Computational Biology* **2014**, *10*, doi:10.1371/journal.pcbi.1003440.
17. Bendl, J.; Musil, M.; Štourač, J.; Zendulka, J.; Damborský, J.; Brezovský, J. PredictSNP2: A Unified Platform for Accurately Evaluating SNP Effects by Exploiting the Different Characteristics of Variants in Distinct Genomic Regions. *PLoS Computational Biology* **2016**, *12*, doi:10.1371/journal.pcbi.1004962.
18. Capriotti, E.; Altman, R.B. Improving the Prediction of Disease-Related Variants Using Protein Three-Dimensional Structure. *BMC Bioinformatics* **2011**, *12*, doi:10.1186/1471-2105-12-S4-S3.
19. Lajic, S.; Nikoshkov, A.; Holst, M.; Wedell, A. *Effects of Missense Mutations and Deletions on Membrane Anchoring and Enzyme Function of Human Steroid 21-Hydroxylase (P450c21)*; 1999;
20. Soardi, F.C.; Barbaro, M.; Lau, I.F.; Lemos-Marini, S.H. v.; Baptista, M.T.M.; Guerra-Junior, G.; Wedell, A.; Lajic, S.; de Mello, M.P. Inhibition of CYP21A2 Enzyme Activity Caused by Novel Missense Mutations Identified in Brazilian and Scandinavian Patients. *The Journal of Clinical Endocrinology & Metabolism* **2008**, *93*, 2416–2420, doi:10.1210/jc.2007-2594.
21. Ohlsson, G.; Muller, J.; Skakkebaek, N.E.; Schwartz, M. Steroid 21-Hydroxylase Deficiency: Mutational Spectrum in Denmark, Three Novel Mutations, and in Vitro Expression Analysis. *Human Mutation* **1999**, *13*, 482–486, doi:10.1002/(SICI)1098-1004(1999)13:6<482::AID-HUMU8>3.0.CO;2-0.
22. Krone, N.; Riepe, F.G.; Grötzinger, J.; Partsch, C.-J.; Sippell, W.G. Functional Characterization of Two Novel Point Mutations in the *CYP21* Gene Causing Simple Virilizing Forms of Congenital Adrenal Hyperplasia Due to 21-Hydroxylase Deficiency. *The Journal of Clinical Endocrinology & Metabolism* **2005**, *90*, 445–454, doi:10.1210/jc.2004-0813.
23. Nunez, B.S.; Lobato, M.N.; White, P.C.; Meseguer, A. Functional Analysis of Four *CYP21* Mutations from Spanish Patients with Congenital Adrenal Hyperplasia. *Biochemical and Biophysical Research Communications* **1999**, *262*, 635–637, doi:10.1006/bbrc.1999.1271.
24. de Paula Michelatto, D.; Karlsson, L.; Lusa, A.L.G.; Silva, C.D.M.; Östberg, L.J.; Persson, B.; Guerra-Junior, G.; Lemos-Marini, S.H.V. de; Barbaro, M.; de Mello, M.P.; *et al.* Functional and Structural

Consequences of Nine *CYP21A2* Mutations Ranging from Very Mild to Severe Effects. *International Journal of Endocrinology* **2016**, 2016, 1–10, doi:10.1155/2016/4209670.

25. Massimi, A.; Malaponti, M.; Federici, L.; Vinciguerra, D.; Manca Bitti, M.; Vottero, A.; Ghizzoni, L.; Maccarrone, M.; Cappa, M.; Bernardini, S.; *et al.* Functional and Structural Analysis of Four Novel Mutations of *CYP21A2* Gene in Italian Patients with 21-Hydroxylase Deficiency. *Hormone and Metabolic Research* **2014**, 46, 515–520, doi:10.1055/s-0034-1371864.
26. Barbaro, M.; Soardi, F.C.; Palandi de Mello, M.; Wedell, A.; Lajic, S. Functional Studies of *CYP21A2* Mutants Complement Structural and Clinical Predictions of Disease Severity in CAH. *Clinical Endocrinology* **2012**, 76, 766–768, doi:10.1111/j.1365-2265.2011.04275.x.
27. Robins, T.; Bellanne-Chantelot, C.; Barbaro, M.; Cabrol, S.; Wedell, A.; Lajic, S. Characterization of Novel Missense Mutations in *CYP21* Causing Congenital Adrenal Hyperplasia. *Journal of Molecular Medicine* **2007**, 85, 247–255, doi:10.1007/s00109-006-0121-x.
28. Tardy, V.; Menassa, R.; Sulmont, V.; Lienhardt-Roussie, A.; Lecointre, C.; Brauner, R.; David, M.; Morel, Y. Phenotype-Genotype Correlations of 13 Rare *CYP21A2* Mutations Detected in 46 Patients Affected with 21-Hydroxylase Deficiency and in One Carrier. *Journal of Clinical Endocrinology and Metabolism* **2010**, 95, 1288–1300, doi:10.1210/jc.2009-1202.
29. Grischuk, Y.; Rubtsov, P.; Riepe, F.G.; Grötzinger, J.; Beljelarskaia, S.; Prassolov, V.; Kalintchenko, N.; Semitcheva, T.; Peterkova, V.; Tiulpakov, A.; *et al.* Four Novel Missense Mutations in the *CYP21A2* Gene Detected in Russian Patients Suffering from the Classical Form of Congenital Adrenal Hyperplasia: Identification, Functional Characterization, and Structural Analysis. *The Journal of Clinical Endocrinology & Metabolism* **2006**, 91, 4976–4980, doi:10.1210/jc.2006-0777.
30. Barbaro, M.; Baldazzi, L.; Balsamo, A.; Lajic, S.; Robins, T.; Barp, L.; Pirazzoli, P.; Cacciari, E.; Cicognani, A.; Wedell, A. Functional Studies of Two Novel and Two Rare Mutations in the 21-Hydroxylase Gene. *Journal of Molecular Medicine* **2006**, 84, 521–528, doi:10.1007/s00109-006-0043-7.
31. Barbaro, M.; Soardi, F.C.; Östberg, L.J.; Persson, B.; de Mello, M.P.; Wedell, A.; Lajic, S. In Vitro Functional Studies of Rare *CYP21A2* Mutations and Establishment of an Activity Gradient for Nonclassic Mutations Improve Phenotype Predictions in Congenital Adrenal Hyperplasia. *Clinical Endocrinology* **2015**, 82, 37–44, doi:10.1111/cen.12526.
32. Robins, T.; Barbaro, M.; Lajic, S.; Wedell, A. Not All Amino Acid Substitutions of the Common Cluster E6 Mutation in *CYP21* Cause Congenital Adrenal Hyperplasia. *The Journal of Clinical Endocrinology & Metabolism* **2005**, 90, 2148–2153, doi:10.1210/jc.2004-1937.
33. Lajić, S.; Robins, T.; Krone, N.; Schwarz, H.P.; Wedell, A. *CYP21* Mutations in Simple Virilizing Congenital Adrenal Hyperplasia. *Journal of Molecular Medicine* **2001**, 79, 581–586, doi:10.1007/s001090100261.
34. Concolino, P.; Mello, E.; Patrosso, M.C.; Penco, S.; Zuppi, C.; Capoluongo, E. P.H282N and p.Y191H: 2 Novel *CYP21A2* Mutations in Italian Congenital Adrenal Hyperplasia Patients. *Metabolism* **2012**, 61, 519–524, doi:10.1016/j.metabol.2011.08.008.

35. Nikoshkov, A.; Lajic, S.; Vlamis-Gardikas, A.; Tranebjærg, L.; Holst, M.; Wedell, A.; Luthman, H. Naturally Occurring Mutants of Human Steroid 21-Hydroxylase (P450c21) Pinpoint Residues Important for Enzyme Activity and Stability. *Journal of Biological Chemistry* **1998**, *273*, 6163–6165, doi:10.1074/jbc.273.11.6163.
36. Bleicken, C.; Loidi, L.; Dhir, V.; Parajes, S.; Quinteiro, C.; Dominguez, F.; Grötzinger, J.; Sippell, W.G.; Riepe, F.G.; Arlt, W.; *et al.* Functional Characterization of Three CYP21A2 Sequence Variants (p.A265V, p.W302S, p.D322G) Employing a Yeast Co-Expression System. *Human Mutation* **2009**, *30*, E443–E450, doi:10.1002/humu.20926.
37. Krone, N.; Riepe, F.G.; Grötzinger, J.; Partsch, C.-J.; Brämswig, J.; Sippell, W.G. The Residue E351 Is Essential for the Activity of Human 21-Hydroxylase: Evidence from a Naturally Occurring Novel Point Mutation Compared with Artificial Mutants Generated by Single Amino Acid Substitutions. *Journal of Molecular Medicine* **2005**, *83*, 561–568, doi:10.1007/s00109-005-0655-3.
38. Lajic, S.; Levo, A.; Nikoshkov, A.; Lundberg, Y.; Partanen, J.; Wedell, A. A Cluster of Missense Mutations at Arg356 of Human Steroid 21-Hydroxylase May Impair Redox Partner Interaction. *Human Genetics* **1997**, *99*, 704–709, doi:10.1007/s004390050436.
39. Chiou, S.H.; Hu, M.C.; CHung, B. A Missense Mutation of Ile172 → Asn or Arg356 → Trp Causes Steroid 21-Hydroxylase Deficiency. *Journal of Biological Chemistry* **1990**, *265*, 3549–3552, doi:10.1016/s0021-9258(19)39804-7.
40. Lajić, S.; Clauin, S.; Robins, T.; Vexiau, P.; Blanché, H.; Bellanne-Chantelot, C.; Wedell, A. Novel Mutations in CYP21 Detected in Individuals with Hyperandrogenism. *The Journal of Clinical Endocrinology & Metabolism* **2002**, *87*, 2824–2829, doi:10.1210/jcem.87.6.8525.
41. Brønstad, I.; Breivik, L.; Methlie, P.; Wolff, A.S.B.; Bratland, E.; Nerموen, I.; Løvås, K.; Husebye, E.S. Functional Studies of Novel CYP21A2 Mutations Detected in Norwegian Patients with Congenital Adrenal Hyperplasia. *Endocrine Connections* **2014**, *3*, 67–74, doi:10.1530/ec-14-0032.
42. Xu, C.; Jia, W.; Cheng, X.; Ying, H.; Chen, J.; Xu, J.; Guan, Q.; Zhou, X.; Zheng, D.; Li, G.; *et al.* Genotype–Phenotype Correlation Study and Mutational and Hormonal Analysis in a Chinese Cohort with 21-hydroxylase Deficiency. *Molecular Genetics & Genomic Medicine* **2019**, *7*, doi:10.1002/mgg3.671.
43. Krone, N.; Riepe, F.; Partsch, C.-J.; Vorhoff, W.; Brämswig, J.; Sippell, W. Three Novel Point Mutations of the CYP21 Gene Detected in Classical Forms of Congenital Adrenal Hyperplasia Due to 21-Hydroxylase Deficiency. *Experimental and Clinical Endocrinology & Diabetes* **2006**, *114*, 111–117, doi:10.1055/s-2005-872841.
44. Nikoshkov, A.; Lajic, S.; Holst, M.; Wedell, A.; Luthman, H. Synergistic Effect of Partially Inactivating Mutations in Steroid 21-Hydroxylase Deficiency¹. *The Journal of Clinical Endocrinology & Metabolism* **1997**, *82*, 194–199, doi:10.1210/jcem.82.1.3678.
45. Concolino, P.; Vendittelli, F.; Mello, E.; Alinovi, C.C.; Minucci, A.; Carrozza, C.; Santini, S.A.; Zuppi, C.; Capoluongo, E. Two Novel CYP21A2 Missense Mutations in Italian Patients with 21-

- Hydroxylase Deficiency: Identification and Functional Characterisation. *IUBMB Life* **2009**, *61*, 229–235, doi:10.1002/iub.147.
46. Riepe, F.G.; Hiort, O.; Grötzinger, J.; Sippell, W.G.; Krone, N.; Holterhus, P.-M. Functional and Structural Consequences of a Novel Point Mutation in the *CYP21A2* Gene Causing Congenital Adrenal Hyperplasia: Potential Relevance of Helix C for P450 Oxidoreductase-21-Hydroxylase Interaction. *The Journal of Clinical Endocrinology & Metabolism* **2008**, *93*, 2891–2895, doi:10.1210/jc.2007-2646.
 47. Taboas, M.; Gómez Acuña, L.; Scaia, M.F.; Bruque, C.D.; Buzzalino, N.; Stivel, M.; Ceballos, N.R.; Dain, L. Functional Studies of p.R132C, p.R149C, p.M283V, p.E431K, and a Novel c.652-2A>G Mutations of the *CYP21A2* Gene. *PLoS ONE* **2014**, *9*, e92181, doi:10.1371/journal.pone.0092181.
 48. Chu, X.; Ding, H.; Cui, G.; Xu, Y.; Wang, D.W.; He, Y. Functional Consequences of a Novel Point Mutation in the *CYP21A2* Gene Identified in a Chinese Han Patient with Nonclassic 21-Hydroxylase Deficiency. *Clinical Endocrinology* **2014**, *80*, 927–928, doi:10.1111/cen.12309.
 49. Concolino, P.; Vendittelli, F.; Mello, E.; Minucci, A.; Carrozza, C.; Rossodivita, A.; Giardina, B.; Zuppi, C.; Capoluongo, E. Functional Analysis of Two Rare *CYP21A2* Mutations Detected in Italian Patients with a Mildest Form of Congenital Adrenal Hyperplasia. *Clinical Endocrinology* **2009**, *71*, 470–476, doi:10.1111/j.1365-2265.2008.03517.x.
 50. Lajić, S.; Clauin, S.; Robins, T.; Vexiau, P.; Blanché, H.; Bellanne-Chantelot, C.; Wedell, A. Novel Mutations in *CYP21* Detected in Individuals with Hyperandrogenism. *The Journal of Clinical Endocrinology & Metabolism* **2002**, *87*, 2824–2829, doi:10.1210/jcem.87.6.8525.
 51. Khajuria, R.; Walia, R.; Bhansali, A.; Prasad, R. Functional Characterization and Molecular Modeling of the Mutations in *CYP21A2* Gene from Patients with Congenital Adrenal Hyperplasia. *Biochimie* **2018**, *149*, 115–121, doi:10.1016/j.biochi.2018.04.012.
 52. Helmberg, A.; Tusie-Luna, M.T.; Tabarelli, M.; Kofler, R.; White, P.C. R339H and P453S: *CYP21* Mutations Associated with Nonclassic Steroid 21-Hydroxylase Deficiency That Are Not Apparent Gene Conversions. *Molecular Endocrinology* **1992**, *6*, 1318–1322, doi:10.1210/mend.6.8.1406709.
 53. Karlsson, L.; de Paula Michelatto, D.; Lusa, A.L.G.; D’Almeida Mgnani Silva, C.; Östberg, L.J.; Persson, B.; Guerra-Júnior, G.; Valente de Lemos-Marini, S.H.; Baldazzi, L.; Menabó, S.; *et al.* Novel Non-Classic *CYP21A2* Variants, Including Combined Alleles, Identified in Patients with Congenital Adrenal Hyperplasia. *Clinical Biochemistry* **2019**, *73*, 50–56, doi:10.1016/j.clinbiochem.2019.07.009.
 54. Hsu, N.-C.; Guzov, V.M.; Hsu, L.-C.; Chung, B. Characterization of the Consequence of a Novel Glu-380 to Asp Mutation by Expression of Functional P450c21 in Escherichia Coli. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1999**, *1430*, 95–102, doi:10.1016/S0167-4838(98)00271-4.

55. Robins, T.; Bellanne-Chantelot, C.; Barbaro, M.; Cabrol, S.; Wedell, A.; Lajic, S. Characterization of Novel Missense Mutations in CYP21 Causing Congenital Adrenal Hyperplasia. *Journal of Molecular Medicine* **2007**, *85*, 247–255, doi:10.1007/s00109-006-0121-x.
56. Higashi, Y.; Hiromasa, T.; Tanae, A.; Miki, T.; Nakura, J.; Kondo, T.; Ohura, T.; Ogawa, E.; Nakayama, K.; Fujii-Kuriyama, Y. Effects of Individual Mutations in the P-450(C21) Pseudogene on the P-450(C21) Activity and Their Distribution in the Patient Genomes of Congenital Steroid 21-Hydroxylase Deficiency1. *The Journal of Biochemistry* **1991**, *109*, 638–644, doi:10.1093/oxfordjournals.jbchem.a123433.
57. Wu, D.A.; Chung, B.C. Mutations of P450c21 (Steroid 21-Hydroxylase) at Cys428, Val281, and Ser268 Result in Complete, Partial, or No Loss of Enzymatic Activity, Respectively. *Journal of Clinical Investigation* **1991**, *88*, 519–523, doi:10.1172/JCI115334.

Capítulo V – Discussão geral

A deficiência da enzima CYP21A2 é uma doença autossômica recessiva causada por defeitos na biossíntese do cortisol. O gene *CYP21A2* possui mais de 380 mutações conhecidas, sendo as variantes *missense* responsáveis por mais da metade de suas alterações que afetam a saúde humana (SIMONETTI *et al.*, 2018). A predição computacional junto com os estudos funcionais constitui-se na principal forma de classificar as mutações e entender as suas relações com os efeitos causadores de doenças.

No presente estudo, caracterizamos seis variantes pontuais no gene *CYP21A2* já relatadas anteriormente, mas não funcionalmente estudadas, p.P35L, p.L199P, p.W202R, p.E352V, p.P433L e p.R484L. Construímos modelos computacionais de estruturas de proteínas CYP21A2 mutadas para entender o efeito dessas variantes na estrutura da proteína. Ademais, expressamos o gene *CYP21A2* selvagem e mutado em células HEK293t para entender os efeitos das mutações na funcionalidade da proteína. As variantes p.P35L e p.L199P mostraram uma conversão parcial de progesterona em DOC de 13% e 10% em relação à selvagem, respectivamente, comparável a bem conhecida mutação NC p.V282L. Portanto, essas duas variantes puderam ser categorizadas como mutação do grupo C e, conseqüentemente, relacionadas ao fenótipo HAC-NC (SPEISER *et al.*, 1993; WEDELL *et al.*, 1994; KRONE *et al.*, 2013). Em cerca de 90% das variantes genéticas relacionadas ao fenótipo HAC-NC, a atividade enzimática está entre 10-78% (SIMONETTI *et al.*, 2013). Por outro lado, p.W202R e p.E352V apresentaram a atividade da CYP21A2 de 1,2% e 1,1% relativas à selvagem, respectivamente, a mesma obtida para a bem conhecida mutação clássica p.I173N. Adicionalmente, p.P433L e p.R484L obtiveram a atividade de 7,5% e 3,4% em relação a enzima

selvagem, respectivamente. Portanto, de acordo com os estudos de correlação genótipo-fenótipo, p.W202R, p.E352V, p.P433L e p.R484L puderam ser categorizadas como variantes do grupo de mutação B e relacionadas ao fenótipo HAC-VS (SPEISER *et al.*, 199; WEDELL *et al.*, 1994; KRONE *et al.*, 2013). Em cerca de 95% das variantes genéticas no gene *CYP21A2* relacionadas às formas clássicas de CAH, a atividade da enzima é < 10% da selvagem (SIMONETTI *et al.*, 2013).

Duas das variantes estudadas, p.P35L e p.R484L, estão localizadas em uma *coil* e possuem alta conservação entre as proteínas da família de citocromo P450 (CYPs) de mamíferos. O número de mutações tolerado dentro de uma *coil* é significativamente menor do que em α -hélices e β -fitas, uma vez que esta região tem menos contatos entre resíduos para alterar as compensações (ABRUSÁN & MARSH, 2016). O resíduo p.P35 está na *coil* N-terminal da proteína, mais especificamente, entre a região transmembranar e a hélice A. No estudo de XU *et al.* (2019), previu-se que a região entre as posições 32 e 35 há a formação de uma dobradiça para a orientação da proteína na membrana. A variante p.P35L troca um resíduo polar que mantém a ligação peptídica (CO-NH) numa conformação rígida por um resíduo hidrofóbico que interage para formar estruturas secundárias. Portanto, sugerimos que a variante p.P35L provoca uma alteração estérica que compromete a correta orientação proteína-membrana, acarretando numa parcial redução da atividade da *CYP21A2* concomitante com diminuição da velocidade da reação, conforme indicado pela constante cinética V_{max} obtida. Da mesma forma, a conhecida mutação NC p.P31L foi previamente descrita reduzir a atividade enzimática e a velocidade de reação (TUSIE-LUNA *et al.*, 1991; WANG *et al.*,

2017). Quanto à variante p.R484L, ela está situada na região C-terminal, dentro do característico cluster hidrofóbico das CYPs. Em WANG *et al.*, (2017), o resíduo p.R484 foi previsto para ter uma conexão de ponte salina com p.D323. A substituição do resíduo longo, flexível, carregado positivamente e fortemente ligado a uma molécula negativa dentro de uma região hidrofóbica por um resíduo curto que tende a interagir com o dobramento afetou gravemente a atividade da enzima CYP21A2 como observado no estudo funcional. Nossa hipótese é que p.R484L cause a perda da organização sólida hidrofóbica, alterando a estabilidade e expressão da proteína, como observamos no ensaio de *western blot*. Os indivíduos portadores de p.P35L (COELI *et al.*, 2010) e p.R484L (SILVEIRA *et al.*, 2009) têm um genótipo complexo com uma mutação grave em *cis*, portanto, qualquer patogenicidade inferida pelo fenótipo era inviável.

O resíduo leucina 199 não é conservado entre as CYPs de mamíferos. Esse resíduo está localizado na hélice F, perto do sítio ativo. A troca da leucina, um resíduo mantedor da estrutura secundária da hélice, pela prolina, resíduo com um anel no N-terminal que enrijece a ligação peptídica, possivelmente perturba a conformação da hélice e causa a quebra da estrutura secundária local. A substituição de uma leucina pela prolina nas posições 168 (hélice E), 262 (hélice H) e 322 (hélice J) também apresentou quebra da estrutura da hélice em (LOKE *et al.*, 2001; TARDY *et al.*, 2010; HAIDER *et al.*, 2013). Portanto, hipotetizamos que a variante p.L199P quebra parcialmente a hélice F, perturbando a ligação da proteína ao esteroide. Os dados funcionais corroborarão com esse dado, pois foi observada a diminuição afinidade da progesterona com a CYP21A2 mutada em uma razão de 6 vezes menor em relação à enzima selvagem. Esta variante foi identificada em um

paciente heterozigoto (SILVEIRA *et al.*, 2009), portanto, não foi possível uma correlação fenotípica.

O resíduo p.W202 está localizado em uma conformação de *turn* entre as hélices F e F'. Em CRUZ *et al.* (2020), p.W202 com p.R234, p.D288 e p.S109 foram descritos como resíduos hidrofóbicos notáveis no sítio catalítico da CYP21A2, com predomínio da carga positiva. Levando em consideração que a variante p.W202R substitui um grande resíduo por um grupo aromático rígido para uma cadeia lateral longa e flexível positivamente carregada perto do grupamento heme, esperaríamos que essa variante resultasse em uma perda completa da atividade da CYP21A2. No entanto, uma atividade residual permaneceu. Portanto, a carga positiva do resíduo substituído possivelmente causou um distúrbio de ligação da proteína ao grupamento heme, provocando com uma ligeira desestabilização identificada pelo cálculo estrutural, mas que não foi o suficiente para abolir a atividade da enzima. Curiosamente, esse resíduo não é conservado entre os CYPs de mamíferos. Esta variante, p.W202R, foi identificada em heterozigose composta com uma grande conversão gênica em um paciente com perda de sal renal (CARVALHO *et al.*, 2012). A principal mutação no grupo B, p.I173N, em homozigose está associada ao fenótipo de HAC-VS na maioria dos casos (NEW *et al.*, 2013; CARVALHO *et al.*, 2016; RIEDL *et al.*, 2019). Entretanto, em genótipos heterozigotos compostos como mutação leve, o fenótipo é variável. Em um grande estudo com uma população heterogênea (NEW *et al.*, 2013), a mutação p.I173N em heterozigose composta com uma mutação do grupo nulo foi associada ao fenótipo HAC-PS em 24,8% dos casos (n = 27) e ao fenótipo HAC-VS em 75% (n = 81). Por outro lado, em estudos de populações específicas esse dado pode apresentar uma variabilidade maior que

descrita, podendo esse genótipo ser encontrado com o fenótipo HAC-VS em 100% (n = 18) dos casos (CARVALHO *et al.*, 2016) ou com o fenótipo HAC-PS em 63,2% (n = 36) dos casos (RIEDL *et al.*, 2019).

A variante p.E352V também mostrou uma variabilidade fenotípica em pacientes heterozigotos compostos no estudo de CARVALHO *et al.* (2016). Um paciente com p.E352V em heterozigose composta com o Cluster E6 (grupo nulo) com o fenótipo HAC-VS e dois pacientes com IVS2-13A / C> G (grupo A) no alelo oposto com o fenótipo HAC-PS. Além desses, um caso de HAC-SV com p.G425S (grupo B) no alelo oposto. O resíduo p.E352 está localizado na hélice K e é parte da tríade ERR, juntamente com p.R355 e p.R409 (E352-R355-R409). A tríade ERR é um motivo em CYPs bem conhecido e altamente conservado que auxilia na estabilização da estrutura terciária local, travando o resíduo crucial p.C429 na posição correta para a ligação da proteína CYP21A2 com o grupamento heme (HAIDER *et al.*, 2013; WANG *et al.*, 2017). Mutações nos resíduos de arginina estão associadas à perda da atividade de hidroxilação, conforme observado com p.R355H, p.R355C e p.R409C (NUNEZ *et al.*, 1999; KRONE *et al.*, 2000; SOARDI *et al.*, 2008). No entanto, estudos com o p.E352 mutado mostraram uma atividade residual da enzima. A enzima com a mutação natural p.E352L tem uma atividade de 1,2% e 1,1% em relação à proteína selvagem para progesterona e 17OHP, respectivamente (KRONE *et al.*, 2000). Além disso, duas mutações artificiais neste resíduo (isoleucina e ácido aspártico) também mostraram uma atividade residual da enzima CYP21A2, mesmo o ácido aspártico sendo o aminoácido mais próximo do ácido glutâmico (KRONE *et al.*, 2005). A variante que estudamos, p.E352V, troca um resíduo longo, ligeiramente flexível e fortemente negativo carregado por um

resíduo curto e sem carga, causando uma ligeira desestabilização da estrutura da CYP21A2. Curiosamente, como observado para as outras variantes na mesma posição, nosso estudo funcional para p.E352V também mostrou atividade residual de CYP21A2 associada à mutação do grupo B (KRONE *et al.*, 2005). No entanto, a troca em 352 de ácido glutâmico por valina mostrou uma diminuição da expressão da proteína, o que não ocorreu em KRONE *et al.* (2005) com lisina, isoleucina e ácido aspártico.

Por último, o resíduo p.P433 está localizado na hélice L, adjacente a uma conformação *turn* (entre as hélices L e M) que contém dois resíduos críticos de ligação ao grupamento heme, a p.C429 e p.R427. A substituição de um resíduo polar que fixa a orientação da volta por um resíduo hidrofóbico afeta o dobramento local. A análise computacional mostrou uma desestabilização estrutural suave. Na análise de realizada por HAIDER *et al.* (2013), considerou-se que a variante p.P433L era capaz de tornar a estrutura mais flexível, evitando que p.C429 fosse apresentado ao grupamento heme. Os estudos funcionais mostraram uma atividade residual da proteína mutada combinada à redução de 6 vezes na velocidade de reação, conforme indicado pela constante cinética aparente V_{max} . Portanto, concluímos que a desestabilização da proteína devido à variante p.P433L faz com que a conformação *turn*, entre as hélices L e M, interrompa a orientação natural dos resíduos de ligação ao grupamento heme p.C429 e p.R427, afetando a eficiência de hidroxilação da enzima. O fenótipo HAC-NC do paciente portador desta variante em *trans* com a bem conhecida mutação NC p.P454S (CARVALHO *et al.*, 2012) corrobora com o impacto observado da variante p.P433L na atividade da CYP21A2.

A análise *in silico* através da análise estrutural mostrou ser uma ferramenta útil para prever os efeitos das mutações em enzimas metabolizadoras de esteroides. Nossos achados corroboraram com a hipótese cujas mutações *missense* que alteram a estrutura secundária ou terciária tem alto potencial de afetar a atividade da proteína (HAIDER *et al.*, 2013; ABRUSÁN & MARSH, 2016).

Além disso, cinco preditores online foram utilizados para prever o dano dessas variantes com efeito desconhecido, Meta-SNP, MutPred2, PredictSNP, PolyPhen-2 e SNAP2. As variantes p.L199P, p.E352V e p.R484L foram corretamente preditas pelos cinco programas. Por outro lado, a variante p.P35L foi apontada como neutra pelo programa PredictSNP, a variante p.W202R pelos programas Meta-SNP e PredictSNP e a variante p.P433L pelos programas Meta-SNP, PolyPhen-2 e PredictSNP. Nossos resultados enfatizaram o uso de diferentes abordagens de preditores de dano para realizar uma análise inicial do efeito de variantes do gene *CYP21A2*, principalmente quando o ensaio funcional não pode ser realizado para a confirmação do dado.

Sendo assim, com o intuito de determinamos os programas de predição *online* que apresentam os melhores desempenhos para discriminar variantes patogênicas e neutras no gene *CYP21A2*, nós selecionamos programas de predição *in silico*. Dezesete programas foram testados, sendo deste treze ferramentas individuais e quatro meta-preditores. Todos os programas selecionados são gratuitos, além de serem descritos na literatura com bom desempenho e/ou populares.

O desempenho dos programas de predição foi testado apenas com variantes *missense* do gene *CYP21A2* caracterizadas funcionalmente afim de se obter um resultado mais próximo do real. No total, nosso banco de dados foi composto por 103 variantes *missense*, sendo 90 patogênicas e 13 neutras. Entre as variantes patogênicas, o grupo de variantes clássico (CL) apresentou uma atividade enzimática residual entre 1,32% e 1,52% em relação a selvagem e o grupo NC entre 37,41% e 42,94%. Já o grupo de variantes neutras obteve uma média de atividade entre 100% e 94% em relação à enzima selvagem. As atividades enzimáticas mais baixas foram observadas na conversão de progesterona em DOC nos três grupos.

Todos os programas de predição online testados foram capazes de identificar variantes *missense* patogênicas, sendo que as variantes do grupo CL obtiveram uma maior taxa de acertos do que as do grupo NC. Esse resultado já era esperado, uma vez que o grupo CL reúne as variantes que causam o maior dano à funcionalidade da enzima. Por outro lado, o programa PANTHER-PSEP (TANG & THOMAS, 2016) foi o único preditor que não foi capaz de identificar variantes neutras. Sendo assim, o PANTHER-PSEP não deve ser utilizado para análise de variantes *missense* do gene *CYP21A2*.

Uma vez que número amostral restrito não permitiu uma distribuição entre grupos balanceada, os dados estatísticos considerados prioritariamente para a avaliação dos preditores foram a acurácia e o *Matthews correlation coefficient* (MCC). A acurácia e o MCC são descritos como os testes mais confiáveis para dados não balanceados em estudos de bioinformática, pois consideram o total de

amostras - verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN) (VIHINEN, 2012; CHICCO *et al.*, 2021). Quanto ao cálculo de sensibilidade e especificidade, que consideram metade da informação dos dados, nós consideramos o equilíbrio entre esses valores para as análises, pois ambos sozinhas não representam um desempenho real quando os grupos amostrais não são balanceados. Por fim, apesar dos valores preditivos positivos (VPP) e negativos (VPN) terem sido apresentados, ambos são altamente sensíveis ao desequilíbrio de dados, portanto, não foram considerados para a avaliação de desempenho (VIHINEN, 2012).

Em relação ao desempenho dos meta-preditores, era esperado obter esses programas entre os com melhor desempenho. Os meta-preditores acoplam diferentes fontes de banco de dados e os resultados de preditores simples para gerar o seu próprio. Entretanto, em comparação com os preditores simples, os meta-preditores apresentaram um desempenho intermediário. Uma possível hipótese é que devido ao desempenho de cada preditor simples variar para cada tipo de proteína, os pesos dados a eles nos meta-preditores podem não corresponder com o ideal. Além disso, o número de ferramentas não teve relação com a eficiência de predição de dano. Os programas Meta-SNP com quatro ferramentas (CAPRIOTTI *et al.*, 2013) e PredictSNP com seis (BENDL *et al.*, 2014) obtiveram melhor desempenho do que PredictSNP2 com seis ferramentas (BENDL *et al.*, 2016) e S3Ds&GO com três (CAPRIOTTI & ALTMAN, 2015). Em comparação com os testes realizados pelos desenvolvedores desses quatro meta-preditores, nós obtivemos uma acurácia semelhante das encontradas para o Meta-SNP e o PredictSNP (CAPRIOTTI *et al.*, 2013; BENDL *et al.*, 2014). Ambos compartilham

três preditores simples (PhD-SNP, SNAP e SIFT). Entretanto, os desempenhos do PredictSNP2 e S3Ds&GO foram inferiores aos observados pelos seus desenvolvedores (BENDL *et al.*, 2016; CAPRIOTTI & ALTMAN, 2015).

Os preditores simples com melhor desempenho para caracterizar a patogenicidade das variantes *missense* do gene *CYP21A2* foram: CADD (VAN DER VELDE *et al.*, 2017), ConSurf (ASHKENAZY *et al.*, 2016), DANN (QUANG *et al.*, 2015) e PolyPhen-2 (ADZHUBEI *et al.*, 2013). O programa CADD obteve a melhor acurácia e MCC para categorizar variantes patogênicas e neutras no geral (Ac 0,94 e MCC 0,75) e inclusive dentro dos grupos CL (Ac 0,97 e MCC 0,9) e NC (Ac 0,88 e MCC 0,71). A especificidade (0,85) e a sensibilidade (0,9) também obtiveram um bom equilíbrio entre seus valores. Curiosamente, esses valores da acurácia e a especificidade obtidas para a *CYP21A2* foram maiores do que os relatados pelos desenvolvedores do programa que utilizaram para o teste o banco de dados ClinVar (Ac 0,85; Ep 0,57) (VAN DER VELDE *et al.*, 2017). Os programas ConSurf, DANN e PolyPhen-2 mostraram um desempenho semelhante e o segundo melhor resultados de desempenho de acordo com a acurácia geral de 0,86-0,89 (CL 0,92 a 0,95; NC 0,77 a 0,81) e o MCC geral de 0,56 a 0,58 (CL 0,75 a 0,83; NC 0,51 a 0,56). Quanto à sensibilidade e especificidade do ConSurf e PolyPhen2, ambos apresentaram valores equilibrados. Por outro lado, DANN obteve 1,3 vezes menos especificidade do que sensibilidade. O artigo original do DANN utilizou o banco de dados ClinVar para o teste de desempenho do programa e apresentou uma área sob a curva (AUC) ROC de 0,95 (QUANG *et al.*, 2015). Já o PolyPhen-2 mostrou um melhor equilíbrio entre sensibilidade e especificidade para as variantes do gene *CYP21A2* do que os valores apresentados no artigo

original (ADZHUBEI *et al.*, 2013). O artigo que testou o PolyPhen-2 utilizou um banco de dados composto por mutações específicas dos genes *BRCA1*, *MSH2*, *MLH1* e *TP52* (ADZHUBEI *et al.*, 2013). A sensibilidade do PolyPhen-2 de 0,85 foi semelhante ao artigo original, mas a especificidade de 0.60 foi menor. O ConSurf não pode ser comparado com dados de outros artigos, uma vez que não há um modelo padrão recomendado para a escolha das sequências de alinhamento, o banco de dados e os algoritmos, como mencionado pelos desenvolvedores do programa em (ASHKENAZY *et al.*, 2016).

Dentre as 103 variantes utilizadas para o teste dos programas de predição, o CADD não acertou a predição de dano de seis variantes (4 FN e 2 FP), o ConSurf 12 variantes (11 FN e 1 FP), o DANN 11 variantes (7 FN e 4 FP) e o PolyPhen-2 14 variantes (11 FN e 3 FP). No entanto, apenas quatro variantes foram erroneamente preditas em comum por esses quatro programas, uma neutra (p.L13M) e três patogênicas (p.P106L, p.R225W e p.M474I). A variante p.P106L foi categorizada corretamente apenas pelos programas SNAP2 (HECHT *et al.*, 2015) e PANTHER-PSEP (TANG & THOMAS, 2016), enquanto as outras três variantes foram pelo PROVEAN (CHOI *et al.*, 2012). O programa de predição PROVEAN obteve um desempenho intermediário (Ac 0,77 e MCC 0,51) em relação aos outros preditores, com uma variação entre a especificidade e sensibilidade de 1.4 vezes. Comparando com os dados obtidos pelos desenvolvedores utilizando variantes do banco de dados UniProt, o desempenho analisado pelo equilíbrio entre sensibilidade e especificidade foi menor com o banco de variantes do nosso estudo, sendo a sensibilidade semelhante e a especificidade mais alta (CHOI *et al.*, 2012). O PROVEAN utiliza as informações de conservação dos resíduos vizinhos

adicionalmente ao resíduo de interesse. Para enzimas, o grau de conservação de resíduos vizinhos pode ser uma armadilha, uma vez que estas apresentam resíduos com alto nível de conservação e responsáveis por ligações críticas no meio de resíduos mais variáveis. Essa variabilidade pode ser vista para nas variantes p.P35L e p.R484L caracterizadas funcionalmente no presente estudo.

Entre as características das variantes utilizadas pelos programas de predição, o dado evolutivo do resíduo de interesse está presente na grande maioria destes. Uma das razões para tal, é que resíduos conservados ao longo do tempo podem indicar pontos críticos para a função da proteína. Dentre os programas avaliados, quatro utilizam apenas desse recurso para realizar os cálculos de predição FATHMM (SHIHAB *et al.*, 2013), PhD-SNP⁹ (CAPRIOTTI & FARISELLI, 2017), PROVEAN (CHOI *et al.*, 2012) e SIFT (VASER *et al.*, 2016). Esses preditores apresentaram desempenhos semelhantes, no entanto, SIFT obteve o balanço entre sensibilidade e especificidade mais equilibrado (Se 0,78 e Ep 0.85), semelhante ao obtido por (VASER *et al.*, 2016). No estudo realizado com variantes dos genes *HSD17B3*, *NR5A1*, *AR* e *LHCGR* (MONTENEGRO *et al.*, 2021), os programas SIFT e PROVEAN foram utilizados e obtiveram também o mesmo desempenho entre eles. No estudo (PSHENNIKOVA *et al.*, 2019), SIFT e PROVEAN mostraram o melhor desempenho entre os nove programas testados para os genes *GJB2*, *GJB3* e *GJB6*. Por outro lado, o programa FATHMM apresentou uma alta taxa de falso negativo em nosso estudo, sendo 30% das variantes patogênicas categorizadas como neutras, semelhantemente aos estudos (PSHENNIKOVA *et al.*, 2019; VASER *et al.*, 2016) cujas acurácia e MCC obtidas foram abaixo da média.

A segunda característica predominante nos preditores selecionados foi a avaliação de alterações estruturais. Variantes *missense* que modificam a estrutura secundária e/ou terciária da proteína têm mais chance de causarem dano na sua funcionalidade, segundo (ASHKENAZY *et al.*, 2016). Entre os programas selecionados, ConSurf (ASHKENAZY *et al.*, 2016), MutPred2 (PEJAVER *et al.*, 2020), PolyPhen-2 (ADZHUBEI *et al.*, 2013) e SNAP2 (HECHT *et al.*, 2015) fazem o uso dessa característica. Adicionalmente, alguns deles incluem outras fontes de dados, como é o caso do ConSurf que inclui relações filogenéticas (ASHKENAZY *et al.*, 2016), do Mutpre2 que inclui propriedades funcionais (PEJAVER *et al.*, 2020) e do SNAP2 (HECHT *et al.*, 2015) que faz o uso de uma matriz de probabilidades de efeito com um método de rede neural. Por outro lado, o programa PolyPhen-2 tem duas opções de banco de dados com os quais foi treinado separadamente, o HumVar e o HumDir (ADZHUBEI *et al.*, 2013). O PolyPhen-2 treinado com o HumVar é sugerido para o diagnóstico de doenças Mendelianas que estão relacionadas a variantes de efeitos drásticos. Já o PolyPhen-2 treinado com o HumDir é indicado para prever o efeito de alelos raros em *locus* potencialmente incluídos em um fenótipo complexo (ADZHUBEI *et al.*, 2013). No entanto, já foi mostrado que o PolyPhen-2, bem como, o SIFT tem baixa dependência do banco de dados para o alinhamento de sequências, apresentando bom desempenho inclusive com outros bancos de dados (HICKS *et al.*, 2011). No nosso estudo, o programa PolyPhen-2 apresentou um bom desempenho de predição (ac 0,88 e MCC 0,64), com a mesma sensibilidade e melhor equilíbrio sensibilidade-especificidade do que o relatado em (HICKS *et al.*, 2011). Semelhante a esse desempenho obtivemos com o programa ConSurf (Ac 0,88 e MCC = 0,58). Por fim,

o MutPred2 e o SNAP2 apresentaram um desempenho razoável, porém, com valores desbalanceados entre a sensibilidade-especificidade, sendo estes de 1,4 vezes com o MutPred2 e 1,5 vezes com o SNAP2. Apesar disso, o desempenho obtido com as variantes do gene *CYP21A2* foi melhor do que o descrito para o MutPred2 com os bancos de dados ClinVar e UniProt em (PEJAVER *et al.*, 2020) e para o SNAP2 com um banco composto por mais de 9.500 variantes em genes humanos em (HECHT *et al.*, 2015).

Capítulo VI: Conclusão

No presente trabalho, utilizamos programas de predição *in silico* para selecionar variantes *missense* de efeito desconhecido e com potencial de causarem dano à CYP21A2 encontradas na população brasileira ou portuguesa. Ademais, determinamos como as seis variantes selecionadas afetaram a atividade da enzima CYP21A2 por meio da modificação estrutural e funcional. Todas as variantes caracterizadas diminuíram a atividade da enzima CYP21A2 e três delas (p.L199P, p.W202R e p.E352V) modificaram também o nível de expressão proteico. As variantes p.P35L e p.L199P foram categorizadas como mutações do grupo C, relacionadas ao fenótipo HAC-NC. Já as variantes p.W202R, p.E352V, p.P433L e p.R484L apresentaram uma atividade residual compatível ao grupo de mutações B e, conseqüentemente, ao fenótipo HAC-VS. Além disso, com a meta-análise para avaliar o desempenho de dezessete programas de predição foi possível identificarmos quatro deles com os melhores desempenhos para variantes da proteína CYP21A2, CADD, ConSurf, DANN e PolyPhen-2. O número de ferramentas empregadas nesses programas não mostrou relação com a eficiência destes. Por fim, os resultados apresentados nestes estudos poderão ser aplicados no futuro para determinar a gravidade do genótipo de pacientes portadores da deficiência na enzima CYP21A2.

Capítulo VII – Referências Bibliográficas

ABRUSÁN, G; & MARSH, J. A. Alpha Helices Are More Robust to Mutations than Beta Strands. *PLoS computational biology*, 12(12):e1005242, 2016.

ADZHUBEI, I; JORDAN, D. M; & SUNYAEV, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics*, Cap. 7, p.20, 2013.

AHMED, A; SCHMIDT, C; & BRUNNER, T. Extra-Adrenal Glucocorticoid Synthesis in the Intestinal Mucosa: Between Immune Homeostasis and Immune Escape. *Frontiers in immunology*, 10, 1438, 2019.

ALVES C. A. D; CARGNIN K. R. N, SILVA C. C. C; PAULA L. C. P; NASCIMENTO M. L; BARBOSA M. E; LIBERATORE-JR L. D. R; PINTO R. M; ARRAIS R. F; MIRANDA M. C; BACHEGA T. S; KOPACEK C. Hiperplasia adrenal congênita: triagem neonatal - Guia Prático de Atualização 2019-2021. *Departamento Científico de Endocrinologia - Sociedade Brasileira de Endocrinologia*, 7:1-16, 2019. Disponível em <https://www.sbp.com.br/fileadmin/user_upload/22106c-GPA_-_Hiperplasia_Adrenal_Congenita-TriagemNeonatal.pdf>

ARAÚJO J. P. B; GOLDBECK A. S; SANTOS E. P; SILVA I. N; OLIVEIRA M. F; NASCIMENTO M. L; MEIRELLES R. M. R; BACHEGA T. S; CARVALHO T. M; *et al.* Triagem neonatal: hiperplasia adrenal congênita. *Ministério da Saúde (MS – OS 2015/0241)*, 1:1-44,2015.

ARAUJO, R. S; BILLERBECK, A. E; MADUREIRA, G; MENDONCA, B. B; & BACHEGA, T. A. Substitutions in the CYP21A2 promoter explain the simple-virilizing form of 21-hydroxylase deficiency in patients harbouring a P30L mutation. *Clinical endocrinology*, 62(2), 132–136, 2005.

ASHKENAZY, H; ABADI, S; MARTZ, E; CHAY, O; MAYROSE, I; PUPKO, T; & BEN-TAL, N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research*, 44(W1), W344–W350, 2016.

BARBARO, M; SOARDI, F. C; ÖSTBERG, L. J; PERSSON, B; DE MELLO, M. P; WEDELL, A; & LAJIC, S. In vitro functional studies of rare CYP21A2 mutations and establishment of an activity gradient for nonclassic mutations improve phenotype predictions in congenital adrenal hyperplasia. *Clinical endocrinology*, 82(1), 37–44, 2015.

BARONIO, F; ORTOLANO, R; MENABÒ, S; CASSIO, A; BALDAZZI, L; DI NATALE, V; TONTI, G; VESTRUCCI, B; & BALSAMO, A. 46,XX DSD due to Androgen Excess in Monogenic Disorders of Steroidogenesis: Genetic, Biochemical, and Clinical Features. *International journal of molecular sciences*, 20(18), 4605, 2019.

BAUMGARTNER-PARZER, S; WITSCH-BAUMGARTNER, M; & HOEPPNER, W. EMQN best practice guidelines for molecular genetic testing and reporting of 21-hydroxylase deficiency. *European journal of human genetics : EJHG*, 28(10), 1341–1367, 2020.

BECHMANN, N; BERGER, I; BORNSTEIN, S. R; & STEENBLOCK, C. Adrenal medulla development and medullary-cortical interactions. *Molecular and cellular endocrinology*, 528, 111258, 2021.

BENDL, J; MUSIL, M; ŠTOURAC, J; ZENDULKA, J; DAMBORSKÝ, J; & BREZOVSKÝ, J. PredictSNP2: A Unified Platform for Accurately Evaluating SNP Effects by Exploiting the Different Characteristics of Variants in Distinct Genomic Regions. *PLoS computational biology*, 12(5), e1004962, 2016.

BENDL, J; STOURAC, J; SALANDA, O; PAVELKA, A; WIEBEN, E. D; ZENDULKA, J; BREZOVSKY, J; & DAMBORSKY, J. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS computational biology*, 10(1), e1003440, 2014.

BRUQUE, C. D; DELEA, M; FERNÁNDEZ, C. S; ORZA, J. V; TABOAS, M; BUZZALINO, N; ESPECHE, L. D; SOLARI, A; LUCCERINI, V; ALBA, L; NADRA, A. D; & DAIN, L. Structure-based activity prediction of CYP21A2 stability variants: A survey of available gene variations. *Scientific reports*, 6, 39082, 2016.

CARVALHO B, PEREIRA M, MARQUES CJ, CARVALHO D, LEÃO M, OLIVEIRA JP, BARROS A, CARVALHO F. Comprehensive genetic analysis and structural characterization of CYP21A2 mutations in CAH patients. *Exp Clin Endocrinol Diabetes*. 2012 Oct;120(9):535-9. doi: 10.1055/s-0032-1323805.

CARVALHO, B; PEREIRA, M; MARQUES, C. J; CARVALHO, D; LEÃO, M; OLIVEIRA, J. P; BARROS, A; & CARVALHO, F. Comprehensive genetic analysis and structural characterization of CYP21A2 mutations in CAH patients. *Experimental and clinical endocrinology & diabetes: official journal, German Society of Endocrinology [and] German Diabetes Association*, 120(9), 535–53, 2012.

CAPRIOTTI, E; ALTMAN, R. B; & BROMBERG, Y. Collective judgment predicts disease-associated single nucleotide variants. *BMC genomics*, 14 Suppl 3(Suppl 3), S2, 2013.

CAPRIOTTI, E; & FARISELLI, P. PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants. *Nucleic acids research*, 45(W1), W247–W252, 2017.

CHICCO, D; TÖTSCH, N; & JURMAN, G. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData mining*, 14(1), 13, 2021.

CHOI, J. H; KIM, G. H; & YOO, H. W. Recent advances in biochemical and molecular analysis of congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Annals of pediatric endocrinology & metabolism*, 21(1), 1–6, 2016.

CHOI, Y; SIMS, G. E; MURPHY, S; MILLER, J. R; & CHAN, A. P. Predicting the functional effect of amino acid substitutions and indels. *PloS one*, 7(10), e46688, 2012.

COSTA-SANTOS, M; KATER, C. E; AUCHUS, R. J; & BRAZILIAN CONGENITAL ADRENAL HYPERPLASIA MULTICENTER STUDY GROUP. Two prevalent CYP17 mutations and genotype-phenotype correlations in 24 Brazilian patients with 17-hydroxylase deficiency. *The Journal of clinical endocrinology and metabolism*, 89(1), 49–60, 2004.

DE CARVALHO, D. F; MIRANDA, M. C; GOMES, L. G; MADUREIRA, G; MARCONDES, J. A; BILLERBECK, A. E; RODRIGUES, A. S; PRESTI, P. F; KUPERMAN, H; DAMIANI, D; MENDONÇA, B. B; & BACHEGA, T. A. Molecular CYP21A2 diagnosis in 480 Brazilian patients with congenital adrenal hyperplasia before newborn screening introduction. *European journal of endocrinology*, 175(2), 107–116, 2016.

EL-MAOUCHE, D; ARLT, W; & MERKE, D. P. Congenital adrenal hyperplasia. *Lancet (London, England)*, 390(10108), 2194–2210, 2017.

FLÜCK C.E. & PANDEY A.V. P450 Oxidoreductase Deficiency (PORD). In: Maria I. New, Oksana Lekarev, Alan Parsa, Tony T. Yuen, Bert W. O'Malley, Gary D. Hammer. Genetic Steroid Disorders. *Academic Press*, Cap. 3H, p. 125-143, 2013. ISBN: 9780124160064.

FLÜCK, C. E; PANDEY, A. V; HUANG, N; AGRAWAL, V; & MILLER, W. L. P450 oxidoreductase deficiency - a new form of congenital adrenal hyperplasia. *Endocrine development*, 13, 67–81, 2008.

GOMES, L. G; HUANG, N; AGRAWAL, V; MENDONÇA, B. B; BACHEGA, T. A; & MILLER, W. L. Extraadrenal 21-hydroxylation by CYP2C19 and CYP3A4: effect on 21-hydroxylase deficiency. *The Journal of clinical endocrinology and metabolism*, 94(1), 89–95, 2009.

GOTO, M; PIPER HANLEY, K; MARCOS, J; WOOD, P. J; WRIGHT, S; POSTLE, A. D; CAMERON, I. T; MASON, J. I; WILSON, D. I; & HANLEY, N. A. In humans, early cortisol biosynthesis provides a mechanism to safeguard female sexual development. *The Journal of clinical investigation*, 116(4), 953–960, 2006.

HAHNER, S; ROSS, R. J; ARLT, W; BANCOS, I; BURGER-STRITT, S; TORPY, D. J; HUSEBYE, E. S; & QUINKLER, M. Adrenal insufficiency. *Nature reviews. Disease primers*, 7(1), 19, 2021.

HAIDER, S; ISLAM, B; D'ATRI, V; SGOBBA, M; POOJARI, C; SUN, L; YUEN, T; ZAIDI, M; & NEW, M. I. Structure-phenotype correlations of human CYP21A2 mutations in congenital adrenal hyperplasia. *Proceedings of the National Academy of Sciences of the United States of America*, 110(7), 2605–2610, 2013.

HANNAH-SHMOUNI, F; CHEN, W; & MERKE, D. P. Genetics of Congenital Adrenal Hyperplasia. *Endocrinology and metabolism clinics of North America*, 46(2), 435–458, 2017.

HANNAH-SHMOUNI, F; MORISSETTE, R; SINAI, N; ELMAN, M; PREZANT, T. R; CHEN, W; PULVER, A; & MERKE, D. P. Revisiting the prevalence of nonclassic congenital adrenal hyperplasia in US Ashkenazi Jews and Caucasians. *Genetics in medicine : official journal of the American College of Medical Genetics*, 19(11), 1276–1279, 2017.

HECHT, M; BROMBERG, Y; & ROST, B. Better prediction of functional effects for sequence variants. *BMC genomics*, 16 (Suppl 8), S1, 2015.

HIGASHI, Y; TANAE, A; INOUE, H; HIROMASA, T; & FUJII-KURIYAMA, Y. Aberrant splicing and missense mutations cause steroid 21-hydroxylase [P-450(C21)] deficiency in humans: possible gene conversion products. *Proceedings of the National Academy of Sciences of the United States of America*, 85(20), 7486–7490, 1988.

JORGE, J. C; ECHEVERRI, C; MEDINA, Y; & ACEVEDO, P. Male gender identity in an XX individual with congenital adrenal hyperplasia. *The journal of sexual medicine*, 5(1), 122–131, 2008.

KARAOĞLAN, M; NACARKAHYA, G; AYTAÇ, E. H; & KESKIN, M. Challenges of CYP21A2 genotyping in children with 21-hydroxylase deficiency: determination of genotype-phenotype correlation using next generation sequencing in Southeastern Anatolia. *Journal of endocrinological investigation*, 44(11), 2395–2405, 2021.

KAUPERT, L. C; LEMOS-MARINI, S. H; DE MELLO, M. P; MOREIRA, R. P; BRITO, V. N; JORGE, A. A; LONGUI, C. A; GUERRA, G; JR, MENDONCA, B. B; & BACHEGA, T. A. The effect of fetal androgen metabolism-related gene variants on external genitalia virilization in congenital adrenal hyperplasia. *Clinical genetics*, 84(5), 482–488, 2013.

KHAJURIA, R; WALIA, R; BHANSALI, A; & PRASAD, R. Functional characterization and molecular modeling of the mutations in CYP21A2 gene from patients with Congenital Adrenal Hyperplasia. *Biochimie*, 149, 115–121, 2018.

KOPACEK, C; DE CASTRO, S. M; PRADO, M. J; DA SILVA, C. M; BELTRÃO, L. A; & SPRITZER, P. M. (Neonatal screening for congenital adrenal hyperplasia in Southern Brazil: a population based study with 108,409 infants. *BMC pediatrics*, 17(1), 22, 2017.

KOPACEK, C; DE CASTRO, S. M; PRADO, M. J; DA SILVA, C. M; BELTRÃO, L. A; & SPRITZER, P. M. Neonatal screening for congenital adrenal hyperplasia in Southern Brazil: a population based study with 108,409 infants. *BMC pediatrics*, 17(1), 22, 2017.

KRONE, N; RIEPE, F. G; GRÖTZINGER, J; PARTSCH, C. J; BRÄMSWIG, J; & SIPPELL, W. G. The residue E351 is essential for the activity of human 21-hydroxylase: evidence from a naturally occurring novel point mutation compared with artificial mutants generated by single amino acid substitutions. *Journal of molecular medicine (Berlin, Germany)*, 83(7), 561–568, 2005.

KRONE, N; ROSE, I. T; WILLIS, D. S; HODSON, J; WILD, S. H; DOHERTY, E. J; HAHNER, S; PARAJES, S; STIMSON, R. H; HAN, T. S; CARROLL, P. V; CONWAY, G. S; WALKER, B. R; MACDONALD, F; ROSS, R. J; ARLT, W; & UNITED KINGDOM CONGENITAL ADRENAL HYPERPLASIA ADULT STUDY EXECUTIVE (CAHASE). Genotype-phenotype correlation in 153 adult patients with congenital adrenal hyperplasia due to 21-hydroxylase deficiency: analysis of the United Kingdom Congenital adrenal Hyperplasia Adult Study Executive (CaHASE) cohort. *The Journal of clinical endocrinology and metabolism*, 98(2), E346–E354, 2013.

LAO, Q; MALLAPPA, A; RUEDA FAUCZ, F; JOYAL, E; VEERARAGHAVAN, P; CHEN, W; & MERKE, D. P. A TNXB splice donor site variant as a cause of hypermobility type Ehlers-Danlos syndrome in patients with congenital adrenal hyperplasia. *Molecular genetics & genomic medicine*, 9(2), e1556, 2021.

LI, N; ZHANG, J; LIAO, D; YANG, L; WANG, Y; & HOU, S. Association between C4, C4A, and C4B copy number variations and susceptibility to autoimmune diseases: a meta-analysis. *Scientific reports*, 7, 42628, 2017

LIN-SU, K; HARBISON, M. D; LEKAREV, O; VOGIATZI, M. G; & NEW, M. I. Final adult height in children with congenital adrenal hyperplasia treated with growth hormone. *The Journal of clinical endocrinology and metabolism*, 96(6), 1710–1717, 2011.

LOKE, K. Y; LEE, Y. S; LEE, W. W; & POH, L. K. Molecular analysis of CYP-21 mutations for congenital adrenal hyperplasia in Singapore. *Hormone research*, 55(4), 179–184, 2001.

MENABÒ, S; POLAT, S; BALDAZZI, L; KULLE, A. E; HOLTERHUS, P. M; GRÖTZINGER, J; FANELLI, F; BALSAMO, A; & RIEPE, F. G. Congenital adrenal hyperplasia due to 11-beta-hydroxylase deficiency: functional consequences of four CYP11B1 mutations. *European journal of human genetics : EJHG*, 22(5), 610–616, 2014.

MERKE, D. P; & AUCHUS, R. J. Congenital Adrenal Hyperplasia Due to 21-Hydroxylase Deficiency. *The New England journal of medicine*, 383(13), 1248–1261, 2020.

MESIANO, S; & JAFFE, R. B. Developmental and functional biology of the primate fetal adrenal cortex. *Endocrine reviews*, 18(3), 378–403, 1997.

MILLER, W. L; & AUCHUS, R. J. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocrine reviews*, 32(1), 81–151, 2011.

MILLER, W. L; & MERKE, D. P. Tenascin-X, Congenital Adrenal Hyperplasia, and the CAH-X Syndrome. *Hormone research in paediatrics*, 89(5), 352–361, 2018.

MILLER W. L. Congenital Adrenal Hyperplasia: Time to Replace 17OHP with 21-Deoxycortisol. *Hormone research in paediatrics*, 91(6), 416–420, 2019.

MONTENEGRO, L. R; LERÁRIO, A. M; NISHI, M. Y; JORGE, A; & MENDONCA, B. B. Performance of mutation pathogenicity prediction tools on missense variants associated with 46,XY differences of sex development. *Clinics (Sao Paulo, Brazil)*, 76, e2052, 2021.

MOURA-MASSARI, V. O; CUNHA, F. S; GOMES, L. G; BUGANO DINIZ GOMES, D; MARCONDES, J. A; MADUREIRA, G; DE MENDONCA, B. B; & BACHEGA, T. A. The Presence of Clitoromegaly in the Nonclassical Form of 21-Hydroxylase Deficiency Could Be Partially Modulated by the CAG Polymorphic Tract of the Androgen Receptor Gene. *PLoS one*, 11(2), e0148548, 2016.

NARASIMHAN, M. L; & KHATTAB, A. Genetics of congenital adrenal hyperplasia and genotype-phenotype correlation. *Fertility and sterility*, 111(1), 24–29, 2019.

NEVES CRUZ, J; DA COSTA, K. S; DE CARVALHO, T; & DE ALENCAR, N. Measuring the structural impact of mutations on cytochrome P450 21A2, the major steroid 21-hydroxylase related to congenital adrenal hyperplasia. *Journal of biomolecular structure & dynamics*, 38(5), 1425–1434, 2020.

NEW, M. I; ABRAHAM, M; GONZALEZ, B; DUMIC, M; RAZZAGHY-AZAR, M; CHITAYAT, D; SUN, L; ZAIDI, M; WILSON, R. C; & YUEN, T. Genotype-phenotype correlation in 1,507 families with congenital adrenal hyperplasia owing to 21-hydroxylase deficiency. *Proceedings of the National Academy of Sciences of the United States of America*, 110(7), 2611–2616, 2013.

NEW MI, YAU M, LEKAREV O, LIN-SU K, PARSA A, PINA C, YUEN T, KHATTAB A. Congenital Adrenal Hyperplasia. [Updated 2017 Mar 15]. In: Feingold KR, Anawalt B, Boyce A, et al; editors. Endotext [Internet]. South Dartmouth (MA): MDText.com, Inc.; 2000-. Disponível em: <<https://www.ncbi.nlm.nih.gov/books/NBK278953/>>.

NUNEZ, B. S; LOBATO, M. N; WHITE, P. C; & MESEGUER, A. Functional analysis of four CYP21 mutations from spanish patients with congenital adrenal hyperplasia. *Biochemical and biophysical research communications*, 262(3), 635–637, 1999.

PALLAN, P. S; WANG, C; LEI, L; YOSHIMOTO, F. K; AUCHUS, R. J; WATERMAN, M. R; GUENGERICH, F. P; & EGLI, M. Human Cytochrome P450 21A2, the Major Steroid 21-Hydroxylase: structure of the enzyme-progesterone substrate complex

and rate-limiting c-h bond cleavage. *The Journal of biological chemistry*, 290(21), 13128–13143, 2015.

PANDEY, A. V; & FLÜCK, C. E. NADPH P450 oxidoreductase: structure, function, and pathology of diseases. *Pharmacology & therapeutics*, 138(2), 229–254, 2013.

PEJAVER, V; URRESTI, J; LUGO-MARTINEZ, J; PAGEL, K. A; LIN, G. N; NAM, H. J; MORT, M; COOPER, D. N; SEBAT, J; IAKOUCHEVA, L. M; MOONEY, S. D; & RADIVOJAC, P Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nature communications*, 11(1), 5918, 2020.

PEJAVER, V; URRESTI, J; LUGO-MARTINEZ, J; PAGEL, K. A; LIN, G. N; NAM, H. J; MORT, M; COOPER, D. N; SEBAT, J; IAKOUCHEVA, L. M; MOONEY, S. D; & RADIVOJAC, P. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nature communications*, 11(1), 5918, 2020.

PIGNATELLI, D; CARVALHO, B. L; PALMEIRO, A; BARROS, A; GUERREIRO, S. G; & MACUT, D. The Complexities in Genotyping of Congenital Adrenal Hyperplasia: 21-Hydroxylase Deficiency. *Frontiers in endocrinology*, 10, 432, 2019.

PRADER A. Der Genitalbefund beim Pseudohermaphroditismus femininus des kongenitalen adrenogenitalen Syndroms; Morphologie, Häufigkeit, Entwicklung und Vererbung der verschiedenen Genitalformen [Genital findings in the female pseudohermaphroditism of the congenital adrenogenital syndrome; morphology, frequency, development and heredity of the different genital forms]. *Helvetica paediatrica acta*, 9(3), 231–248, 1954.

PRADO, M. J; DE CASTRO, S. M; KOPACEK, C; DE MELLO, M. P; RISPOLI, T; GRANDI, T; DA SILVA, C; & ROSSETTI, M. Development of CYP21A2 Genotyping Assay for the Diagnosis of Congenital Adrenal Hyperplasia. *Molecular diagnosis & therapy*, 21(6), 663–675, 2017.

PSHENNIKOVA, V. G; BARASHKOV, N. A; ROMANOV, G. P; TERYUTIN, F. M; SOLOV'EV, A. V; GOTOVTSEV, N. N; NIKANOROVA, A. A; NAKHODKIN, S. S; SAZONOV, N. N; MOROZOV, I. V; BONDAR, A. A; DZHEMILEVA, L. U; KHUSNUTDINOVA, E. K; POSUKH, O. L; & FEDOROVA, S. A. Comparison of Predictive *In Silico* Tools on Missense Variants in *GJB2*, *GJB6*, and *GJB3* Genes Associated with Autosomal Recessive Deafness 1A (DFNB1A). *TheScientificWorldJournal*, 2019, 5198931, 2019

RIEDL, S; RÖHL, F. W; BONFIG, W; BRÄMSWIG, J; RICHTER-UNRUH, A; FRICKE-OTTO, S; BETTENDORF, M; RIEPE, F; KRIEGSHÄUSER, G; SCHÖNAU, E; EVEN, G; HAUFFA, B; DÖRR, H. G; HOLL, R. W; MOHNIKE, K; & AQUAPE CAH STUDY GROUP. Genotype/phenotype correlations in 538 congenital adrenal hyperplasia patients from Germany and Austria: discordances in milder genotypes and in screened versus prescreening patients. *Endocrine connections*, 8(2), 86–94, 2019.

RIEPE, F. G; HIORT, O; GRÖTZINGER, J; SIPPELL, W. G; KRONE, N; & HOLTERHUS, P. M. Functional and structural consequences of a novel point mutation in the CYP21A2 gene causing congenital adrenal hyperplasia: potential relevance of helix C for P450 oxidoreductase-21-hydroxylase interaction. *The Journal of clinical endocrinology and metabolism*, 93(7), 2891–2895, 2008.

RODRIGUES, N. R; DUNHAM, I; YU, C. Y; CARROLL, M. C; PORTER, R. R; & CAMPBELL, R. D. Molecular characterization of the HLA-linked steroid 21-hydroxylase B gene from an individual with congenital adrenal hyperplasia. *The EMBO journal*, 6(6), 1653–1661, 1987.

SANTOS-SILVA, R; CARDOSO, R; LOPES, L; FONSECA, M; ESPADA, F; SAMPAIO, L; BRANDÃO, C; ANTUNES, A; BRAGANÇA, G; COELHO, R; BERNARDO, T; VIEIRA, P; MORAIS, R; LEITE, A. L; RIBEIRO, L; *et al.* CYP21A2 Gene Pathogenic Variants: A Multicenter Study on Genotype-Phenotype Correlation from a Portuguese Pediatric Cohort. *Hormone research in paediatrics*, 91(1), 33–45, 2019.

SECCIA, T. M; CAROCCIA, B; GOMEZ-SANCHEZ, E. P; GOMEZ-SANCHEZ, C. E; & ROSSI, G. P. The Biology of Normal Zona Glomerulosa and Aldosterone-Producing Adenoma: Pathological Implications. *Endocrine reviews*, 39(6), 1029–1056, 2018.

SHIHAB, H. A; GOUGH, J; COOPER, D. N; STENSON, P. D; BARKER, G. L; EDWARDS, K. J; DAY, I. N; & GAUNT, T. R. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Human mutation*, 34(1), 57–65, 2013.

SILVEIRA, E. L; ELNECAVE, R. H; DOS SANTOS, E. P; MOURA, V; PINTO, E. M; VAN DER LINDEN NADER, I; MENDONÇA, B. B; & BACHEGA, T. A. Molecular analysis of CYP21A2 can optimize the follow-up of positive results in newborn screening for congenital adrenal hyperplasia. *Clinical genetics*, 76(6), 503–510, 2009.

SIMONETTI, L; BRUQUE, C. D; FERNÁNDEZ, C. S; BENAVIDES-MORI, B; DELEA, M; KOLOMENSKI, J. E; ESPECHE, L. D; BUZZALINO, N. D; NADRA, A. D; & DAIN, L. CYP21A2 mutation update: Comprehensive analysis of databases and published genetic variants. *Human mutation*, 39(1), 5–22, 2018.

SPEISER, P. W; DUPONT, J; ZHU, D; SERRAT, J; BUEGELEISEN, M; TUSIE-LUNA, M. T; LESSER, M; NEW, M. I; & WHITE, P. C. Disease expression and molecular genotype in congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *The Journal of clinical investigation*, 90(2), 584–595, 1992.

SOARDI, F. C; BARBARO, M; LAU, I. F; LEMOS-MARINI, S. H; BAPTISTA, M. T; GUERRA-JUNIOR, G; WEDELL, A; LAJIC, S; & DE MELLO, M. P. Inhibition of CYP21A2 enzyme activity caused by novel missense mutations identified in Brazilian and Scandinavian patients. *The Journal of clinical endocrinology and metabolism*, 93(6), 2416–2420, 2008.

TANG, H; & THOMAS, P. D. PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinformatics (Oxford, England)*, 32(14), 2230–2232, 2016.

TARDY, V; MENASSA, R; SULMONT, V; LIENHARDT-ROUSSIE, A; LECOINTRE, C; BRAUNER, R; DAVID, M; & MOREL, Y. Phenotype-genotype correlations of 13 rare CYP21A2 mutations detected in 46 patients affected with 21-hydroxylase deficiency and in one carrier. *The Journal of clinical endocrinology and metabolism*, 95(3), 1288–1300, 2010.

TAVES, M. D; GOMEZ-SANCHEZ, C. E; & SOMA, K. K. Extra-adrenal glucocorticoids and mineralocorticoids: evidence for local synthesis, regulation, and function. *American journal of physiology. Endocrinology and metabolism*, 301(1), E11–E24, 2011.

TUSIE-LUNA, M. T; SPEISER, P. W; DUMIC, M; NEW, M. I; & WHITE, P. C. A mutation (Pro-30 to Leu) in CYP21 represents a potential nonclassic steroid 21-hydroxylase deficiency allele. *Molecular endocrinology (Baltimore, Md.)*, 5(5), 685–692, 1991.

TUSIÉ-LUNA, M. T; & WHITE, P. C. Gene conversions and unequal crossovers between CYP21 (steroid 21-hydroxylase gene) and CYP21P involve different mechanisms. *Proceedings of the National Academy of Sciences of the United States of America*, 92(23), 10796–10800, 1995.

TUTUNCULER, F; SAKA, N; ARKAYA, S. C; ABBASOGLU, S; & BAS, F. Evaluation of adrenomedullary function in patients with congenital adrenal hyperplasia. *Hormone research*, 72(6), 331–336, 2009.

QUANG, D; CHEN, Y; & XIE, X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics (Oxford, England)*, 31(5), 761–763, 2015.

VAN DER VELDE, K. J; DE BOER, E. N; VAN DIEMEN, C. C; SIKKEMA-RADDATZ, B; ABBOTT, K. M; KNOPPERTS, A; FRANKE, L; SIJMONS, R. H; DE KONING, T. J; WIJMENGA, C; SINKE, R. J; & SWERTZ, M. A. (2017). GAVIN: Gene-Aware Variant INterpretation for medical sequencing. *Genome biology*, 18(1), 6, 2017.

VASER, R; ADUSUMALLI, S; LENG, S. N; SIKIC, M; & NG, P. C. SIFT missense predictions for genomes. *Nature protocols*, 11(1), 1–9, 2016.

VERMA, S; GREEN-GOLAN, L; VANRYZIN, C; DRINKARD, B; MEHTA, S. P; WEISE, M; EISENHOFER, G; & MERKE, D. P. Adrenomedullary function in patients with nonclassic congenital adrenal hyperplasia. *Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et métabolisme*, 42(8), 607–612, 2010.

VIHINEN M. How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis. *BMC genomics*, 13 (Suppl 4), S2, 2012.

WANG, C; PALLAN, P. S; ZHANG, W; LEI, L; YOSHIMOTO, F. K; WATERMAN, M. R; EGLI, M; & GUENGERICH, F. P. Functional analysis of human cytochrome P450 21A2 variants involved in congenital adrenal hyperplasia. *The Journal of biological chemistry*, 292(26), 10767–10778, 2017.

WEDELL A. Molecular genetics of 21-hydroxylase deficiency. *Endocrine development*, 20, 80–87, 2011.

WHITE, P. C; & SPEISER, P. W. Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Endocrine reviews*, 21(3), 245–291, 2000.

WILSON D, BORDONI B. Embryology, Mullerian Ducts (Paramesonephric Ducts) [Updated 2021 Jul 30] *Treasure Island (FL): StatPearls*, 2021. Disponível em: <<https://www.ncbi.nlm.nih.gov/books/NBK557727/>>.

XU, C; JIA, W; CHENG, X; YING, H; CHEN, J; XU, J; GUAN, Q; ZHOU, X; ZHENG, D; LI, G; & ZHAO, J. Genotype-phenotype correlation study and mutational and hormonal analysis in a Chinese cohort with 21-hydroxylase deficiency. *Molecular genetics & genomic medicine*, 7(6), e671, 2019.

YAMAZAKI, H; & SHIMADA, T. Progesterone and testosterone hydroxylation by cytochromes P450 2C19, 2C9, and 3A4 in human liver microsomes. *Archives of biochemistry and biophysics*, 346(1), 161–169, 1997.

YAU M, GUJRAL J, NEW MI. Congenital Adrenal Hyperplasia: Diagnosis and Emergency Treatment. [Updated 2019 Apr 16]. In: *Feingold KR, Anawalt B, Boyce A, et al; editors. Endotext [Internet]. South Dartmouth (MA): MDText.com, Inc.; 2000-*. Disponível em: <<https://www.ncbi.nlm.nih.gov/books/NBK279085/>>.

Zhou, Y; Ingelman-Sundberg, M; & Lauschke, V. M. (2017). Worldwide Distribution of Cytochrome P450 Alleles: A Meta-analysis of Population-scale Sequencing Projects. *Clinical pharmacology and therapeutics*, 102(4), 688–700, 2017.

Apêndice I – Certificado de aceite do artigo 1



Apêndice II – Submissão do artigo 2

The Journal of Clinical Endocrinology & Metabolism // A manuscript number has been assigned to Meta-analysis of variant predictions in congenital adrenal hyperplasia caused by mutations in CYP21A2 (jc.2021-03306)

JCEM <em@editorialmanager.com>

Sat 12/25/2021 17:41

To: Mayara Jorgens Prado <mayjorgens@hotmail.com>

You are being carbon copied ("cc:d") on an e-mail "To" "Amit V Pandey"

amit@pandeylab.org;amit.pandey@dbmr.unibe.ch

CC: "Mayara Jorgens Prado" mayjorgens@hotmail.com, "Rodrigo Ligabue-Braun"

ligabue.braun@gmail.com, "Arnaldo Zaha" zaha.arnaldo@gmail.com, "Maria Lucia Rosa Rossetti"

mrossett@terra.com.br

JCEM THE JOURNAL OF CLINICAL
ENDOCRINOLOGY & METABOLISM

12/25/2021

Re: Meta-analysis of variant predictions in congenital adrenal hyperplasia caused by mutations in CYP21A2

Dear Dr. Pandey,

Thank you for sending your paper, Meta-analysis of variant predictions in congenital adrenal hyperplasia caused by mutations in CYP21A2, to *The Journal of Clinical Endocrinology & Metabolism*. We will assign your manuscript to an Associate Editor, who will decide within three business days whether to send it out for peer review. After that, you can track the status of your manuscript by logging into your author portal at the following link: <https://www.editorialmanager.com/jclinendometab/>

Your manuscript has been assigned the following number: jc.2021-03306, which should be used in any future communications with the editors or editorial office.

Please note that Supplemental Data is no longer allowed as uploads submitted with a manuscript. Supplemental data must instead be submitted to a repository and cited in the manuscript bibliography (or incorporated into the paper body). Please note that if a revision of your manuscript is requested, you will be required to deposit such information, if not in the paper body, into a data repository and cite it appropriately in the text. For more information, please see [our author guidelines](#).

If you have any questions or concerns, please contact Tim Beardsley, Executive Editor, at tbeardsley@endocrine.org.

We greatly appreciate the opportunity to review your work.

Sincerely,

Paul M. Stewart, MD FRCP FMedSci

Editor-in-Chief, *The Journal of Clinical Endocrinology and Metabolism*

Apêndice III – Colaboração em artigo

Molecular Biology Reports
<https://doi.org/10.1007/s11033-020-05677-z>

SHORT COMMUNICATION



Novel frameshift variant of the *CFTR* gene: S511Lfs*2 from phenotype to molecular predictions

Thaiane Rispoli^{1,2} · Grazielle Motta Rodrigues^{2,3} · Mayara Jorgens Prado^{1,2} · Leonardo Araújo Pinto⁴ · Marcelo Tadday Rodrigues⁵ · Cynthia Rocha Dullius⁵ · Tarciana Grandi² · Cláudia Maria Dornelles da Silva² · José Eduardo Vargas⁶ · Maurício Menegatti Rigo⁷ · Maria Lucia Rossetti^{1,8}

Received: 27 March 2020 / Revised: 9 July 2020 / Accepted: 19 July 2020
 © Springer Nature B.V. 2020

Abstract

Cystic fibrosis (CF) is a genetic disease caused by variants in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene. There are over 2,000 different pathogenic and non-pathogenic variants described in association with a broad clinical heterogeneity. In this work, we identified a novel variant S511Lfs*2 in *CFTR* gene that has not been reported in patients with CF. The patient was a female genotyped with c.1000C>T (legacy name: R334W) variant (pathogenic, CF-causing) and the novel variant (S511Lfs*2). We verified the amino acid sequence, the protein structure, and predicted the pathogenicity employing computational analysis. Our findings showed that S511Lfs*2 is a frameshift variant and suggest that it is associated with severe CF phenotype, as it leads to a lack of *CFTR* protein synthesis, and consequently the loss of its functional activity.

Keywords Cystic fibrosis · S511Lfs*2 variant · *CFTR* gene · *CFTR* protein · Frameshift variant

Introduction

Cystic fibrosis (CF-OMIM 219700) is an autosomal recessive inherited disorder caused by variants in the Cystic Fibrosis Transmembrane Conductance Regulator (*CFTR*) gene [1]. The *CFTR* gene encodes a protein of 1480 residues in length (UniProt ID: P13569). This protein belongs to the ATP-binding cassette (ABC) transporter superfamily and its function is related to the transport of chloride ions across cell membranes [2, 3], regulation of airway fluid homeostasis [4–6], and defense against pathogens by the regulation of pH and ion content in the airway surface fluid layer [4, 6, 7].

The three-dimensional structure of *CFTR* protein is composed by five domains. Two transmembrane domains (TMDs), responsible for the chloride ion translocation pathway, two cytoplasmic Nucleotide Binding Domains (NBDs) with ATPase activity that controls channel gating, and a regulatory (R) domain that mediates *CFTR* activity upon phosphorylation [8]. More than 2000 variants have been already described in the *CFTR* gene (*CFTR1*, available in <http://www.genet.sickkids.on.ca/>), and around 300 variants are considered pathogenic (*CFTR2*, available in <https://www.cftr2.org/>). These pathogenic variants are classified into major seven classes (IA, IB, II, III, IV, V, and VI), according to the molecular mechanisms' disruption:

✉ Thaiane Rispoli
 thaianerispoli@gmail.com

- ¹ Programa de Pós-Graduação em Biologia Celular e Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil
- ² Centro de Desenvolvimento Científico e Tecnológico, Secretaria da Saúde do Estado do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil
- ³ Faculdade de Farmácia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil
- ⁴ Hospital São Lucas da PUCRS, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil
- ⁵ Serviço de Fibrose Cística - Adultos, Serviço de Pneumologia, Hospital São Lucas da PUCRS, Porto Alegre, Rio Grande do Sul, Brazil
- ⁶ Instituto de Ciências Biológicas, Universidade de Passo Fundo, Passo Fundo, Rio Grande do Sul, Brazil
- ⁷ Programa de Pós-Graduação em Pediatria e Saúde da Criança, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil
- ⁸ Programa de Pós-Graduação em Biologia Celular e Molecular Aplicada à Saúde, Universidade Luterana do Brasil, Canoas, Rio Grande do Sul, Brazil

Published online: 25 July 2020

Springer

synthesis, traffic, or function of the protein [9, 10]. The classes IA, IB, II, and III are associated with a more severe prognosis due to defective protein production, processing, and regulation; the classes IV, V and VI are associated with milder prognosis, because CFTR protein maintains residual activity [9]. However, a mutation expanded classification can be considered with a combination of classes once a single mutation can cause defects in the CFTR protein by multiple mechanisms [11]. Therefore, the study of variants is fundamental to understand the patient prognostic and also to develop new treatments based on gene variants [12].

In this study, we described a novel frameshift variant (S511Lfs*2) resulted from the deletion of two thymines. The frameshift yielded a codon that specifies the amino acid leucine followed by a premature termination codon (PTC). We have combined the patient's clinical features and structural bioinformatics data to propose the variant's effects in the CF phenotype.

Case report

Patient data collection

The subject from a CF Reference Center (Hospital São Lucas da PUCRS) was a 23-year-old female with a confirmed CF diagnosis at the age of twelve. She had the first sweat test intermediate (57 mmol/L) and the second abnormal (97 mmol/L). The patient had a history of respiratory infections with the airway colonized by *Staphylococcus aureus*, severe bronchiectasis—diagnosed by chest computed tomography (CT)—and mucoid impaction of the large airways. This subject also showed a mild airflow limitation with FVC = 2.75 (87% predicted), FEV1 = 2.12 (75% predicted) and a reduced FEV1/FVC ratio = 0.77 (86% predicted), recurrent pancreatitis, pancreatic insufficiency, and critical failure to thrive (weight Z-score of - 0.88 and height Z-score of - 2.04).

Molecular genetic analysis of the CFTR gene

Genomic DNA extraction from peripheral blood was performed by salting out assay [13]. The DNA was submitted to a molecular screening that analyses the presence of the eleven variants usually found in CF patients from south of Brazil and the main frequency variant c.1521_1523delCTT (legacy: F508del) found in CF patients worldwide. The eleven variants panel was carried out by SNaPshot [14] and the c.1521_1523delCTT was analysed by Sanger sequencing method as following. A PCR reaction was performed to amplify the exon 11 of the CFTR gene using specific primers (5'-TGAATCCGTGAGCGTGATTG-3' and 5'-TGGGTA GTGTGAAGGGTTCAT-3'). The PCR product was purified

with ExoSAP-IT PCR (Affymetrix, USA) according to the manufacture's protocol and used in the Sanger sequencing that was carried out with BigDye Terminator Cycle Sequencing kit v3.1 (Applied Biosystems, USA). Then the product was precipitated by ethanol/ethylenediaminetetraacetic acid (EDTA) protocol recommended by Applied Biosystems chemistry guide [15]. The sequencing system used was the Genetic Analyzer 3130xl (Applied Biosystems, USA). Sequences obtained were aligned with the CFTR gene sequence (e.g. RefSeq NG_016465.4) and examined using SeqMan software v.17 (DNASTAR, USA).

Cloning of the novel variant

To precisely characterize the deletion, an allele-specific cloning and sequencing method was carried out. The exon 11 of the CFTR gene was amplified by PCR using specific primers that contains a target sequence at the 3'-end of the primer, a sequence for restriction enzyme (BamHI in the forward primer and EcoRI in the reverse) and 6pb at the 5'-end of the primer to ensure efficient DNA cleavage by the restriction enzymes (CFTR_BamHI_F 5'-tttttGGATCCTGAATC CTGAGCGTGATTG-3' and CFTR_EcoRI_R 5'-tttttGAA TTGATGCTTTGATGACGCTTCTG-3'). The PCR product was precipitated using the sodium acetate/ethanol method [16]. Each 1 µg of DNA from PCR product purified and pUC18 vector were sequentially digested with EcoRI and BamHI (Thermo Fisher Scientific) according to the manufacturer's instructions. After that, the DNA reactions were precipitated with sodium acetate/ethanol method. Using a 1:3 molar ratio of vector:insert DNA (100 ng of cut plasmid and 26 ng of cut insert) and 3 units of T4 DNA Ligase (Promega) the ligation reaction was performed according to manufacturer's recommendations and incubated for 18 h at 22 °C. The DNA was precipitated with sodium acetate/ethanol method and then transformed into XL1-Blue competent cells that were plated on 25 mL of LB solid medium with 0.85 mg X-Gal, 1 mM IPTG and 2.5 mg ampicillin. The plate was incubated overnight at 37 °C. White colonies resulted from transformation were used as a template to perform a PCR with the M13 primers to confirm the correct ligation. The Sanger sequencing with M13 primers was performed as described above in the "Molecular genetic analysis of the CFTR gene" section.

Structure and pathogenicity prediction

The ExpASy software was used to obtain the translation of the nucleotide sequence to a protein sequence [17]. The putative protein structure was built using UCSF chimera software based on the structure found in Protein Data Bank (PDB ID: 5UAK) [18]. The MutationTaster software was used to predict the pathogenicity considering evolutionary

conservation, mRNA, and protein structure/function through phyloP/phastCons scores [19]. The PhyloP measure the conservation or acceleration, either lineage-specific or across all branches and PhastCons the conservation scoring and identification of conserved elements [19].

Results

The SNaPshot genetic analysis identified a compound heterozygote. One variant was c.1000C>T (legacy: R334W), which has been previously reported in the CFTR2 database (Fig. 1). The other variant was a novel deletion variant found by Sanger sequencing; however, it was challenging to identify the deletion correctly (Fig. 2a). Thereby, to characterize this novel variant, we isolated the sequences by allele-specific cloning (Fig. 2b).

The novel variant, S511Lfs*2 was discovered in exon 11 of the *CFTR* gene, where two thymines at position g.98818_98819del (NG_016465.4) were deleted (Fig. 2b). As a result, this deletion formed the premature termination codon UGA after replacing serine for leucine being

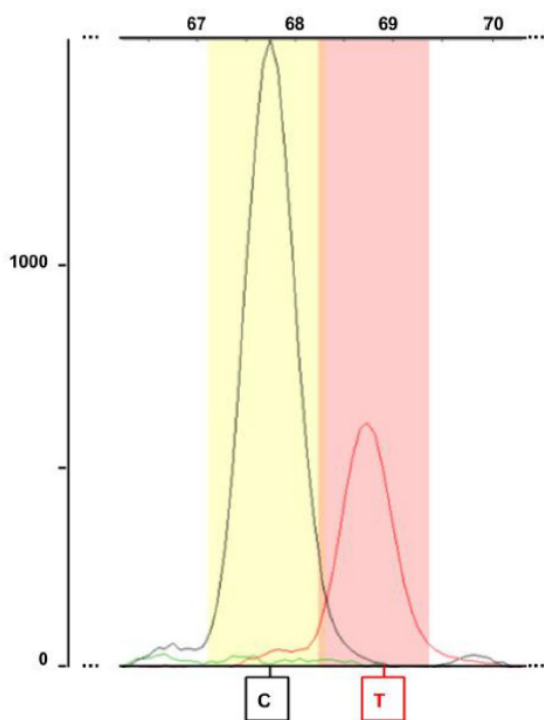


Fig. 1 Detection of c.1000C>T (legacy name: R334W) variant in heterozygosity in *CFTR* gene. Electropherogram of SNaPshot methodology: C allele (wild-type; black peak) and T allele (variant; red peak)

a frameshift variant type (Fig. 2c). The novel variant was deposited in Cystic Fibrosis Mutation Database (CFTR1) as well as in the public archive of human variations and phenotypes (ClinVar; SCV000987313) and in the public archive of Single Nucleotide Polymorphism database (dbSNP; rs1562898471).

In silico analysis in MutationTaster software suggests that this deletion can modify functions of key regions such as the ATP-binding domain and the PDZ-binding domain (Fig. 3). The value obtained of two conservation parameters used, PhyloP (1.528–4.75 value) and PhastCons (1 value), revealed that deleted thymines are in a conserved region in the *CFTR* gene corroborating with pathogenic features. Furthermore, this PTC promotes a loss of approximately 65% of full-length *CFTR* protein (Fig. 2c), possibly resulting in nonsense-mediated mRNA decay (NMD). The structure of the protein lacking amino acids after the termination codon is shown in Fig. 3.

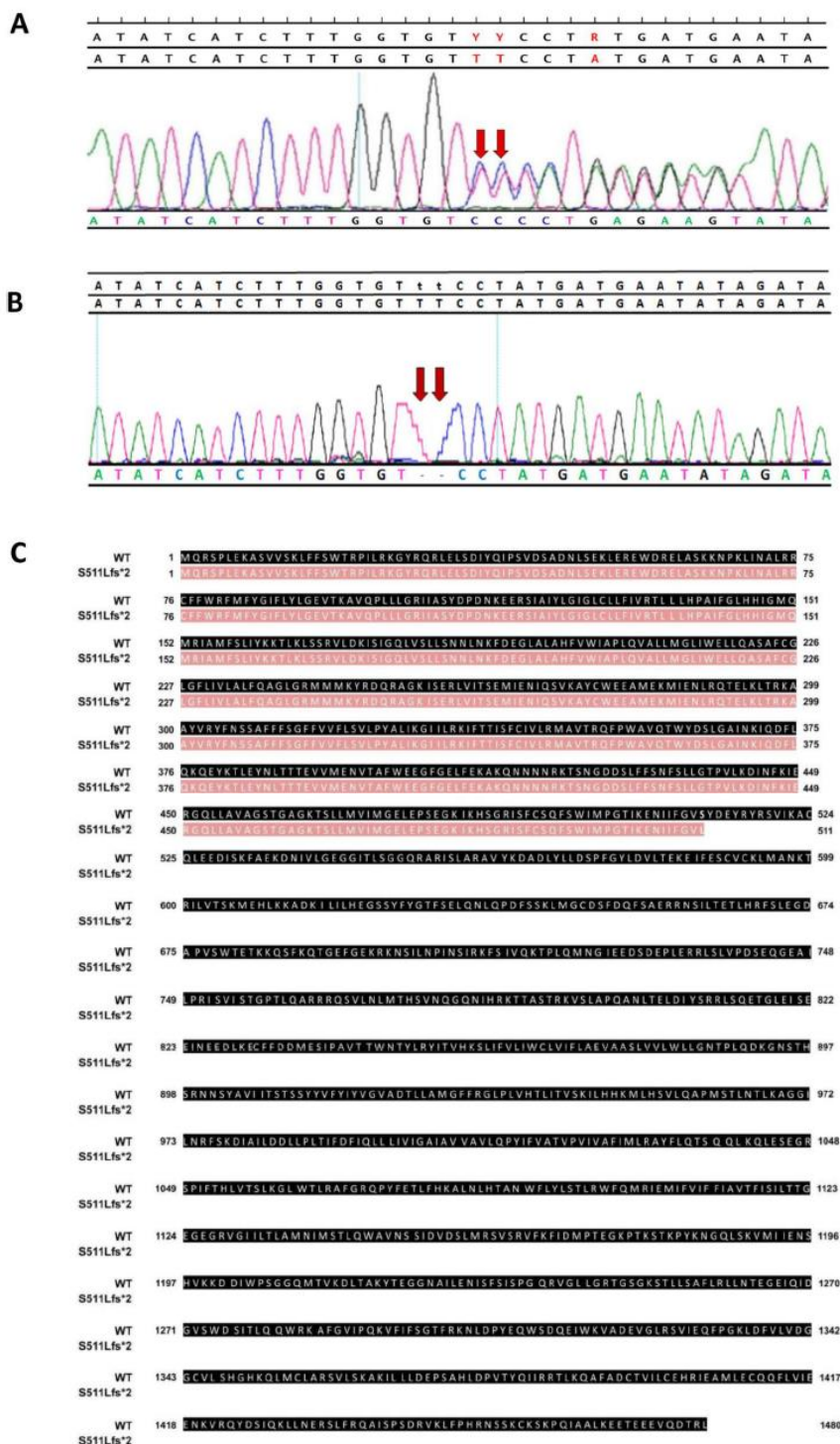
Discussion

This is the first study worldwide to describe the S511Lfs*2 variant identified in a Brazilian CF patient. Variants identification is essential for diagnosis purposes, specific treatment approaches, and genetic counseling [21]. We also demonstrate an in silico model to elucidate the protein alteration at a structural level.

The S511Lfs*2 variant leads to a change in the amino acids sequence (S511L) and promotes the appearance of a PTC. The PTC leads to the production of truncated and non-functional *CFTR* protein. It is possible that this modification may activate a post-transcriptional mechanism called nonsense-mediated mRNA decay (NMD) which detects and degrades the mutated transcript [22–24]. However, more studies would be required to evidence this mechanism. Thus, it is reasonable to classify this variant as a class IB due to the possible NMD activation, associated with pathogenicity prediction values, clinical data from the patient, and defective protein production. [9, 10]. It is important to emphasize that independently of the NMD activation, the protein structure would be strongly affected due to the absence of full-length functional *CFTR* protein (Figs. 2c and 3). Consequently, protein functionality would be lost.

Structural analysis of *CFTR* protein shows that the novel variant S511Lfs*2 is in the NBD1 domain, near the c.1521_1523delICTT (legacy: F508del) variant. The in silico analysis in this region showed a high conservation grade. Thus, any alterations in this domain have a deleterious impact on the protein by to affect essential regions, such as ATP-binding domain, which is essential to allow the open-closed status of the channel, and the PDZ-binding

Fig. 2 From nucleotide sequence to the amino acid sequence of S511Lfs*2 protein. **a** Sequence electropherogram of the exon 11 (*CFTR* gene) showing the novel variant S511Lfs*2 in heterozygous state. **b** Detection of the novel S511Lfs*2 (deletion of two thymines) variant in one of the alleles after allele-specific cloning. The two red arrows confirm the deletion of two thymines in one allele. **c** CFTR protein amino acid sequence obtained from ExPASy software. The wild-type (black) and the mutated (pink) CFTR amino acid sequence are represented



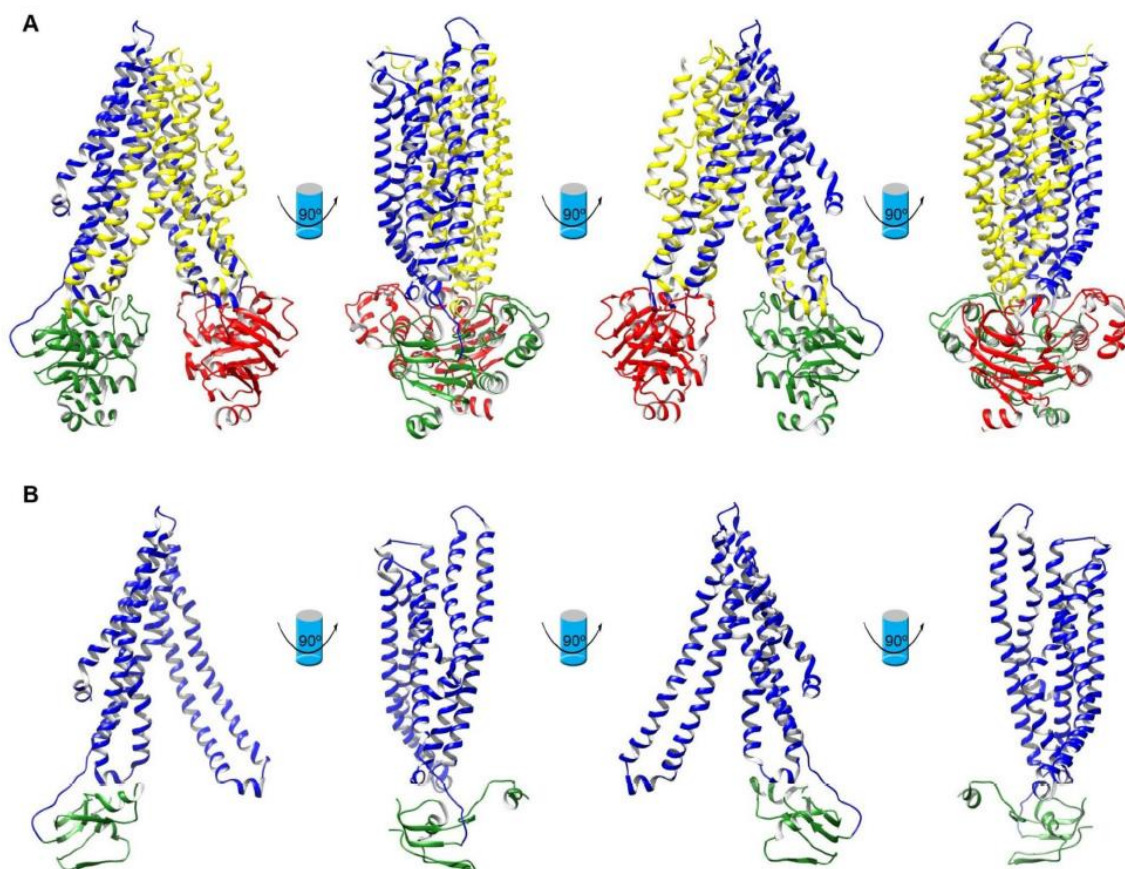


Fig. 3 CFTR three-dimensional structures. The dephosphorylated CFTR structure was modeled using the template 5UAK, from Protein DataBank [20]. The R domains are not represented. **a** Ribbon representation of the CFTR wild-type protein and **b** the CFTR S511Lfs*2

protein. The structure was rotated 90° to highlight all the faces of the protein. The functional domains TMD1 (blue), TMD2 (yellow), NBD1 (green), and NBD2 (red) are represented

domain, that play a key role in anchoring proteins in the cell membrane [25].

Additional studies demonstrated that a single-residue deletion in the region 491–525 elicits an array of abnormalities in CFTR function [26]. Although it was found that specifically the amino acids V510 and S511 deletions did not cause problems in the processing of the CFTR protein, our data show that the new mutation S511Lfs*2 is caused by the deletion of two T nucleotides in different codons. This modification is a frameshift variant, causing the loss of the S511 amino acid and changing the reading phase, thus altering all subsequent positions within the critical region 511–525, in addition to also causing the formation of a premature codon stop.

Furthermore, this novel variant would potentially eliminate the R domain (Fig. 3), which is responsible by the channel activation and thus essential for the functionality

of the protein. The loss of this domain was predicted based on sequence and structural analysis, since this domain was not resolved by the Electron Cryomicroscopy (cryo-EM) structure [20, 26].

We identified the S511Lfs*2 variant in compound heterozygosity with c.1000C>T (legacy: R334W), which represents the fourth most frequent disease-causing variant among Brazilian CF patients and leads to a mild CF phenotype (class IV) [10, 27]. The severity of the clinical features indicates that, although the S511Lfs*2 variant is associated with no CFTR protein synthesis, the c.1000C>T (legacy: R334W) variant confers the residual CFTR channel activity, which could explain the delayed appearance of the symptoms and the late diagnosis of this patient [28, 29].

The knowledge of new variants, their classification, and mechanisms are essential to understand CF pathogenicity and drugs discovery able to correct defective proteins [10,

23]. Furthermore, *in silico* tools are a helpful strategy to predict the degree pathogenicity of the uncertain variants, as shown in this work.

Conclusions

In this study, Sanger sequencing detected a novel variant in a compound heterozygote that was confirmed by allele-specific analysis *in vitro*. The *in silico* analysis shows that S511Lfs*2 is a frameshift variant and generates a PTC which leads to the production of truncated and non-functional CFTR protein. Moreover, CF patients with this novel variant could benefit from therapies that improve the expression of full-length CFTR protein.

Acknowledgements The authors are grateful to Arnaldo Zaha for providing the laboratory facilities to carry out the allele-specific cloning assay of the novel variant.

Author contributions Thaiane Rispoli and Grazielle Rodrigues: designed the experiments, performed the experiments, analysed, and interpreted the data, drafted the manuscript, read, and approved the final manuscript. Mayara J. Prado: designed, drafted, and performed the *in vitro* allele-specific cloning, revised the manuscript, read, and approved the final manuscript. Leonardo Araújo Pinto, Marcelo Tadday Rodrigues, Cynthia Rocha Dullius: collected phenotypic data from patients, revised the manuscript, read, and approved the final manuscript. Tarciana Grandi, Cláudia Maria Dornelles da Silva, José Eduardo Vargas: revised the manuscript, read, and approved the final manuscript. Maurício Menegatti Rigo, Maria Lucia Rossetti: supervised the study, read, and approved the final manuscript.

Funding The study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES – Finance Code 001) and Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS, 1260-2551/13-0).

Compliance with ethical standards

Conflict of interest Authors declared no conflict of interest.

Ethical approval The subject enrolled in the study was a volunteer from the CF Reference Center (Hospital São Lucas da PUCRS). The study was conducted according to the 1964 Helsinki declaration and all procedures involving human subjects were approved by the Local Ethics Committee of the Hospital São Lucas – PUCRS and Secretaria da Saúde do Estado do Rio Grande do Sul.

Informed consent Written informed consent was obtained from the subject enrolled in the study. Declaration of consent in the study from the participant is available from the correspondence author.

References

- Ratjen F, Bell SC, Rowe SM, Goss CH, Quittner AL, Bush A (2015) Cystic fibrosis. *Nat Rev Dis Prim* 1:15010
- Gadsby DC, Vergani P, Csanády L (2006) The ABC protein turned chloride channel whose failure causes cystic fibrosis. *Nat Rev* 440:477–483
- Meng X, Wang Y, Wang X, Wrennall JA, Rimington TL, Li H, Cai Z, Ford RC, Sheppard DN (2017) Two small molecules restore stability to a subpopulation of the cystic fibrosis transmembrane conductance regulator with the predominant disease-causing mutation. *J Biol Chem* 292:3706–3719
- Blouquit S, Regnier A, Dannhoffer L, Fermanian C, Naline E, Boucher R, Chinnet T (2006) Ion and fluid transport properties of small airways in cystic fibrosis. *Am J Respir Crit Care Med* 174:299–305
- Zhang L, Button B, Gabriel SE et al (2009) CFTR delivery to 25% of surface epithelial cells restores normal rates of mucus transport to human cystic fibrosis airway epithelium. *PLoS Biol* 7:e1000155
- Shah VS, Meyerholz DK, Tang XX et al (2016) Airway acidification initiates host defense abnormalities in cystic fibrosis mice. *Science* 351:503–507
- Coakley RD, Grubb BR, Paradiso AM, Gatzky JT, Johnson LG, Kreda SM, O'Neal WK, Boucher RC (2003) Abnormal surface liquid pH regulation by cultured cystic fibrosis bronchial epithelium. *Proc Natl Acad Sci USA* 100:16083–16088
- Hwang T-C, Yeh J-T, Zhang J, Yu Y-C, Yeh H-I, Destefano S (2018) Structural mechanisms of CFTR function and dysfunction. *J Gen Physiol* 150:539–570
- Marson FAL, Bertuzzo CS, Ribeiro JD (2016) Classification of CFTR mutation classes. *Lancet Respir Med* 4:e37–e38
- De Boeck K, Amaral MD (2016) Progress in therapies for cystic fibrosis. *Lancet Respir Med* 4:662–674
- Veit G, Avramescu RG, Chiang AN et al (2016) From CFTR biology toward combinatorial pharmacotherapy: expanded classification of cystic fibrosis mutations. *Mol Biol Cell* 27:424–433
- Martins RS, Campos Junior M, dos Santos Moreira A, Marques Zembrzusi V, da Fonseca ACP, Abreu G, de Cabello M, de Cabello PH (2019) Identification of a novel large deletion and other copy number variations in the CFTR gene in patients with cystic fibrosis from a multiethnic population. *Mol Genet Genomic Med* 7:e645
- Mousquer GT, Maciel LP, Pompeu Saraiva AC, Dalla Costa ER, Rosa Rossetti ML (2018) Validation of a quick and low-cost DNA extraction protocol applicable to long-stored blood samples. *Anal Biochem* 561–562:47–51
- Rispoli T, Martins de Castro S, Grandi T, Prado M, Filippon L, Dornelles da Silva CM, Vargas JE, Rossetti LMR (2018) A low-cost and simple genetic screening for cystic fibrosis provided by the Brazilian Public Health System. *J Pediatr* 199:272–277.e3
- Yang DL, Sauvageot R, Pentoney SL (2007) DNA sequencing by capillary electrophoresis. *Electrophoresis*. <https://doi.org/10.1002/elps.200900218>
- Zeugen JA, Hartley JL (1985) Ethanol precipitation of DNA. *Focus (Madison)* 7:1–2
- Artimo P, Jonnalagedda M, Arnold K et al (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res* 40:W597–W603
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera: a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612
- Schwarz JM, Cooper DN, Schuelke M, Seelow D (2014) MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* 11:361–362
- Liu F, Zhang Z, Csanády L, Gadsby DC, Chen J (2017) Molecular structure of the human CFTR Ion channel. *Cell* 169:85–95.e8
- Patch C, Middleton A (2018) Genetic counselling in the era of genomic medicine. *Br Med Bull* 126:27–36

22. Lejeune F (2017) Nonsense-mediated mRNA decay at the crossroads of many cellular pathways. *BMB Rep* 50:175–185
23. Zainal Abidin N, Haq IJ, Gardner AI, Brodrie M (2017) Ataluren in cystic fibrosis: development, clinical studies and where are we now? *Expert Opin Pharmacother* 18:1363–1371
24. Clarke LA, Awatade NT, Felício VM et al (2018) The effect of premature termination codon mutations on CFTR mRNA abundance in human nasal epithelium and intestinal organoids: a basis for read-through therapies in cystic fibrosis. *Hum Mutat* 40:23692
25. Guggino WB (2004) The cystic fibrosis transmembrane regulator forms macromolecular complexes with PDZ domain scaffold proteins. *Proc Am Thorac Soc* 1:28–32
26. Chen X, Zhu S, Zhenin M, Xu W, Bose SJ, Wong MPF, Leung GPH, Senderowitz H, Chen JH (2019) A defective flexible loop contributes to the processing and gating defects of the predominant cystic fibrosis-causing mutation. *FASEB J* 33:5126–5142
27. Grupo Brasileiro de Estudos de Fibrose Cística (2016) The Brazilian Cystic Fibrosis Patient Registry
28. Kristidis P, Bozon D, Corey J, Danuta Markiewicz M, Rommens J, Tsui L-C, Durie P (1992) Genetic determination of exocrine pancreatic function in cystic fibrosis. *Am J Hum Genet* 50:1178–1184
29. Sheppard DN, Rich DP, Ostedgaard LS, Gregory RJ, Smith AE, Welsh MJ (1993) Mutations in CFTR associated with mild-disease-form Cl⁻ channels with altered pore properties. *Nature* 362:160–164

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

CURRICULUM VITÆ resumido

Prado, M.J.

1. DADOS PESSOAIS

Nome: Mayara Jorgens Prado

Local e data de nascimento: Quaraí/RS, 01/02/1991

Endereço profissional: Av. Bento Goncalves, 9500. Centro de Biotecnologia.
CEP 91501970, Porto Alegre/RS.

E-mail: mayjorgens@hotmail.com

ORCID: 0000-0003-0647-4429

2. FORMAÇÃO:

2018/1- atual: Doutorado em Biologia Celular e Molecular no PPGBCM (UFRGS) com estágio no LABPADEY (UniBE/ Suíça).

2014/2-2017/1: Mestrado em Biologia Celular e Molecular no PPGBCM (UFRGS).

2010/01-2013/2: Graduação em Biotecnologia com ênfase e biotecnologia molecular (UFRGS).

3. ESTÁGIOS:

- 2020/2-2021/1: Bolsa de pesquisa de 12 meses no LABPANDEY do programa *Swiss Government Excellence Scholarship* (ESKAS) sob orientação do Professor Amit Pandey.

- 2013/2: O estágio curricular, que ocorreu no oitavo semestre da graduação, foi realizado no Centro de Desenvolvimento Científico e Tecnológico (CDCT), sob orientação da Prof. Dra. Maria Lucia Rosa Rossetti e supervisão do Dr. Rodrigo Rodenbusch. Durante esse estágio, acompanhei o desenvolvimento da tese do doutorando Rodrigo Rodenbusch intitulada “Análise de SNPs em genes de pigmentação humana em indivíduos fenotipicamente distintos para este caráter”, a qual gerou um artigo que está submetido na revista *International Journal of Molecular Sciences* (fator de impacto 3.257).
- 2011/1-2013/2: Bolsista de Iniciação Científica do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) sob orientação da Prof. Dra. Célia Carline e coorientação da Dra. Fernanda Stanisçuaski no Laprotox. Durante o estágio desenvolvi o projeto intitulado “Tolerância diferencial ao alumínio em soja: um estudo sobre o possível envolvimento de aquaporinas”. Nesse mesmo período, pude acompanhar o projeto da doutoranda Angela Menegassi intitulado “Análise de expressão e caracterização estrutural e funcional de aquaporinas de soja (*Glycine max*)”.
- 2010/2: Estágio voluntário de Iniciação Científica sob orientação da Prof. Dra. Célia Carline e coorientação da Dra. Fernanda Stanisçuaski no Laboratório de Proteínas Tóxicas (Laprotox). Durante esse estágio, acompanhei o projeto intitulado “Identificação de genes de aquaporinas em feijão”.

4. EXPERIÊNCIA PROFISSIONAL OU DIDÁTICA ANTERIOR

- Monitora de Curso prático de extensão para Professores e Estudantes do Ensino Médio. “Curso de Férias – Edição: A Célula. Entre 2015/2 e 2016/1 (40 horas).

- Aula de biologia molecular na Universidade Luterana do Brasil, sub supervisão da Prof.^a regente da disciplina, Dra. Maria Lucia Rosa Rossetti. Entre 2019-2021. Carga horária: 30 horas aula.

5. ARTIGOS COMPLETOS PUBLICADOS

PRADO MJ, SINGH S, LIGABUE-BRAUN R, MENEGHETTI BV, RISPOLI T, KOPACEK C, MONTEIRO K, ZAHA A, ROSSETTI MLR, PANDEY AV. Characterization of Mutations Causing CYP21A2 Deficiency in Brazilian and Portuguese Populations. *International Journal of Molecular Sciences*, 23(1):296, 2022.

RISPOLI, T., RODRIGUES, G. M., PRADO, M. J., PINTO, L. A., RODRIGUES, M. T., DULLIUS, C. R., GRANDI, T., DA SILVA, C., VARGAS, J. E., RIGO, M. M., & ROSSETTI, M. L. Novel frameshift variant of the CFTR gene: S511Lfs*2 from phenotype to molecular predictions. *Molecular biology reports*, 47(8), 6463–6469, 2020.

RISPOLI, T., RODRIGUES, G. M., DE CASTRO, S. M., PRADO, M. J., DA SILVA, C., GRANDI, T., FISCHER, G. B., PINTO, L. A., MARÓSTICA, P., SCORTEGAGNA, L., MOCELIN, H. T., VARGAS, J. E., & ROSSETTI, M. Cystic Fibrosis: A Simple and Customized Strategy for Genetic Screening Able to Detect Over 90% of Identified Mutated Alleles in Brazilian Newborns. *Molecular diagnosis & therapy*, 24(3), 315–325, 2020.

KOPACEK, C., PRADO, M. J., DA SILVA, C., DE CASTRO, S. M., BELTRÃO, L. A., VARGAS, P. R., GRANDI, T., ROSSETTI, M., & SPRITZER, P. M. Clinical and molecular profile of newborns with confirmed or suspicious congenital adrenal hyperplasia detected after a public screening program implementation. *Jornal de pediatria*, 95(3), 282–290, 2019.

RISPOLI, T., MARTINS DE CASTRO, S., GRANDI, T., PRADO, M., FILIPPON, L., DORNELLES DA SILVA, C. M., VARGAS, J. E., & ROSSETTI, L. A Low-Cost and

Simple Genetic Screening for Cystic Fibrosis Provided by the Brazilian Public Health System. *The Journal of pediatrics*, 199, 272–277.e3, 2018.

PRADO, M. J., DE CASTRO, S. M., KOPACEK, C., DE MELLO, M. P., RISPOLI, T., GRANDI, T., DA SILVA, C., & ROSSETTI, M. Development of CYP21A2 Genotyping Assay for the Diagnosis of Congenital Adrenal Hyperplasia. *Molecular diagnosis & therapy*, 21(6), 663–675, 2017.

KOPACEK, C., DE CASTRO, S. M., PRADO, M. J., DA SILVA, C. M., BELTRÃO, L. A., & SPRITZER, P. M. Neonatal screening for congenital adrenal hyperplasia in Southern Brazil: a population based study with 108,409 infants. *BMC pediatrics*, 17(1), 22, 2017.

6. RESUMOS E TRABALHOS APRESENTADOS EM CONGRESSOS (durante o período do doutorado)

PRADO, MAYARA J.; DA SILVA, CLAUDIA M.D, LIGABUE-BRAUN, RODRIGO.; KOPACEK, CRISTIANE ; DE CASTRO, SIMONE M. ; WIEST, PALOMA. ; RISPOLI, THAIANE ; PONCES, BETANIA. ; GRANDI, TARCIANA ; ZAHA, ARNANDO ; ROSSETTI, M.L.R. Analysis of the effect of mutations in the CYP21A2 gene on the structure of the 21-hydroxylase protein. 2019. (7th International DSD Symposium).

PONCES, BETÂNIA, PRADO, MAYARA J.; WIEST, PALOMA., DE CASTRO, SIMONE M.; BELTRÃO, LUCIANA A.; GRANDI, TARCIANA; ROSSETTI, M.L.R; DA SILVA, CLAUDIA M.D.; KOPACEK, CRISTIANE. CYP11B1 Mutations in patients with suspected Congenital Adrenal Hyperplasia (HAC) in the public neonatal screening in South Brazil: two case reports. 2019. (7th International DSD Symposium).

PRADO, MAYARA; DA SILVA, CLAUDIA M.D. ; KOPACEK, CRISTIANE ; DE CASTRO, SIMONE M. ; WIEST, P. ; RISPOLI, THAIANE ; PONCES, B. ; GRANDI, TARCIANA ; ZAHA, A ; ROSSETTI, M.L.R . Congenital Adrenal Hyperplasia:

genetic analysis of children that had borderline elevation of 17OHP in the newborn screening. 2018. (XXII International Congress of Genetics).

GRANDI, TARCIANA ; DA SILVA, CLAUDIA M.D. ; RISPOLI, THAIANE ; PRADO, M. J. ; RODRIGUES, G ; PONCES, B. ; DE CASTRO, SIMONE M. ; ROSSETTI, LUCIA MARIA ROSA . Application of Molecular Diagnosis of Hemoglobin Disorders in the Public Health System in Southern Brazil. 2018. (XXII International Congress of Genetics).

RISPOLI, THAIANE ; DE CASTRO, SIMONE M. ; Rodrigues, G ; AIRES, G. ; PRADO, MAYARA J. ; MOCELIN, T. ; CHAPPER, M. ; GRANDI, T. ; DA SILVA, CLAUDIA M.D. ; VARGAS, J. E. ; Rossetti, M.L.R . Genetic Screening for Cystic Fibrosis Provided by the Brazilian Public Health System: SNaPshot Panel of Eleven Variants and the Main Mutations Found. 2018. (XXII International Congress of Genetics).

PRADO, M. J.; Dornelles C.M ; KOPACEK, C. ; CASTRO, S. M. ; MELLO, M. P. ; SERRANO, T. R. ; GRANDI, T. ; Rossetti, M.L.R . Congenital Adrenal Hyperplasia: mutation detection of CYP21A2 gene in cases from southern Brazil. 2017. (13th International Congress of Inborn Errors of Metabolism).

CASTRO, S. M. ; Dornelles C.M ; PRADO, M. J. ; GRANDI, T. ; BELTRAO, L. ; CHAPPER, M. ; Rossetti, M.L.R ; VARGAS, P. ; KOPACEK, C. . Congenital Adrenal Hyperplasia neonatal screening in Southern Brazil: the challenges until confirmatory diagnosis. 2017. (13th International Congress of Inborn Errors of Metabolism).

CASTRO, S. M. ; SERRANO, T. R. ; PRADO, M. J. ; GRANDI, T. ; Dornelles C.M ; Rodrigues, G ; MOCELIN, H. ; FISCHER, G. ; CHAPPER, M. ; MACEDO, J. ; KOPACEK, C. . Molecular biology laboratory as an important support in the final diagnosis of the diseases triaged in newborn screening in Southern Brazil. 2017. (13th International Congress of Inborn Errors of Metabolism).