

**UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA**

RAQUEL MARTINEZ RAMOS

**Expansão do diagnóstico molecular da Síndrome de Insensibilidade
aos Andrógenos**

São Paulo 2023

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA

RAQUEL MARTINEZ RAMOS

**Expansão do diagnóstico molecular da Síndrome de Insensibilidade
aos Andrógenos**

**Dissertação apresentada à Faculdade de Medicina da Universidade de São Paulo
para a obtenção do título de mestre em Ciências Médicas**

Área de concentração: Distúrbios Genéticos de Desenvolvimento e Metabolismo

Orientador: Rafael Loch Batista

São Paulo 2023

Dados Internacionais de Catalogação na Publicação (CIP)

Preparada pela Biblioteca da
Faculdade de Medicina da Universidade de São Paulo

©reprodução autorizada pelo autor

Ramos, Raquel Martinez
Expansão do diagnóstico molecular da Síndrome de
Insensibilidade aos Andrógenos / Raquel Martinez
Ramos. -- São Paulo, 2023.
Dissertação (mestrado)--Faculdade de Medicina da
Universidade de São Paulo.
Programa de Ciências Médicas. Área de
Concentração: Distúrbios Genéticos de
Desenvolvimento e Metabolismo.
Orientador: Rafael Loch Batista.

Descritores: 1.Síndrome da insensibilidade a
andrógenos 2.Diferenças do desenvolvimento sexual
3.Diagnóstico molecular 4.Receptores androgênicos
5.Variantes genéticas 6.Pequenas indels

USP/FM/DBD-443/23

Responsável: Erinalva da Conceição Batista, CRB-8 6755

Este trabalho foi desenvolvido na Unidade de Endocrinologia do Desenvolvimento, Laboratório de Hormônios e Genética Molecular LIM/42 da Disciplina de Endocrinologia do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo.

Agradecimentos

Primeiramente aos meus pais, Sandra e Claudio, por me proporcionar a vida, educação, amor e por terem me incentivado nela em todas as suas fases. Sem eles, nada disso teria sido possível.

Á toda minha família pelo apoio mesmo nos momentos de ausência em que tive que me dedicar.

Ao meu orientador Dr. Rafael Loch Batista pela paciência, dedicação, conhecimento e pela confiança para realizar esse trabalho. Tenho certeza de que continuaremos a alcançar grandes conquistas juntos.

À Dra. Berenice Mendonça, que é um exemplo de liderança e me motivou a seguir carreira no mundo acadêmico, expresso minha sincera gratidão.

Á toda equipe do Hospital das Clínicas, LIM 42, em especial para as minhas colegas da secretaria Val e Rô por tornarem o ambiente de trabalho mais acolhedor e amigável. Á Mirian pelo suporte de bancada e gentileza.

Aos meus colegas de laboratório, Anna, Augusto, Bárbara e Elina que se tornaram verdadeiros amigos ao longo desta jornada, e ao Felipe, por ter aceitado o desafio de me ensinar as técnicas de bancada do laboratório logo que cheguei.

Á todos os meus amigos pelo constante apoio e fonte de incentivo, resgatando minha confiança inúmeras vezes ao longo da realização deste trabalho.

Á Caju pelos momentos de relaxamento e alívio do estresse. Seu carinho e presença foram essenciais.

Á todos os pacientes presentes nesse estudo e suas famílias.

Á CAPES, pelo financiamento.

“Não fiz o melhor, mas fiz tudo para que o melhor fosse feito. Não sou o que deveria ser, mas não sou o que era antes.” — Martin Luther King

Resumo

Ramos, RM. Expansão do diagnóstico molecular da Síndrome de Insensibilidade aos Andrógenos. [dissertação]. São Paulo. Faculdade de Medicina, Universidade de São Paulo; 2023.

Introdução: A Síndrome da Insensibilidade aos Andrógenos (AIS) é a causa mais comum de Distúrbios do Desenvolvimento Sexual (DDS) em pacientes com cariótipo 46,XY. AIS se caracteriza por déficit de virilização em um feto do sexo cromossômico masculino, resultando em uma genitália atípica ou feminina ao nascimento. Essa desordem resulta de disfunção no Receptor Androgênico (AR), causando um quadro de resistência hormonal. A expressão desta síndrome ocorre sobre 3 fenótipos conhecidos: completo, parcial e leve (CAIS, PAIS e MAIS, respectivamente). No entanto, enquanto variantes no AR são identificadas em cerca de 90% dos CAIS, isso é verdade para apenas 28 – 50% dos PAIS. **Objetivo:** Expandir o diagnóstico molecular da Síndrome de Insensibilidade aos Andrógenos. **Metodologia:** Seleccionamos pacientes com suspeita de AIS. Realizamos a extração de DNA genômico e conduzimos o sequenciamento por Sanger. Todas as variantes identificadas foram posteriormente reclassificadas e analisadas quanto ao seu perfil mutacional. Para nossa revisão sistemática de pequenas *indels*, consultamos bancos de dados relacionados ao receptor de androgênio (AR) e as classificamos de acordo com o tipo de *indel* e a localização genômica. **Resultados:** Identificamos 55 variantes distintas no AR. Dessas, 30 estavam presentes em pacientes com CAIS, incluindo 50% de variantes *missense*, 36,7% de variantes *nonsense* (n=11) e 13,3% de variantes em sítios de splicing (n=4). Por outro lado, em pacientes com PAIS, todas as variantes identificadas eram do tipo *missense*. Há duas variantes classificadas como VUS duas delas em um paciente PAIS e uma dela em um paciente com fenótipo CAIS. Identificamos 82 *small indels* diferentes no gene AR em indivíduos com AIS. A maioria desses casos (n=78; 95,1%) estava associada à CAIS, enquanto quatro casos foram relacionados à MAIS. Também identificamos uma dupla variante e uma grande CNV causando CAIS. **Conclusão:** Nosso estudo ampliou o diagnóstico molecular da Síndrome de Insensibilidade aos Andrógenos, identificando uma variedade de variantes genéticas, aumentando o entendimento desta complexa condição genética.

Palavras-chave: Síndrome da insensibilidade a andrógenos. Diferenças do desenvolvimento sexual. Diagnóstico molecular. Receptores androgênicos. Variantes genéticas. Pequenas *indels*.

Abstract

Ramos, RM. Expansion of molecular diagnosis of Androgen Insensitivity Syndrome..[dissertação].São Paulo. “Faculdade de Medicina, Universidade de São Paulo”; 2023.

Introduction: Androgen Insensitivity Syndrome (AIS) is the most common cause of Disorders of Sex Development (DSD) in patients with a 46,XY karyotype. AIS is characterized by a lack of virilization in a male chromosomal fetus, resulting in atypical or female genitalia at birth. This disorder results from dysfunction in the Androgen Receptor (AR), leading to a hormonal resistance condition. AIS presents in three known phenotypes: complete (CAIS), partial (PAIS), and mild (MAIS). However, while AR variants are identified in about 90% of CAIS cases, this is true for only 28-50% of PAIS cases. **Objective:** To expand the molecular diagnosis of Androgen Insensitivity Syndrome. **Methodology:** We selected patients with suspected AIS, performed genomic DNA extraction, and conducted Sanger sequencing. All AR variants identified were subsequently reclassified and analyzed for their pathogenicity potential. For the systematic review of AR small indels, we classified the small indels based on the type and genomic location. **Results:** We identified 55 distinct AR variants. Of these, 30 were present in patients with CAIS (50% *missense* variants, n=15; 36.7% *nonsense*, n=11, and 13.3% splicing variants, n=4). All AR variants identified in PAIS patients were *missense non-synonymous* variants. Two AR variants were classified as VUS. We identified 82 different *small indels* in the AR gene in individuals with AIS. The majority of these cases (n=78; 95.1%) were associated with CAIS, while four cases were related to MAIS. We also identified a double variant in an individual with CAIS and a large CNV causing CAIS. **Conclusion:** Our study expanded the molecular diagnosis of AIS, identifying novel AR allelic variants, enhancing the understanding of this complex genetic condition. **Keywords:** Androgen Insensitivity Syndrome; Disorders of Sex Development; Molecular Diagnosis

Keywords: Androgen insensitivity syndrome. Sexual development differences. Molecular diagnosis. Androgen receptor. Genetic variants. Small indels.

Lista de Abreviaturas

ACMG : *American College of Medical Genetics and Genomics*

AF1: *Activation function-1*

AF2: *Activation Function-2*

AIS: Síndrome da Insensibilidade aos Andrógenos

AR: Receptor Androgênico

CAIS: Síndrome da Insensibilidade Androgênica Completa

CNV: *Copy number variantion*

DDS: Diferenças do desenvolvimento sexual

DBD: *DNA binding domain*

FATHMM: *Functional Analysis through Hidden Markov Models*

GERP++: Genomic Evolutionary Rate Profiling Plus Plus

LBD: *Ligand Binding Domain*

MLPA: *Multiplex Ligation-dependent Probe Amplification*

MAIS: Síndrome da Insensibilidade aos Andrógenos Leve

NTD: *N-terminal Domain*

PAIS: Síndrome da Insensibilidade Androgênica Parcial

PCR: Reação em cadeia da polimerase

SRC: Steroid Receptor Coactivator

WES: *Whole Exome Sequencing*

Lista de Figuras

Figura 1 - Estrutura do gene e proteína do receptor androgênico humano. O AR está localizado no cromossomo X (Xq11.2) sendo composto por oito exons e dividido nos seguintes domínios: domínio de ligação ao DNA (DBD) ou no domínio de ligação ao ligante (LBD), NTD: domínio N-terminal, HR : Região Hinge. AF-1: <i>activation function-1</i> ; AF-2: <i>activation function-2</i>	19
Figura 2 - Etapas da técnica de MLPA (MRC-Holland, adaptado).	28
Figura 3 - Distribuição das variantes identificadas no AR por domínios do AR entre os fenótipos completo e parcial.	32
Figura 4 - CNV no gene AR em paciente com CAIS identificada por MLPA.	38
Figura 5 - Heredograma das <i>small indels</i> inéditas no gene do receptor androgênico identificadas neste estudo.	39
Figura 6 - <i>Small indels</i> identificadas em pacientes com AIS.	40
Figura 7 - A visão geral das pequenas inserções e deleções (indels) dentro do gene do receptor androgênico na Síndrome de Insensibilidade Androgênica. A) Frequência de indels NFS (n=18; 21%) e indels FS (n=79; 79%). B) Localização dos indels – em região não.	41

Lista de Tabelas

Tabela 1 – Classificação das diferenças do desenvolvimento sexual.....	16
Tabela 2 - Exemplos dos tipos de indels nas regiões codificantes – adaptado de Montgomery et al. ³²	29
Tabela 3 - Variantes alélicas do tipo <i>missense</i> inéditas identificadas em pacientes com AIS neste estudo.	33
Tabela 4 - Classificação ACMG da dupla variante.....	35
Tabela 5 - Variantes alélicas previamente reportadas em resíduos que compõem o AF-2 do AR em indivíduos com AIS.	35
Tabela 6 - Fenótipos de AIS em pacientes com variante do tipo <i>missense</i> no AR que resultou em uma prolina.	36
Tabela 7 - Small indels inéditas no gene do receptor androgênico em pacientes com Síndrome da Insensibilidade Androgênica identificadas neste estudo.....	39
Tabela 8 - Variantes encontradas em pacientes com AIS em nossa coorte e reclassificação (continua).	70
Tabela 9 - Variantes do tipo small indels identificadas em pacientes com AIS (continua).....	74

Sumário

1	INTRODUÇÃO	15
1.1	Diferenças do Desenvolvimento Sexual (DDS).....	16
1.2	Síndrome da Insensibilidade aos Andrógenos	17
1.3	Diagnóstico molecular da AIS.....	17
1.4	Gene Receptor Androgênico (AR)	18
2	JUSTIFICATIVA	20
3	OBJETIVOS	22
3.1	Objetivo Principal	23
3.2	Objetivos específicos	23
4	METODOLOGIA.....	24
4.1	Considerações éticas.....	25
4.2	Pacientes	25
4.3	Extração do DNA genômico	25
4.4	Sequenciamento por Sanger.....	26
4.5	Análise <i>in silico</i> e classificação das variantes	26
4.6	MLPA para CNV no AR.....	27
4.7	Revisão Sistemática das variantes <i>indels</i> no AR.....	28
5	RESULTADOS	31
5.1	Variantes no AR encontradas em pacientes com AIS.....	32
5.1.1	Variantes inéditas encontradas em pacientes com AIS.....	33
5.2	Dupla variante no gene AR em paciente com CAIS	34
5.3	Variantes na região AF2 do gene do AR.....	35
5.4	CNV no gene do Receptor Androgênico	37
5.5	<i>Small Indels</i> identificadas no gene AR	38
5.5.1	<i>Small indels</i> inéditas no AR identificadas em nossa coorte de pacientes com AIS	38
5.5.2	Revisão sistemática das <i>small indels</i> no AR identificadas na literatura e em bancos de dados.....	39

6	DISCUSSÃO	42
7	CONCLUSÕES.....	53
	REFERÊNCIAS	55
	ANEXOS	67
	APÊNDICES	81

1 INTRODUÇÃO

1 INTRODUÇÃO

O desenvolvimento sexual humano envolve três etapas: a determinação do sexo cromossômico, do sexo gonadal e do sexo fenotípico ^{1; 2}. Na determinação do sexo cromossômico a presença dupla do cromossomo X resulta em um cariótipo 46, XX, enquanto que a presença do cromossomo Y resultará no cariótipo 46, XY ³. A seguir, a gônada bipotencial se diferenciará em uma gônada feminina (ovário) ou masculina (testículo), a partir da expressão orquestrada de genes envolvidos na diferenciação gonadal ⁴. Quando a gônada bipotencial se diferencia em um testículo, as células de Leydig produzirão andrógenos, sobretudo a testosterona, os quais promoverão a diferenciação dos ductos de Wolff em epidídimos, ductos deferentes e vesículas seminais caracterizando a virilização da genitália interna ⁵. A testosterona produzida pelo testículo é, então, convertida ao seu metabólito mais potente, a dihidrotestosterona, a qual tem importante papel na virilização da genitália externa e na determinação do sexo fenotípico ⁶.

1.1 Diferenças do Desenvolvimento Sexual (DDS)

As diferenças do desenvolvimento sexual (DDS) incluem condições congênitas nas quais o desenvolvimento do sexo cromossômico, gonadal ou fenotípico é atípico ⁷. DDS pode resultar de variações numéricas ou estruturais nos cromossomos sexuais, bem como de alterações em quaisquer dos genes relacionados ao desenvolvimento gonadal, a síntese, a metabolização ou a ação dos esteroides sexuais ⁸. Diversos mecanismos fisiopatológicos estão implicados na etiologia do DDS, que podem ser categorizados em três principais grupos: DDS resultantes de anormalidades cromossômicas (DDS por anormalidade cromossômica), DDS 46XY e DDS 46,XX. (Tabela 1).

Tabela 1 – Classificação das diferenças do desenvolvimento sexual.

Diferenças do desenvolvimento sexual
DDS por anormalidade do sexo cromossômico
<ul style="list-style-type: none"> • 45,X (Síndrome de Turner e variantes) • 47,XXY (Síndrome Klinefelter e variantes) • 45,X/46,XY (Disgenesia gonadal mista) • 46,XX/46XY (DDS ovotesticular-quimera)
DDS 46,XY
<ul style="list-style-type: none"> • Distúrbios do desenvolvimento testicular (agenesia e disgenesia gonadal completa e parcial, síndrome da

regressão testicular embrionária)

- Distúrbios da síntese da testosterona (defeito no receptor de LHCG, defeitos da síntese de esteroides da suprarrenal e testicular)
- Distúrbios da metabolização da testosterona (deficiência da enzima 5 α -redutase 2)
- Distúrbios da ação da testosterona (insensibilidade androgênica completa e parcial)
- Outros (associados a malformações urogenitais)
- Origem indeterminada

DDS 46,XX

- Distúrbios do desenvolvimento ovariano (disgenesia gonadal, DDS ovotesticular, DDS testicular)
 - Por excesso de andrógenos (de origem fetal, fetoplacentaria, ou materna)
 - Outros associado a malformações urogenitais
 - Origem indeterminada
-

1.2 Síndrome da Insensibilidade aos Andrógenos

A síndrome da insensibilidade androgênica (AIS) é a causa mais comum de DDS em indivíduos com cariótipo 46,XY⁹. AIS apresenta-se clinicamente sob três fenótipos: Completo (CAIS), onde os indivíduos afetados apresentam genitália externa tipicamente feminina ao nascimento; Parcial (PAIS), onde indivíduos afetados apresentam genitália atípica ao nascimento e Leve (MAIS), onde os indivíduos afetados apresentam genitália externa tipicamente masculina e a suspeita diagnóstica deste fenótipo ocorre pela ausência de virilização na puberdade, presença de ginecomastia e/ou infertilidade^{10; 11}.

Embora AIS seja uma doença tipicamente monogênica, variantes alélicas no AR são descritas em cerca de 90% dos casos de CAIS mas apenas em 50% dos casos com fenótipo parcial¹², o que sugere que alterações na regulação, ação ou tradução do AR possam estar implicados no mecanismo molecular da síndrome de insensibilidade androgênica.

1.3 Diagnóstico molecular da AIS

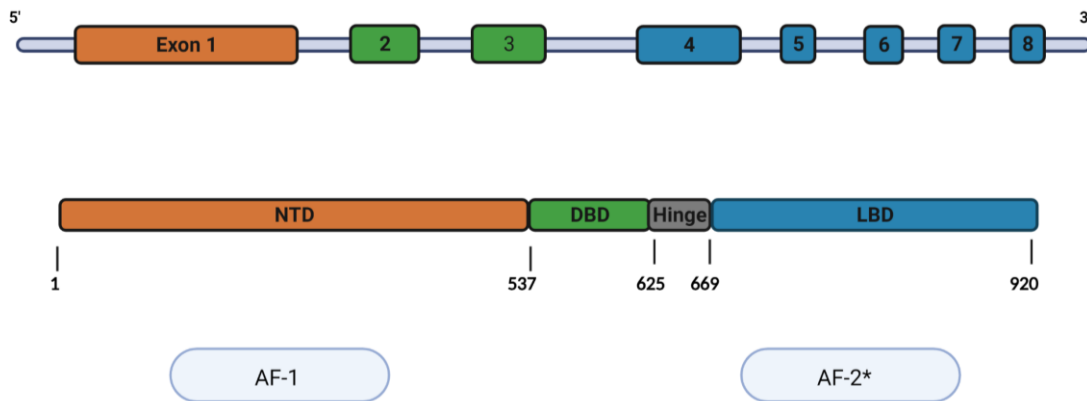
O diagnóstico da Síndrome de Insensibilidade aos Andrógenos (AIS) é confirmado pela presença de variantes alélicas no AR^{9; 12}. Cerca de 30% das mutações no AR na AIS são de novo¹². Há mais de 1000 variantes no AR descritas em um banco de dados associado à AIS e câncer de próstata (<http://www.mcgill.ca/androgendb>), e cerca de 600 delas foram descritas em pacientes com AIS¹³. As variantes no AR identificadas em pacientes com AIS são reportadas ao longo de todo o gene AR, sendo mais frequentes no exon 1 (o maior exon do AR, que codifica o

domínio N-terminal, NTD) ¹⁴. Variantes no domínio NTD são mais frequentes em pacientes com CAIS, enquanto variantes nos exons 5 e 6 (que codificam o domínio de ligação ao ligante, LBD) são mais frequentes em pacientes com PAIS ¹³. Quase todas as variantes no AR reportadas em indivíduos com MAIS foram encontradas no domínio NTD, mas proporcionalmente há um baixo número de mutações reportadas no AR relacionadas a este fenótipo ¹¹.

As variantes alélicas mais comuns do AR em todos os fenótipos da AIS são variantes pontuais não sinônimas do tipo *missense*. Inserções e deleções que causam uma mudança de leitura de *frame* levando a um códon de parada prematuro são relatadas com mais frequência em pacientes com CAIS ¹⁵. Raramente, variantes alélicas sinônimas deletérias foram descritas em pacientes com PAIS e em indivíduos com CAIS ^{16;17}.

1.4 Gene Receptor Androgênico (AR)

O gene AR codifica o receptor androgênico, o qual é um receptor nuclear que desempenha um papel crítico na resposta do corpo aos hormônios androgênicos, como a testosterona e a di-hidrotestosterona (DHT) ¹⁸. O gene do AR está localizado no cromossomo X, no locus Xq11-Xq12, e é composto por 2757 nucleotídeos. Possui 8 éxons, com íntrons de tamanho que variam entre 0.7 a 2.6 kb e codifica uma proteína de 920 aminoácidos ¹⁹. Assim como outros membros da superfamília de receptores nucleares, o AR consiste em 3 domínios funcionais: *N-terminal domain* (NTD; resíduos 1 - 555), *DNA binding domain* (DBD; resíduos 555 - 623) e *Ligand Binding Domain* (LBD; resíduos 665 - 920) ^{20;21}. O éxon 1 codifica o NTD, os éxons 2 e 3 codificam o DBD e os éxons de 4 a 8 codificam o LBD e a região hinge (4), sendo todos importantes para a função do receptor ²². O AR é semelhante aos demais receptores nucleares nos domínios DBD, LBD e na região hinge ²³. No entanto, o AR apresenta um domínio NTD bem mais extenso, com um número polimórfico de repetições CAG e GGC (poliglutamina e poliglicina, respectivamente) ²⁴. O domínio NTD do AR abriga a região AF1 (*transcriptional activation function-1*), enquanto o LBD abriga a região AF2 (*transcriptional activation function-2*) (Figura 1) ²⁵. A região AF2 desempenha um papel fundamental na ativação do AR, servindo como um ponto de ancoragem para várias proteínas coativadoras, como os membros da família p160 (SRC, coativadores de receptores de esteroides) e da família ARA (proteínas associadas ao receptor de andrógeno) ²⁶. Devido à importância da região AF2 na ativação do AR, há um crescente interesse em desenvolver medicamentos que possam inibir a ação do AR através da ligação a essa região como potenciais agentes antiandrogênicos ²⁶.



Resíduos AF-2:718,714,731,734,735,898,894,895

Figura 1 - Estrutura do gene e proteína do receptor androgênico humano. O AR está localizado no cromossomo X (Xq11.2) sendo composto por oito exons e dividido nos seguintes domínios: domínio de ligação ao DNA (DBD) ou no domínio de ligação ao ligante (LBD), NTD: domínio N-terminal, HR : Região Hinge. AF-1: *activation function-1*; AF-2: *activation function-2*.

2 JUSTIFICATIVA

2 JUSTIFICATIVA

Considerando que AIS é a condição molecular mais prevalente entre indivíduos com DDS 46,XY e o receptor androgênico se relaciona a diversas condições endócrinas, explorar o diagnóstico molecular dos casos de AIS permitirá expandir o conhecimento sobre o papel deste receptor hormonal no desenvolvimento sexual através do seu impacto fenotípico.

3 OBJETIVOS

3 OBJETIVOS

3.1 Objetivo Principal

Expandir o diagnóstico molecular da Síndrome de Insensibilidade aos Andrógenos.

3.2 Objetivos específicos

- Realizar o sequenciamento do gene *AR* em pacientes suspeitos de Síndrome de Insensibilidade Androgênica (AIS) em acompanhamento no Hospital das Clínicas da Universidade de São Paulo (HCFMUSP).
- Realizar uma revisão abrangente das pequenas inserções e deleções (*small indels*) previamente descritas no gene *AR* na literatura científica.
- Classificar as variantes alélicas identificadas no *AR* com base nos critérios estabelecidos pela *American College of Medical Genetics and Genomics* (ACMG).
- Investigar um caso de grande deleção identificada no gene *AR*.
- Analisar variantes alélicas identificadas na região AF2 do gene *AR* quanto ao perfil mutacional e ao impacto fenotípico.

4 METODOLOGIA

4. METODOLOGIA

4.1 Considerações éticas

O estudo foi conduzido de acordo com princípios éticos seguindo as orientações contidas na declaração de Helsinki e pelos termos descritos pela Portaria 196/96 do Conselho Nacional de Saúde. O estudo foi aprovado pelo comitê de ética e pesquisa do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, número 66367717.6.0000.0068. O Termo de consentimento livre e esclarecido foi aplicados a todos os pacientes envolvidos neste estudo.

4.2 Pacientes

Critérios de seleção dos pacientes: Os pacientes selecionados são acompanhados no Ambulatório da Unidade de Endocrinologia do Desenvolvimento (Unidade A1MN1003) da Disciplina de Endocrinologia da FMUSP. Pacientes com suspeita de insensibilidade androgênica foram submetidos a avaliação clínica a qual incluiu exame físico e exames laboratoriais. Todos os pacientes foram submetidos a avaliação do cariótipo.

Os critérios laboratoriais que caracterizaram um caso como sendo suspeito de AIS são: níveis elevados de testosterona sérica, níveis elevados de gonadotrofinas (especialmente LH), presença de genitália atípica ao nascimento (nos casos de PAIS), cariótipo 46,XY, histórico de amenorreia primária e/ou hérnia inguinal (em casos de CAIS) e presença de ginecomastia (em indivíduos do sexo social masculino).

4.3 Extração do DNA genômico

As amostras de DNA dos pacientes selecionados foram obtidas a partir de leucócitos de sangue periférico. O DNA foi extraído pela técnica de *Salting out* e armazenado de acordo com procedimento padronizado no Laboratório de Hormônios e Genética Molecular/LIM42 da Unidade de Endocrinologia do Desenvolvimento, Faculdade de Medicina da Universidade de São Paulo.

A concentração do DNA extraído foi obtida por leitura em espectrofotômetro no comprimento de onda de 260 nm (1 unidade DO 260 = 50 µg/mL). A relação ideal entre as leituras em 260 e 280 nm para a caracterização da pureza do material deverá ser superior a 1,75.

Caso o número fosse maior que 1,75 a amostra foi diluída em água. As amostras serão mantidas congeladas a 20°C negativos até seu uso.

Todos os pacientes com suspeita de AIS incluídos neste estudo foram submetidos ao sequenciamento do *AR* utilizando a técnica de Sanger.

4.4 Sequenciamento por Sanger

Foram utilizados primers para amplificar por reação em cadeia da polimerase (PCR) as regiões que flanqueiam as variantes, como previamente descrito^{27; 28}. As amplificações foram acompanhadas de um controle negativo e os produtos amplificados serão submetidos a uma purificação enzimática com a enzima ExoSAP-IT, conforme a recomendação do fabricante. O sequenciamento foi realizado com o kit ABI Prism™ Big Dye Terminator Cycle Sequencing Ready Reaction Kit 3.1 (*Life Technologies, Life Technologies, CA, EUA*) e os produtos desta reação foram submetidos a uma eletroforese capilar no sequenciador automático ABI Prism 3100 Genetic Analyzer (*Life Technologies, Life Technologies, CA, EUA*). A leitura dos eletroferogramas foi realizada através do programa Sequencher 4.10.1 (*GeneCodes Corporation, Ann Arbor, EUA*).

4.5 Análise *in silico* e classificação das variantes

Todas as variantes no *AR* identificadas foram classificadas de acordo com a patogenicidade segundo os critérios da ACMG (*American College of Medical Genetics and Genomics, 2015*)²⁹ utilizando a plataforma Franklin (<https://franklin.genoox.com/clinical-db/home>). Segundo este critério, as variantes foram classificadas como patogênica, provavelmente patogênica, variante de significado incerto (VUS), provavelmente benigna e benigna²⁹. Os critérios ACMG estão detalhados no anexo 5. Para cada variante identificada que fosse classificada ao menos como VUS (ou seja, excluindo os casos de variantes classificadas como potencialmente benignas ou benignas), o potencial deletério destas variantes identificadas também foi avaliado pelas ferramentas REVEL e MetaLR^{30; 31}. Para a predição de possível impacto no splicing foi utilizada a plataforma Splice AI (<https://spliceailookup.broadinstitute.org>).

Elegemos a ferramenta REVEL pois trata-se de um método de conjunto para prever a patogenicidade de variantes de sentido incerto com base em uma combinação de pontuações de 13 ferramentas individuais: MutPred, FATHMM v2.3, VEST 3.0, PolyPhen-2, SIFT,

PROVEAN, MutationAssessor, MutationTaster, LRT, GERP++, SiPhy, phyloP e phastCons. O REVEL foi desenvolvido para avaliar a patogenicidade de variantes do tipo *missense*. O escore REVEL para cada variante varia de 0 a 1, sendo que pontuações mais altas refletem uma maior probabilidade de que a variante seja patogênica. Com base na recomendação da ClinGen, uma pontuação acima de 0,932 corresponde a forte evidência patogênica, uma pontuação na faixa (0,773 - 0,932) é considerada moderadamente patogênica, uma pontuação na faixa (0,644 - 0,773) é considerada como tendo algum potencial de patogenicidade, uma pontuação na faixa (0,183 - 0,29) indica possível benignidade, uma pontuação na faixa (0,016 - 0,183) é considerada moderadamente benigna, uma pontuação na faixa (0,003 - 0,016) é considerada fortemente benigna e uma pontuação abaixo de 0,003 é considerada muito fortemente benigna³⁰.

O MetaLR é uma ferramenta cuja predição de patogenicidade se baseia em regressão logística (LR), a qual incorpora 10 outras ferramentas de predição *in silico* (SIFT, PolyPhen-2 HDIV, PolyPhen-2 HVAR, GERP++, MutationTaster, Mutation Assessor, FATHMM, LRT, SiPhy, PhyloP), considerando também a frequência observada nas populações a partir do *1000 Genomes* (<https://www.internationalgenome.org/>). A pontuação varia entre 0 a 1, sendo que valores mais altos são mais propensos a serem deletérios.

4.6 MLPA para CNV no AR

As alterações estruturais identificadas no AR através da técnica de Sanger foram confirmadas com a técnica de Multiplex-ligation dependent probe amplification (MLPA). Para essa técnica utilizamos o kit SALSA MLPA Reagent KIT com sondas específicas para o AR (SALSA MLPA Probemix P074 AR).

A técnica de MLPA envolve quatro etapas principais, conforme ilustrado na Figura 2. Primeiro, ocorre a desnaturação do DNA, no qual a amostra é submetida ao termociclador por 5 minutos. Em seguida, são adicionadas sondas específicas para hibridização com o DNA target. Cada sonda consiste em dois oligonucleotídeos de DNA de cadeia simples, com sequências de anelamento de 60 a 80 nucleotídeos. Após a hibridização, a etapa de ligação ocorre, a qual a enzima SALSA Ligase-65 une os oligos anelados. Na etapa seguinte, a amplificação das sondas aneladas é realizada por PCR, com a adição do reagente da polimerase contendo a enzima SALSA Polimerase e primers.

Finalmente, os dados brutos obtidos são utilizados como entrada no software Coffalyser.net, desenvolvido pela MRC Holland. O Coffalyser.net normaliza os dados de

MLPA, comparando cada amostra de teste com um conjunto de amostras controle ou de referência, resultando no cálculo de razões para cada sonda de cada amostra. Essas razões são visualizadas graficamente para cada sonda de cada amostra. Essas razões são visualizadas graficamente para análise.

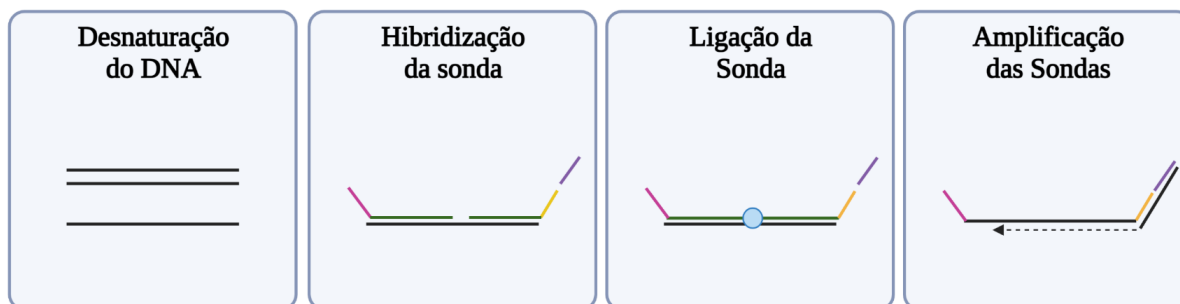


Figura 2 - Etapas da técnica de MLPA (MRC-Holland, adaptado).

4.7 Revisão Sistemática das variantes *indels* no AR

Para identificar *small indels* já descritas no AR em pacientes com AIS, buscamos no PubMed desde o seu início até 15 de abril de 2023, sem restrições de idioma ou regiões geográficas, utilizando os termos de busca "Receptor Androgênico" E "Síndrome de Insensibilidade aos Andrógenos" OU "Distúrbios do Desenvolvimento Sexual" OU "DDS". Em paralelo, realizamos uma busca por variantes alélicas do Receptor Androgênico em dois grandes bancos de dados: o HGMD (<https://www.hgmd.cf.ac.uk/>) e o Androgen Receptor Database (<http://androgendb.mcgill.ca/>).

Para analisar a patogenicidade das pequenas *indels* no gene AR, utilizamos diversas ferramentas computacionais, incluindo PolyPhen2, SIFT, REVEL, MetaLR, MetaRNN, MetaSVM, Mutation Assessor, Mutation Taster, FATHMM, FATHMM-MKL, LRT, M-CAP e DAAN a partir da plataforma Franklin (<https://franklin.genoox.com/clinical-db/home>). Todas as *small indels* identificadas no gene AR foram classificadas de acordo com os critérios da ACMG (REF). Para garantir uma descrição correta e uniforme das variantes, utilizamos a ferramenta Mutalyzer (<https://mutalyzer.nl/>).

Comparamos a frequência de *small indels* entre os pacientes com AIS a partir do HGMD (<https://www.hgmd.cf.ac.uk/>) e o Androgen Receptor Database (<http://androgendb.mcgill.ca/>) com dados populacionais de 2 bancos de dados genômicos públicos: o gnomAD (<https://gnomad.broadinstitute.org>) e o Arquivo Brasileiro Online de Mutações (Abraom; <https://abraom.ib.usp.br>).

Na análise dos dados populacionais, utilizamos o *Genome Aggregation Database* (gnomAD) como uma fonte abrangente para investigar as variantes que impactam o gene *AR*. Com o objetivo de fornecer evidências adicionais da associação genética entre pequenas inserções e deleções no gene *AR* e o fenótipo da Síndrome de Insensibilidade aos Andrógenos (AIS), comparamos essas *small indels* em todo o gene *AR* entre os casos de AIS e comparamos com a frequência destas variantes entre os indivíduos 46,XY do banco de dados público gnomAD, o qual foi utilizado como grupo de controle. Para a comparação entre a frequência de *small indels* entre indivíduos com AIS e os controles 46,XY utilizamos o teste exato de Fisher, sendo a significância estatística definida com um valor de p inferior a 0,05.

As *small indels* (inserções ou deleções de até 50 nucleotídeos) identificadas nessas análises foram então classificadas como *frameshift* (FS) e *não-frameshift* (NFS). Foram consideradas *frameshift indels* (inserções ou deleções) aquelas em que o número de bases alterado na sequência de DNA não era múltiplo de três (43).

Entre as *small indels* no *AR* identificadas, realizamos uma análise detalhada do contexto genômico dessas *indels* envolvendo sequências repetitivas, homopolímeros e repetições trinucleotídicas. As *indels* foram classificadas como associadas a alteração na contagem de cópias" (CCC) ou não-CCC.

Para dividir as *small indels* a partir do contexto genômico na qual elas se localizam dividimos as *small indels* de acordo com a localização em sequências repetitivas ou não repetitivas do genoma. Aquelas localizadas em regiões repetitivas foram classificadas em: 1) *small indels* localizadas em sequência de homopolímeros (homopolimers run - HR), que são sequências de seis ou mais nucleotídeos idênticos e 2) *small indels* localizadas em sequências de repetições do tipo tandem (Tandem Repeats - TR).

As *small indels* presentes em regiões não repetitivas foram avaliadas quanto a mudança local na contagem de cópias (CCC), sendo classificadas como NR-CCC quando alteravam a contagem de cópias e como NR não-CCC quando não alteravam a contagem de cópias no local de inserção. Essa classificação seguiu os critérios estabelecidos por Montgomery³² e estão exemplificadas na tabela 2:

Tabela 2 - Exemplos dos tipos de indels nas regiões codificantes – adaptado de Montgomery et al.³²

Gene	Indel	Contexto Genômico	Tipo
OR5AC2	+A	ATTCTG a AAAAAAAAAAGTCTAAAAGGGCA	HR
CAPN11	+GGCTGCC	AGGGGACTGCTGGCTGCT ggctgcc GGCTGCCATCG	TR
AGAP1	+CAGG	CCTGGC cagg CAGG GCCGAGGGGTGGTCAGAC	NR, CCC
MED4	-AT	GAACAC AT GAGACCTTTCTGGAAGCCAG	NR, non-CCC

5 RESULTADOS

5 RESULTADOS

5.1 Variantes no AR encontradas em pacientes com AIS

Dos pacientes acompanhados no ambulatório de desenvolvimento (AIMN1003) da unidade de Endocrinologia do HCFMUSP identificaram-se um total de 55 variantes genéticas distintas em 94 pacientes com quadro clínico de insensibilidade androgênica e alteração molecular identificada no AR. Dessas, 30 foram observadas em pacientes exibindo um fenótipo completo e 25 com fenótipo parcial. No total foram identificadas 35 variantes do tipo *missense* (63%), 6 variantes do tipo *nonsense* (11%), 6 variantes *small indel* (11%), 4 variantes em sítio de *splicing* (7%), 2 variantes sinônimas (4%), uma variante em região genômica (5' UTR) e uma CNV compreendendo uma deleção incluindo os exons 6 a 8 do AR. Dentre os pacientes apresentando fenótipo parcial, 23 (92%) eram variantes do tipo *missense*. Mutações em sítio de *splicing*, *small indels* e CNV foram identificadas apenas em pacientes com CAIS.

Entre o total de pacientes com CAIS, identificamos dez casos com variantes no domínio N-terminal (NTD), 4 casos no domínio de ligação ao DNA (DBD) e dezesseis casos no domínio de ligação ao ligante (LDB). Em contraste, observou-se uma prevalência de 76% (n=19) de variantes localizadas no LDB (Figura 3).

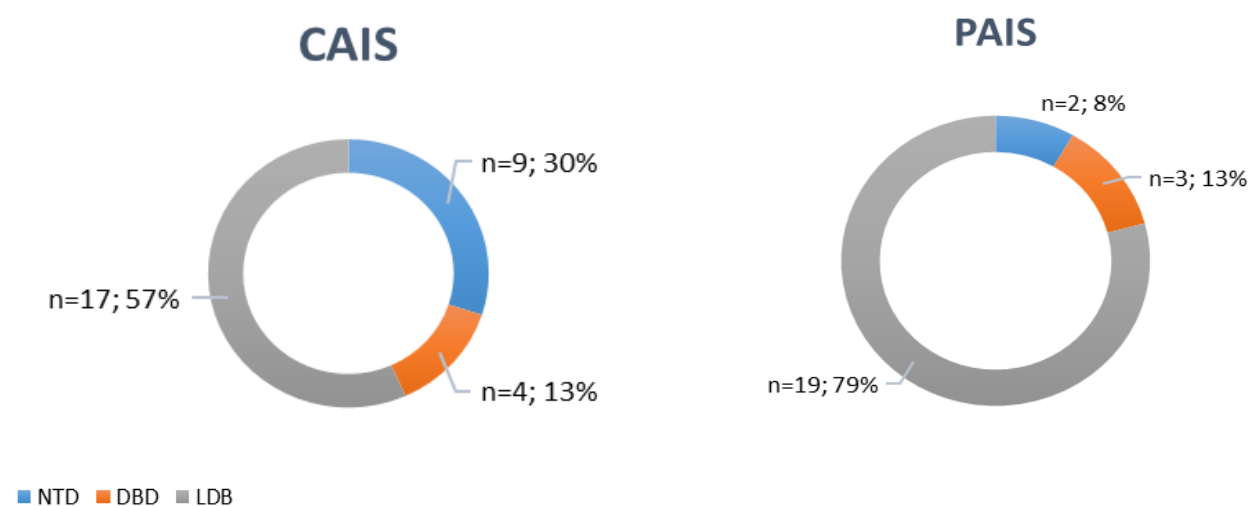


Figura 3 - Distribuição das variantes identificadas no AR por domínios do AR entre os fenótipos completo e parcial.

As 55 variantes foram avaliadas através dos critérios ACMG, das quais 38 foram classificadas como provavelmente patogênicas (69%), 14 como patogênicas (25%) e 3 como Variantes de Significado Desconhecido (VUS) (6%).

Em conjunto, 52 (94.5%) foram classificadas como patogênicas ou potencialmente patogênicas. Houveram três variantes classificadas como VUS, sendo duas delas em pacientes com PAIS (c.1215G>T; p.Gln405His e c.1885G>A; p.Ala629Thr) e uma delas em um paciente com CAIS (c.1530C>T; p.Ser510Ser). A variante sinônima classificada como VUS obteve um *score* de 0,62 atribuído pela ferramenta de análise de *splicing* conhecida como Splice AI.

A tabela com as variantes no *AR* encontradas nos pacientes e sua classificação segundo os critérios ACMG e análise de potencial deletério das variantes *missenses* estão no anexo 1.

Na análise do potencial deletério das variantes *missense*, (n=36) o Revel classificou 12 delas como deletérias moderadas, enquanto 19 variantes foram classificadas como deletérias fortes. Uma única variante foi categorizada como incerta pelo Revel. Variantes são apontadas como incertas quando não há dados experimentais suficientes, quando as ferramentas de predição genética oferecem resultados conflitantes ou quando a variante é rara e não foi estudada extensivamente. Por outro lado, o Meta LR identificou um total de 35 variantes como deletérias e apenas uma variante como benigna, sugerindo um alto grau de impacto funcional na maioria das variantes analisadas. O Revel se mostrou ser uma ferramenta mais sensível para essas variantes.

5.1.1 Variantes inéditas encontradas em pacientes com AIS

Foram identificadas onze novas variantes no *AR* em indivíduos com AIS. Sendo uma delas uma CNV, compreendendo uma deleção que incluía os exons 6 a 8 do *AR*, três variantes do tipo *small indels*, e sete variantes do tipo *missense* não sinônimas. Dentre as variantes *missense*, duas foram encontradas em indivíduos com CAIS, enquanto outras cinco foram observadas em pacientes com PAIS. Todas as variantes foram avaliadas segundo os critérios estabelecidos pelo ACMG, resultando na classificação de sete delas como provavelmente patogênicas e uma variante (c.1885G>A, p.Ala629Thr), sendo categorizada como de significado incerto (VUS). Entretanto, é pertinente ressaltar que todas essas variantes foram classificadas como deletérias a partir das ferramentas disponíveis no site Revel e no Meta LR (Tabela 3).

Tabela 3 - Variantes alélicas do tipo *missense* inéditas identificadas em pacientes com AIS neste estudo.

Domínio	Fenótipo	Posição do codon	Posição da proteína	ACMG	Revel	Meta LR
DBD	PAIS	c.1885G>A	p.Ala629Thr	VUS	Deletéria (Moderada) (0.9)	Deletéria (0.77)
LDB	PAIS	c.2226G>C	p.Trp742Ser	Provavelmente Patogênica	Deletéria (0.97) (Forte)	Deletéria (0.96)
LDB	PAIS	c.2342T>G	p.Met781Arg	Provavelmente Patogênica	Deletéria (0.97) (Forte)	Deletéria (0.99)
LDB	CAIS	c.2595C>G	p.Asp865Glu	Provavelmente Patogênica	Deletéria (Moderada) (0.86)	Deletéria (0.93)
LDB	PAIS	c.2629T>G	p.Phe887Val	Provavelmente Patogênica	Deletéria (0.96) (Forte)	Deletéria (0.8)
LDB	CAIS	c.2674T>G	p.Phe892Val	Provavelmente Patogênica	Deletéria (0.94) (Forte)	Deletéria (0.99)
LDB	PAIS	c.2741C>T	p.Pro914Leu	Provavelmente Patogênica	Deletéria (0.95) (Forte)	Deletéria (0.99)

5.2 Dupla variante no gene *AR* em paciente com CAIS

Identificamos a presença de uma dupla variante em uma paciente com CAIS. Esta variante foi identificada em um indivíduo do sexo social feminino de 13 anos de idade, a qual foi encaminhada para avaliação endocrinológica por queixa de amenorreia primária. Investigação laboratorial demonstrou nível de FSH 6.2 mIU/mL (1.3-19.3), LH 11.4 mIU/mL (1.2-5.6), estradiol 23, testosterona 699 ng/mL (241 – 827: referência masculina). Ao exame físico, apresentava mamas em estágio M3 e ausência de pelos axilares e pubianos.

Frente a suspeita de AIS, o cariótipo foi realizado (46,XY) e o gene *AR* foi sequenciado, sendo identificadas duas variantes alélicas em hemizigose neste gene. A primeira (c.1644G>T; p.Leu548Phe) localizada no exon 2 do gene e a segunda (c.2155T>G; p.Trp719Gly) localizada no exon 4, sendo ambas as variantes localizadas no mesmo alelo. A primeira variante alélica apresenta uma frequência alélica <0.001% (Exome Aggregation Consortium population variant database - rs139524801) e já havia previamente reportada em um paciente com PAIS (31). A segunda é uma variante inédita. No entanto, já foi reportada uma outra variante alélica no

mesmo resíduo (719), mas trocando um Triptofano (Trp) por Ácido glutâmico (Glu) em um paciente com PAIS.

Ambas as variantes foram classificadas como provavelmente patogênicas pelos critérios ACMG. A variante encontrada no exon 2 foi classificada como benigna fraca pelo Meta LR e teve um escore REVEL incerto (0,5). Já a variante encontrada no exon 4 apresentou um potencial deletério mais forte, com o score máximo no Meta LR e um REVEL com pontuação de 0,98. (tabela 5).

Tabela 4 - Classificação ACMG da dupla variante.

Exon	Posição proteína	Posição codon	ACMG	Meta LR	REVEL
2	p.Leu548Phe	c.1644G>T	Provavelmente Patogênica	Benigna(Fraca) (0.38)	Incerto (0,5)
4	p.Trp719Gly	c.2155T>G	Provavelmente Patogênica	Deletéria (1)	Deletério (Forte) (0.98)

5.3 Variantes na região AF2 do gene do AR

Em uma paciente com CAIS, identificamos uma variante na região AF2 do AR (c.2138C>T; p.Leu713Pro). Considerando a importância da região AF2 para a atividade do receptor androgênico, realizamos uma busca sistemática na literatura e identificamos outras dezessete variantes alélicas nesta mesma região do AR (Tabela 6). Verificamos que a variante alélica que identificamos (p.Leu713Pro) já foi previamente reportada, mas em uma paciente com PAIS. Entre as variantes identificadas na região AF2 de gene AR, as variantes p.Leu713Pro, p.Phe726Leu e Ile899Thr são recorrentes, ocorrendo em pacientes não relacionados na literatura. Apenas em quatro casos (23,5%) o fenótipo foi CAIS.

Tabela 5 - Variantes alélicas previamente reportadas em resíduos que compõem o AF-2 do AR em indivíduos com AIS.

Fenótipo	Posição da proteína	Posição do codon	ACMG Franklin	Referência
PAIS	p.Leu713Pro	c.2138T>C	Provavelmente Patogênico	33
CAIS	p.Leu713Val	c.2137C>G	Provavelmente Patogênico	34
PAIS	p.Phe726Leu	c.2176T>C	Patogênico	35
PAIS	p.Val731Met	c.2191G>A	Patogênico	36
PAIS	p.Gln739Arg	c.2216A>G	Patogênico	37
PAIS	p.Glu898Gly	c.2693A>G	Provavelmente Patogênico	10

PAIS	p.Ile899Phe	c.2695A>T	Provavelmente Patogênico	38
PAIS	p.Ile899Thr	c.2696T>C	Provavelmente Patogênico	33
PAIS	p.Arg727Cys	c.2179C>T	Provavelmente Patogênico	39
PAIS	p.Gln734His	c.2202G>C	Provavelmente Patogênico	40
CAIS	p.Gln734Term	c.2200C>T	Provavelmente Patogênico	41
PAIS	p.Ile738Met	c.2214T>G	Provavelmente Patogênico	35
PAIS	p.Ile738Thr	c.2213T>C	Provavelmente Patogênico	42
CAIS	p.Ile738Phe	c.2212A>T	Provavelmente Patogênico	43
CAIS	p.Met895Thr	c.2684T>C	Provavelmente Patogênico	44
PAIS	p.Glu898Asp	c.2694G>T	Provavelmente Patogênico	45

Considerando que em nosso caso com variante na região AF2 do AR o fenótipo foi CAIS e este fenótipo respondeu por apenas cerca de ¼ dos fenótipos observados entre os casos descritos na literatura, realizamos uma revisão sistemática dos casos reportados com variantes do tipo *missense* no AR que alteraram o aminoácido para prolina usando o Clinvar, PubMed e o Androgen Receptor database. No total, identificamos 15 diferentes variantes do tipo *missense* em que a substituição de aminoácidos resultou em uma mudança para prolina (Tabela 7). Em 11 casos (73,5%), essa mudança levou a um fenótipo completo (CAIS). Todas as substituições por prolina localizadas no domínio de ligação ao ligante (LBD) causaram CAIS, enquanto variantes com substituições de prolina no domínio de ligação ao DNA (DBD) e no domínio N-terminal (NTD) resultaram em fenótipos completos e parciais. Entre as variantes que resultaram em CAIS, 7 (63,6%) são resultados da troca de leucina para prolina.

Tabela 6 - Fenótipos de AIS em pacientes com variante do tipo *missense* no AR que resultou em uma prolina.

Fenótipo	Exon/ Domínio	Posição do Codon	Posição da Proteína	ACMG Franklin	Referencia
PAIS	1 Nterm	p.Ser244Pro	c.730T>C	VUS	46
CAIS	1 Nterm	p.Leu263Pro	c.788T>C	VUS	47
CAIS	2 DBD	p.Ala574Pro	c.1720G>C	Patogênico	48
PAIS	3 DBD	p.Thr603Pro	c.1807A>C	Provavelmente Patogênico	38
PAIS	3 DBD	p. Arg616Pro	c.1847G>C	Provavelmente Patogênico	33
CAIS	3 DBD	p. Leu617Pro	c.1850T>C	Patogênico	49
PAIS	3 DBD	p. Arg618Pro	c.1853G>C	Provavelmente Patogênico	50
CAIS	4 LBD	p. Leu678Pro	c.2033T>C	Provavelmente Patogênico	51
CAIS	4 LBD	p. His690Pro	c.2069A>C	Provavelmente Patogênico	52

CAIS	5	LBD	p. Arg753Pro	c.2258G>C	Provavelmente Patogênico	10
CAIS	5	LBD	p. Leu769Pro	c.2306T>C	Provavelmente Patogênico	53
CAIS	6	LBD	p. Leu791Pro	c.2372T>C	Provavelmente Patogênico	54
CAIS	6	LBD	p. Leu813Pro	c.2438T>C	Patogênico	55
CAIS	7	LBD	p. Ser866Pro	c.2596T>C	Provavelmente Patogênico	53
CAIS	8	LBD	p. Leu882Pro	c.2645T>C	VUS	56

5.4 CNV no gene do Receptor Androgênico

Em um caso de um paciente com fenótipo completo da nossa casuística, o sequenciamento por Sanger identificou a ausência de amplificação dos exons 6, 7 e 8 do *AR*. Posteriormente, para confirmar a presença da deleção, a paciente foi submetida ao teste de MLPA.

Na figura 4, as sondas utilizadas no MLPA (anexo 4) que correspondem aos exons deletados estão representadas por esferas vermelhas, indicando que a sequência alvo não foi detectada, ao contrário do que ocorre no paciente controle.

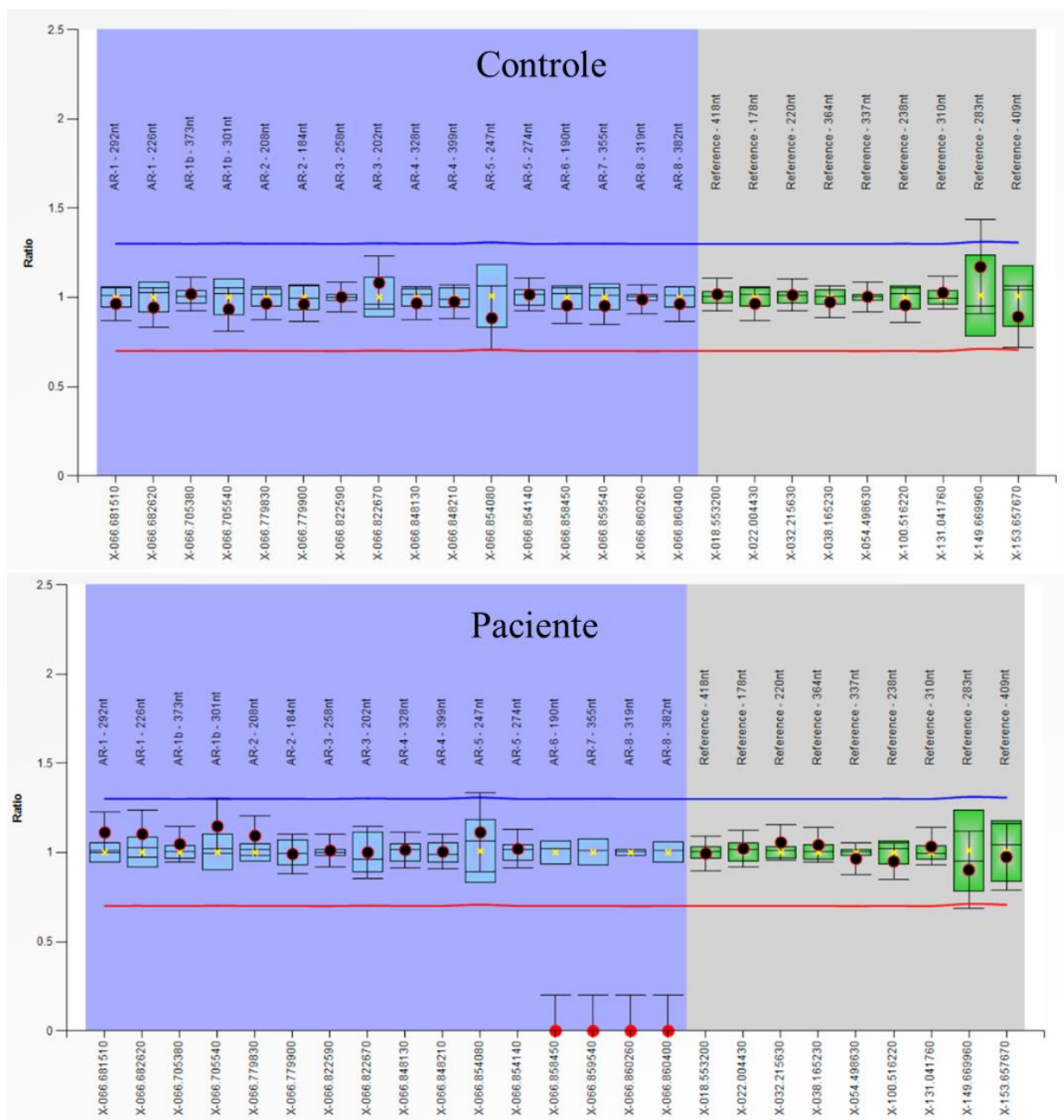


Figura 4 - CNV no gene AR em paciente com CAIS identificada por MLPA.

5.5 *Small Indels* identificadas no gene AR

5.5.1 *Small indels* inéditas no AR identificadas em nossa coorte de pacientes com AIS

Identificamos nove *small indels* em nossa coorte, todos os casos apresentando CAIS, a seguir: p.Gly144Glufs*25, p.Gly130Leufs*27, p.Phe439His*62, p.Gln59Term, p.Gln73Term, p.Gln88Term, p.Ser121Term, p.Asp768Ilefs*21, p.Glu128Term. Das nove, três delas eram

inéditas e estão localizadas no exon 1 do gene *AR*, representadas e classificadas pelos critérios ACMG na Tabela 7. Os heredogramas estão representados na figura 5.

Tabela 7 - Small indels inéditas no gene do receptor androgênico em pacientes com Síndrome da Insensibilidade Androgênica identificadas neste estudo.

Exon	Fenótipo	Posição do codon	Posição da proteína	ACMG
1	CAIS	c.1314_1315delCT	p.Phe439His*62	Patogênica
1	CAIS	c.384_385delGA	p.Gly130Leufs*27	Provavelmente Patogênica
1	CAIS	c.431_452delinsAAA	p.Gly144Glufs*25	Provavelmente Patogênica

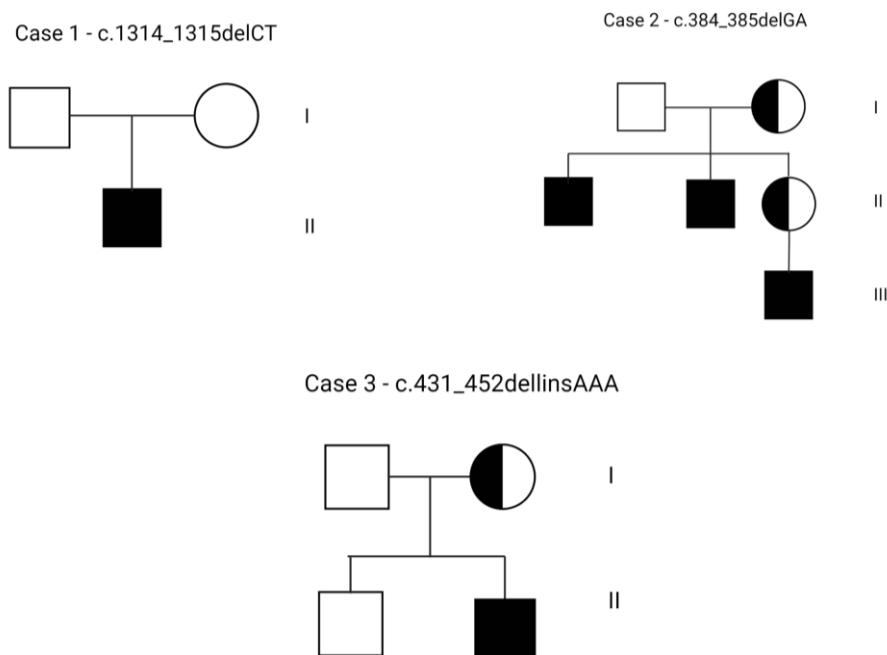


Figura 5 - Heredograma das *small indels* inéditas no gene do receptor androgênico identificadas neste estudo.

5.5.2 Revisão sistemática das *small indels* no *AR* identificadas na literatura e em bancos de dados

Identificamos 82 diferentes *small indels* (incluindo esses três novos) no gene *AR* relatados em indivíduos com AIS (Figura 4) analisamos seu fenótipo, tipo de variante, localização e classificação de acordo com os critérios ACMG (Anexo 1). Na maioria dos casos ($n = 78$; 95,1%), o fenótipo era CAIS. Foram relatados quatro *indels* no gene *AR* em indivíduos com MAIS. A *indel* c.2077_2079del foi relatada em dois indivíduos da mesma família, apresentando o fenótipo PAIS e CAIS. A maioria das *indels* curtas no gene *AR* entre os

indivíduos com AIS foram do tipo *frameshift* (n = 64; 79%). As *indels* foram distribuídas em todo o gene AR, com predominância no domínio NTD (n = 45; 54,9%), seguido pelo domínio LBD (n = 28; 34,1%), domínio DBD (n = 8; 9,7%) e região Hinge (n = 1). Houve uma predominância de pequenas deleções (n = 49; 59,7%), seguidas por pequenas inserções (n = 29; 35,3%) e *indels* complexas (ins + del) do tipo *frameshift* (n = 4; 4,8%). A maioria das *indels* era curta (<10 pb; n = 81 = 98,8%). A maioria das *indels* se localizou em sequências não repetitivas (n = 68; 82,9%). Treze *indels* foram identificados dentro de regiões de trinucleotídeos (15,8%), enquanto sete *small indels* se localizavam em corridas de homopolímeros ou homocopolímeros (8,5%).

Em relação à mudança na contagem de cópias, trinta e nove (47,5%) eram *indels* que modificavam a contagem de cópias no local da inserção (CCC). As *indels* relatadas em indivíduos com MAIS estavam todas localizadas em regiões de repetições em tandem, enquanto que todas as *indels* em sequências não repetitivas causaram CAIS. Foram avaliadas todas as *indels* que ocorrem nas regiões em tandem para mapear a frequência de *indels* dentro de repetições trinucleotídicas no gene AR. Foram identificadas treze *small indels* causando AIS nessas regiões (15,8%). Três se localizam numa sequência de seis prolinas consecutivas no exon 1 (poli-P), seis na sequência de poli-glutamina no éxon (poli-Q) e as três restantes numa sequência de quatro leucinas localizada imediatamente antes poli-Q no éxon 1 (poli-L). Todas as *small indels* que causaram MAIS se localizaram em regiões de polirepetições (poli-L e poli-Q) e todas eram do tipo NFS (não-*frameshift*).

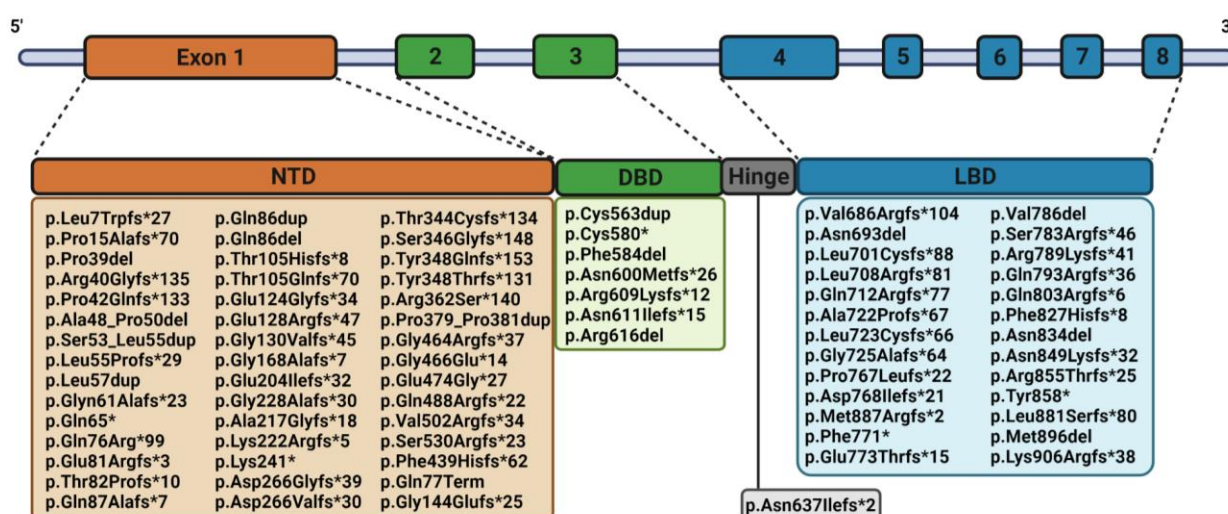


Figura 6 - *Small indels* identificadas em pacientes com AIS.

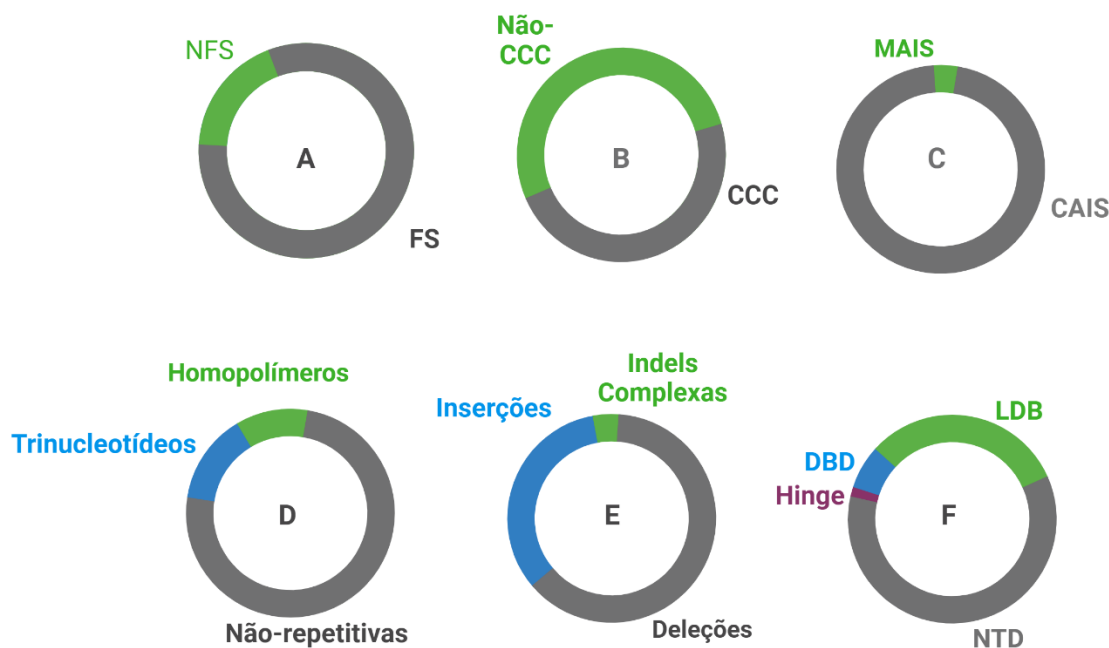


Figura 7 - A visão geral das pequenas inserções e deleções (indels) dentro do gene do receptor androgênico na Síndrome de Insensibilidade Androgênica. A) Frequência de indels NFS (n=18; 21%) e indels FS (n=79; 79%). B) Localização dos indels – em região não -CCC (n=43, 52%) e região CCC (n=39, 48%). C) Fenótipo de AIS – Síndrome de Insensibilidade Androgênica Completa (CAIS) (n=78; 95%) e Síndrome de Insensibilidade Androgênica Leve (MAIS) (n=4, 5%). D) Características exônicas das *indels* – localizadas em sequências de trinucleotídeos (n=13; 16%), em sequências de homopolímeros (n=7; 8,5%) e em sequências não repetitivas (n=62, 70,5%). E) Natureza dos indels – inserções (n=29; 35%), deleções (n=49, 60%) e indels complexas (n=4, 5%). F) Distribuição dos indels – Domínio N-terminal (NTD) (n=45, 55%), Domínio de Ligação ao Ligante (LDB) (n=28, 34%), Região Hinge (n=1) e Domínio de Ligação ao DNA (DBD) (n=8, 10%).

Na análise comparativa, identificamos 79 e 118 *indels* relatadas nos bancos de dados do *AR* e no gnomAD (indivíduos 46, XY), respectivamente. As *indels* do tipo FS apresentaram um enriquecimento substancial em indivíduos com AIS (63/79, 79,9%), em comparação com a população de controle masculina do gnomAD (14/118, 11,9%) ($p < 0,00001$). Além disso, as deleções foram significativamente representadas em indivíduos com AIS (64/79, 81%), em comparação aos controles masculinos (36/118, 30,5%) - $p < 0,00001$. No contexto de *small indels* abrangendo entre um a três nucleotídeos, observamos um enriquecimento significativo em indivíduos com AIS (61/79, 77,2%) em relação aos controles 46,XY (40/118, 33,9%) ($p < 0,00001$).

6 DISCUSSÃO

6 DISCUSSÃO

Dos 55 casos avaliados, 38 (69%) foram classificados como "provavelmente patogênicas" e 14 variantes (25%) foram categorizadas como "patogênicas", perfazendo um total de 94% das variantes sendo classificadas como patogênicas ou provavelmente patogênicas. Três variantes (6%) foram identificadas como "Variantes de Significado Incerto" (VUS). Em estudos moleculares avaliando indivíduos com DDS 46,XY, o percentual de diagnóstico molecular, através da identificação de variantes patogênicas ou provavelmente patogênicas, é elevado para casos de DDS 46,XY associados a defeitos de metabolização da testosterona ou defeitos de ação androgênica. Nesses casos, o percentual de variantes alélicas identificadas nos genes *AR* e *SRD5A2* classificadas como patogênicas ou provavelmente patogênicas é significativamente elevada.

Entre as três variantes classificadas como VUS em nossa casuística, duas delas foram identificadas em pacientes com PAIS (c.1215G>T; p.Gln405His e c.1885G>A; p.Ala629Thr) e uma em uma paciente com CAIS (c.1530C>T; p.Ser510Ser). Foi interessante observar que a variante sinônima c.1530C>T; p.Ser510Ser foi classificada como VUS. Esta variante foi identificada por nosso grupo em uma família multigeracional de pacientes com CAIS¹⁶. Através da hipótese de que uma variante sinônima pode afetar o *splicing*, avaliamos o cDNA desta família e comprovamos o impacto desta variante no *splicing*, o qual gerava uma proteína AR anômala, extremamente curta. Analisamos, então outra variante sinônima reportada em um paciente com PAIS que foi capaz também de alterar o *splicing*¹⁷. Nesse caso, a variante também foi classificada como VUS. No entanto, a análise do potencial impacto no *splicing* através da análise *in silico* utilizando a ferramenta SpliceAI atribuiu scores sugestivos de alterações no *splicing* em ambos os casos. Isso sugere que variantes sinônimas em genes de interesse podem ser avaliadas com mais cuidado, sobretudo utilizando ferramentas de predição de impacto no *splicing*, especialmente se foram classificadas como VUS.

A região AF2 do AR, localizada dentro do domínio de ligação ao ligante (LBD) da porção C-terminal do AR, desempenha um papel crucial na mediação da transcrição gênica após a ligação do ligante²⁵. Ela serve como um local de ancoragem para proteínas coativadoras ou corepressoras que modulam a atividade de transcrição do receptor. A ligação do ligante ao AR induz alterações conformacionais que facilitam a recrutamento de proteínas coativadoras para a região AF2²⁶. Por esse motivo, essa região do AR tem despertado interesse no desenvolvimento de novos fármacos antiandrogênicos.

Considerando a importância desta região, as características fenotípicas manifestadas por indivíduos com AIS que abrigam mutações na região AF2 do AR podem oferecer uma perspectiva interessante sobre a relevância clínica dessa região específica. Nossa análise identificou a presença de 18 variantes alélicas distintas dentro dessa região específica do gene AR entre pacientes com AIS.

Foi interessante notar que, dentre essas 18 mutações, apenas uma minoria ($n = 5$) resultou em CAIS. Nossa hipótese inicial era que variantes ocorrendo na região AF2 predominantemente levariam à CAIS. No entanto, o grande percentual de variantes na região AF2 causando PAIS pode sugerir que a alteração nessa sequência (ou o bloqueio dessa sequência considerando essa sequência como um potencial alvo terapêutico para bloqueio androgênico) não promova um bloqueio completo da atividade do AR.

Como em nosso caso com variante na região AF2 o fenótipo foi CAIS e a variante resultou em uma troca de uma leucina para uma prolina, investigamos o fenótipo de todas as variantes do tipo *missense* que resultaram em troca para prolina na literatura e em bancos de dados.

Identificamos 15 variantes do tipo *missense* que resultou em uma mudança para prolina. Em 11 casos (73,5%), essa mudança resultou em CAIS. Todas as substituições por prolina no domínio de ligação ao ligante (LBD) causaram o fenótipo completo, enquanto variantes com substituições por prolina no domínio de ligação ao DNA (DBD) e no domínio N-terminal (NTD) resultaram em fenótipos tanto completos quanto parciais. Isso sugere que as substituições por prolina na região de ligação ao ligante podem ser mais deletérias, como observamos no caso da nossa casuística. O fenótipo foi CAIS em todos os casos de substituição de leucina por prolina, independentemente da localização da variante entre os domínios do AR.

Essa maior patogenicidade associada a troca por prolina talvez se deva por características intrínsecas ao aminoácido prolina. A prolina possui uma composição típica, caracterizada por uma cadeia lateral cíclica. Essa característica particular tem o potencial de perturbar o arranjo ordenado de estruturas secundárias, como as alfa-hélices, que dependem de um padrão consistente de ligação de hidrogênio entre os átomos da proteína⁵⁷.

Quando a prolina está posicionada no interior de uma alfa-hélice, ela confere uma curvatura à estrutura helicoidal, desestabilizando-a. Essa desestabilização ocorre devido à incapacidade da cadeia lateral cíclica da prolina de estabelecer as ligações de hidrogênio necessárias com os demais átomos, o que dificulta a manutenção da configuração convencional da alfa-hélice. Consequentemente, a presença de resíduos de prolina na região central de uma alfa-hélice pode induzir a desestabilização da estrutura geral da proteína⁵⁷. Portanto, variantes

que resultam na introdução de prolina podem afetar significativamente a estabilidade da proteína.

Considerando uma única variante em um gene pode ser patogênica e causar doença, é razoável presumir que mais de uma variante em um único gene também pode afetar o fenótipo. O impacto de variantes duplas no fenótipo foi demonstrado na cardiomiopatia hipertrófica ⁵⁸. Várias variantes duplas foram relatadas esta condição nos genes *MYBPC3* e *TNNT2* ^{58; 59}. Essas variantes duplas foram associadas a um fenótipo mais grave, incluindo início mais precoce da doença, hipertrofia ventricular esquerda mais severa, maior prevalência de insuficiência cardíaca avançada e maior risco de morte cardíaca súbita.

Da mesma forma que na AIS, variações genótipo-fenótipo são relatadas na fibrose cística, uma condição genética que resulta de mutações no gene *CFTR* ⁶⁰. Na fibrose cística, variantes no gene *CFTR* são divididas em variantes leves ou graves de acordo com a função residual do gene, a qual geralmente é estimada com base no fenótipo do paciente ⁶⁰. Por exemplo, duas variantes leves associadas à fibrose cística resultaram em fibrose cística grave (p.Arg347His-p.Asp979Ala) quando herdadas em combinação ⁶⁰. Curiosamente, cada variante no *CFTR* funcionou de maneira diferente: a primeira causou um canal de cloreto defeituoso, enquanto a segunda reduziu a proteína CFTR na membrana celular. A combinação de ambos levou a uma fibrose cística grave. Desde o primeiro relato de uma mutação dupla no gene *CFTR* em 2011, outras variantes alélicas duplas no gene *CFTR* foram relatadas, e seu impacto no fenótipo foi caracterizado. Coletivamente, esses exemplos reforçam a evidência de que mutações duplas podem afetar o fenótipo final.

Identificamos um paciente com AIS com duas variantes no *AR*. Ambas são variantes do tipo *missense* não sinônimas, e uma delas já foi relacionada anteriormente à PAIS ¹³. Portanto, é possível que ambas as variantes, ao trabalharem juntas, levaram ao fenótipo completo. Há um relato prévio de dupla variante no *AR* identificadas em dois gêmeos monozigóticos (p.Phe857Leu e p.Ser866Pro), ambas localizadas no éxon 7 do LBD do *AR* ⁶¹. Embora a posição 856 seja extremamente conservada e envolvida em interações hidrofóbicas com a hélice 9 do *AR*, a variante p.Phe857Leu não alterou a ligação ao andrógeno e mostrando uma afinidade de ligação semelhante ar *AR* selvagem. Por outro lado, a variante p.Ser866Pro alterou severamente a ligação ao andrógeno.

Curiosamente, apesar da raridade de duplas variantes no *AR* em AIS, elas têm sido frequentemente relatadas no câncer de próstata ⁶², o que pode sugerir que as duplas mutações no gene *AR* são mais comuns em AIS do que tem sido relatado. Isso pode ocorrer pois, em casos de variantes duplas, variantes classificadas como VUS ou como provavelmente benignas podem ser

descartadas, uma vez que a maioria das variantes detectadas em estudos genéticos é regularmente avaliada quanto ao seu potencial patogênico. Na AIS, também é possível que polimorfismos, como o número de repetições CAG, atuem em conjunto com uma variante alélica do AR, contribuindo para o fenótipo^{63; 64}. De fato, a contribuição de polimorfismos, além de uma variante *missense*, pode causar a deficiência da 5-alfa redutase tipo 2, outra condição 46,XY DSD, o que sugere que os polimorfismos podem desempenhar um papel patogênico quando associados a uma variante patogênica^{65; 66}. Outra possibilidade é a presença de variantes alélicas sinônimas que possam atuar em conjunto com uma variante *missense* não sinônima. As mutações sinônimas têm sido amplamente negligenciadas como causadoras de condições genéticas, em grande parte porque são consideradas irrelevantes devido à sua incapacidade de alterar os aminoácidos. No entanto, existem muitas maneiras pelas quais variantes sinônimas podem ser deletérias, principalmente pela capacidade de alterar *splicing*⁶⁷. De fato, variantes sinônimas no AR já foram relatadas na AIS, tanto em CAIS quanto em PAIS^{17; 68}.

Nas últimas décadas, a expansão no campo da biologia molecular tem trazido grandes avanços na genética humana e médica. Dentre esses avanços, está a identificação de variantes estruturais no genoma, como por exemplo, as Variações de Número de Cópias (do inglês, Copy number variations, CNVs). As CNVs são inserções e deleções em maior escala, e referem-se a diferenças na dosagem de segmentos genômicos, variando em tamanho de 1 kilobases (kb) para muitas megabases (Mb)⁶⁹. Até recentemente, a maioria dos estudos genômicos tinha como foco principal a descrição de pequenas variantes, como as SNVs (Variação de nucleotídeo único), já que a detecção de CNVs era mais limitada⁷⁰. No entanto, devido ao seu tamanho, as CNVs abrangem mais bases que as SNVs e podem ter maior influência na expressão e estrutura de um determinado gene⁷¹. CNVs tem sido raramente descritas em AIS. Na literatura, identificamos um total de quinze casos descritos de grandes CNVs (>1kb) no AR. Todas essas CNVs, exceto uma, causaram CAIS. A exceção foi a identificação de uma deleção de 9.4 kb, a qual compreendia todo o éxon 4 em um paciente com azoospermia e com genitália externa masculina descrita como normal⁷². Entre as demais CNVs reportadas em indivíduos com CAIS, elas compreendem deleções de toda a região exônica do AR, do éxon 2, do éxon 3, do éxon 4, do éxon 5 e deleção compreendendo os éxons 4 a 8, além de duplicação do éxon 2. Identificamos uma deleção compreendendo os éxons 6, 7 e 8 do AR, o que expande o número de grandes CNS causando AIS. Com esse novo achado, há dezesseis grandes CNVs identificadas em pacientes com AIS, o que é um número proporcionalmente bem inferior ao número de variantes de outra natureza identificadas no AR em pacientes com AIS.

Uma possível explicação para essa baixa frequência pode ser devido ao fato de que o diagnóstico molecular de AIS envolve primeiramente o sequenciamento direto de todos os oito éxons do gene *AR* ^{38; 48}. No entanto, mesmo utilizando uma abordagem de identificação de variantes do *AR* por meio de um painel de genes direcionado, uma CNV em uma paciente com quadro fenotípico de AIS não foi detectada, apesar do painel ter uma cobertura de 97% para o *AR* ³⁶. Tanto o sequenciamento por Sanger quanto o sequenciamento paralelo de painel de genes são mais eficientes na detecção de SNVs (variantes de nucleotídeo único) ou *small indels* (pequenas inserções e deleções) em vez de CNVs (variações no número de cópias). Uma alternativa seria a realização de *Whole Exome Sequencing* (WES) em combinação com pipelines de bioinformática adequados para detectar CNVs. Com esta abordagem, houve sucesso na detecção de uma grande CNV compreendendo o éxon 2 do *AR* em uma paciente com CAIS. Isso reforça que a ausência de variantes no *AR* em casos de suspeita de AIS quando realizados o sequenciamento do *AR* por Sanger o sequenciamento paralelo de genes não necessariamente exclui a possibilidade de CNVs e que metodologias adicionais podem auxiliar na identificação das alterações moleculares relacionadas ao *AR* nesses casos.

Indels são inserções ou deleções de um ou mais nucleotídeos na sequência do DNA, sendo consideradas *small indels* inserções ou deleções menores que 50pb. *Indels* são altamente abundantes nos genomas humanos, representando de 15% a 21% dos polimorfismos humanos ⁷³. Além de sua contribuição para variações genéticas em populações saudáveis, *small indels* deletérias têm sido cada vez mais implicadas na patogenicidade de numerosas doenças genéticas, afetando tanto regiões codificantes quanto não codificantes ^{74; 75}. Na AIS, as *indels* representam o segundo tipo mais prevalente de variantes genéticas identificadas no gene *AR* ^{9; 14}. Tyutyusheva et al. relataram 66 casos de CAIS e compararam os tipos de variantes genéticas identificadas com 2 bancos de dados genômicos ⁷⁶, corroborando que *small indels* no *AR* são o segundo tipo mais comum de variantes genéticas identificadas em AIS.

Nas regiões codificadoras, *indels* podem resultar em variantes do tipo FS ou NFS ⁷³. *Indels* NFS envolvem a inserção ou deleção de múltiplos de 3 pares de bases, mantendo inalterada a sequência de proteína restante. Por outro lado, *indels* do tipo FS alteram o quadro de leitura a partir do ponto de inserção ou deleção, levando a proteínas anômalas geralmente com término prematuro (71).

Em relação à frequência de *indels frameshift* (FS) versus *indels não frameshift* (NFS), há mais *indels* FS do que *indels* NFS no genoma humano (2). No entanto, dentro das regiões codificantes, a distribuição das *indels* é fortemente enriquecida por *indels* NFS (2). Essa observação provavelmente ocorre porque as proteínas anômalas resultantes de *indels* do tipo FS

podem estar sujeitas à degradação mediada por RNA, através de um mecanismo descrito como mRNA decay (70). Curiosamente, cerca de 80% das *small indels* em indivíduos com AIS que identificamos são *indels* do tipo FS, o que indica que essas *indels* podem escapar de certos mecanismos de reparo de DNA.

Em nossa análise, no que diz respeito à relação genótipo-fenótipo das *small indels* no gene *AR*, a maioria dos casos (95,1%) apresentou o fenótipo CAIS. Notavelmente, todas as *indels frameshift* (FS) identificadas em indivíduos com AIS resultaram em CAIS. Essa observação é intrigante, uma vez que tradicionalmente se considera que a AIS não apresenta uma forte correlação genótipo-fenótipo forte ¹⁰. A importância de compreender a relação genótipo-fenótipo na AIS é destacada pela possibilidade de confirmação genética do diagnóstico, possibilitando prognósticos mais precisos e facilitando a atribuição de sexo para indivíduos 46,XY com AIS, especialmente aqueles com PAIS ⁷⁷. Além disso, esse conhecimento auxilia no aconselhamento genético para mulheres carreadoras de variantes no gene *AR* (22).

Além disso, informações sobre a expressão fenotípica de variantes específicas do *AR* podem ser utilizadas para construir mapas de domínios funcionais dentro do *AR*. É importante notar que certas variantes alélicas do *AR* no banco de dados de mutações do *AR* podem levar a mais de um fenótipo de AIS, incluindo a maioria das variantes alélicas não sinônimas ⁹. As razões para isso ainda não estão claras, mas o mosaïcismo somático pode desempenhar um papel nesse fenômeno ⁷⁸. No entanto, o mosaïcismo somático no gene *AR* entre indivíduos com AIS foi raramente relatado, com apenas 14 casos relatados até o momento ⁷⁹. No entanto, nossos resultados sugerem uma relação genótipo-fenótipo bem definida em relação as *small indels* do tipo *frameshift* no gene *AR*, já que todos os casos identificados resultaram no fenótipo completo.

Identificamos indivíduos com MAIS que apresentavam *small indels* no *AR*. Todos os 4 casos apresentaram *small indels* localizadas em regiões repetitivas em tandem (26, 30, 68).

Especificamente, em 2 casos, as *small indels* ocorreram em uma sequência de 4 leucinas antes da repetição CAG do *AR* no exon 1 (68). A sequência de mRNA do *AR* que codifica essas 4 leucinas complementa a sequência de mRNA do *AR* que codifica a repetição CAG, formando uma estrutura de RNA estável. Curiosamente, uma dessas inserções de leucina (g.1283_1284insAGTTTGCTG) foi reportada como alterando a estrutura secundária da proteína *AR* (68). Outra evidência que sustenta o papel patogênico da alteração da sequência de poli-leucina do *AR* vêm de uma outra variante alélica não sinônima no *AR* que substitui uma glicina por uma leucina (c.173A>T; p.Gln58Leu), adicionando uma leucina extra à sequência de quatro leucinas e causando infertilidade masculina em 2 homens finlandeses não relacionados (69). No entanto, em todos os casos de indivíduos com MAIS apresentando *small indels* no gene *AR*, as

indels são do tipo não *frameshift* (NFS) e se localizam em regiões repetitivas em tandem. Esses relatos sugerem que pequenas *indels* em repetições em tandem dentro do gene AR podem levar a AIS completa ou leve. Ainda assim, o fenótipo final parece ser melhor definido pelo fato de o pequeno indel ser de mudança de quadro (FS) ou não de mudança de quadro (NFS), em vez da localização dos *indels* em repetições em tandem.

Embora pacientes com *small indels* na região codificadora do gene AR possam apresentar tanto fenótipo leve quanto completo de AIS, *small indels* são raras em casos de fenótipo parcial. Identificamos apenas uma *small indel* no gene AR (c.2074_2076del) em um indivíduo com PAIS. Esse caso foi reportado em uma família com 3 indivíduos com AIS. Curiosamente, os outros 2 membros da família apresentavam CAIS, demonstrando que variabilidade fenotípica pode ocorrer mesmo em AIS resultante de *small indels* no AR (22).

Na análise comparativa de frequências de *small indels* entre indivíduos com AIS e a população masculina controle (*comparative burden test*), observamos um enriquecimento significativo de *small indels*, especialmente as do tipo FS, entre os indivíduos com AIS. Além disso, nossa investigação revelou uma notável e significativa super-representação de deleções em indivíduos com AIS, sugerindo que pequenas deleções podem estar mais frequentemente associadas à patogenicidade em comparação com pequenas inserções. Além disso, observamos um enriquecimento significativo de *small indels* abrangendo entre 1 a 3 nucleotídeos em indivíduos com AIS, indicando que a maioria das *small indels* patogênicas no gene AR é causada por pequenas variações.

Quanto a localização das *small indels* em regiões codificadoras no genoma humano, enriquecimento de *small indels* nas regiões N-terminal ou C-terminal (64). No contexto da AIS, também observamos um padrão semelhante, pois a maioria das *small indels* estão localizadas no domínio N-terminal (NTD) e no domínio de ligação a ligantes (LBD), com uma preponderância particular no domínio NTD. Essa preponderância pode ser atribuída ao fato de que o NTD do AR contém 2 regiões longas de poli-repetições de trinucleotídeos (poli-Q e poli-G).

Repetições de trinucleotídeos são sequências repetitivas de 3 nucleotídeos encontradas em diferentes números de cópias no genoma humano (72). Assim como outras regiões repetitivas no genoma, as repetições de trinucleotídeos são propensas a eventos de *slippage* (73). Dado que o *slippage* tem sido implicado na geração de *indels*, repetições de trinucleotídeos dentro de regiões exônicas podem representar *hotspots* para a formação de *indels*. No gene AR, existem duas repetições de trinucleotídeos bem conhecidas. A primeira é a repetição CAG (trecho de poli-Q) localizada no exon 1 (abrangendo do aminoácido 57 ao 80; NM_000044.6), e a segunda é a repetição CCG (trecho de poli-G) também dentro do exon 1 (abrangendo do

aminoácido 451 ao 473). O gene *AR* tem uma alta taxa de variação individual a cada geração, tornando-o um excelente gene para analisar padrões de variações ao longo das gerações. Com base nisso, Montgomery et al. identificaram que repetições de trinucleotídeos longos dentro da região codificante do *AR* são *hotspots* para ocorrência de *small indels* em uma população saudável, explicando a natureza polimórfica dessas poli-repetições em humanos³².

No entanto, identificamos que apenas 15,8% das *small indels* patogênicas reportadas em indivíduos com AIS estão localizadas dentro de poli-repetições do *AR* (Figura 7; D). Isso sugere que os *hotspots* para *small indels* dentro do gene *AR* observados na população geral saudável não representam completamente os mecanismos subjacentes as *small indels* patogênicas identificados em indivíduos com AIS.

Além das clássicas sequências de poli-Q e poli-G, o gene *AR* contém mais 2 repetições de trinucleotídeos fixas (não polimórficas). Essas sequências são uma segunda sequência de poliglutamina (entre os aminoácidos 86 a 91) e uma sequência de oito prolina (entre os aminoácidos 374 a 381). Embora menores em tamanho, essas 2 poli-repetições ainda não foram extensivamente estudadas quanto aos seus papéis individuais na atividade do receptor de androgênio. No entanto, nossa análise revelou a presença de três *small indels* dentro da região de poliprolina e seis dentro da segunda sequência de poliglutamina no *AR*, o que sugere que estas sequências sejam *hotspots* para a ocorrência de *small indels* no *AR*.

Considerando a elevada prevalência de repetições de trinucleotídeos em genes humanos, compreender o impacto potencial dessas pequenas poli-repetições pode ser crucial para entender a patogenicidade das variantes identificadas nessas regiões e avançar nossa compreensão dos mecanismos de mutagênese ligados às *small indels*.

No que diz respeito ao mecanismo de mutagênese, o *slippage* da polimerase é considerado o mecanismo predominante para a geração de *small indels* no genoma (12, 74). Esse fenômeno ocorre durante a replicação e normalmente resulta na contração ou expansão de regiões de repetição em tandem, como as sequências de trinucleotídeos. No genoma humano, o *slippage* da polimerase explica aproximadamente 75% dos eventos em regiões de *hotspot* repetitivas e cerca de 50% das *indels* em áreas não *hotspot* (19). Notavelmente, cerca de 45% das *small indels* estão concentradas em apenas cerca de 4% do genoma e podem ser amplamente atribuídos ao *slippage* da polimerase (75). Sequências propensas ao *slippage* da polimerase, como regiões de repetição em tandem e sequências de homopolímeros, abrigam cerca de 70% das *indels* (12).

Como o *slippage* da polimerase ocorre frequentemente em sequências repetitivas, eventos CCC provavelmente são causados por esse mecanismo⁸⁰. Em nossa análise,

classificamos cada evento de *indel* como evento CCC ou não-CCC para estimar a proporção de *small indels* que poderiam estar ocorrendo em decorrência de slippage. Curiosamente, descobrimos que apenas 52% das pequenas *indels* do AR são eventos não-CCC, sugerindo que outros mecanismos moleculares podem desempenhar um papel na geração dessas *small indels*. Embora esses mecanismos ainda sejam amplamente desconhecidos, perturbações na clivagem do DNA podem estar envolvidas na geração de *indels*, pois as intensidades de clivagem do DNA próximas aos pontos de início e término de pequenas *indels* foram significativamente menores do que em outras regiões genômicas. No entanto, estudos adicionais são necessários para elucidar esses mecanismos moleculares e sua contribuição para eventos patogênicos de pequenas *indels* no gene *AR*.

Em uma busca por uma sequência de consenso genômico propensa à geração de *indels*, Krawczak e Cooper identificaram uma sequência consenso de 6 pares de bases (TG A/G A/G G/T A/C) que ocorre frequentemente em locais de pequenas deleções germinativas em várias doenças genéticas (76). Essa sequência consenso também foi identificada como um *hotspot* para a ocorrência de *small indels* no gene TP53 (77). Para determinar se as *small indels* patogênicas no gene *AR* ocorrem dentro dessa sequência consenso, analisamos o contexto genômico de cada *small indel* que identificamos no *AR* e comparamos a sequência de nucleotídeos que cada *small indel* no *AR* se localizava e comparamos com a sequência consenso. Identificamos três pequenas *indels* dentro da sequência consenso (c.2317_2321del, c.2356_2357del e c.2478_2479ins). No total, as *small indels* nesta sequência consenso compreenderam 3,65% de todas as *small indels* do *AR*, o que é semelhante à frequência de *small indels* nesta sequência consenso no gene TP53.

Homopolímeros ou homocopolímeros são sequências contínuas de purinas ou pirimidinas (por exemplo, AAAAAA, CCTTCC, AAAGGG) (77). Essas regiões foram relatadas como regiões de risco para a ocorrência de *small indels* (75). Por exemplo, no gene TP53, *small indels* são relativamente frequentes dentro de sequências homopoliméricas de 4 ou mais nucleotídeos (78). No *AR*, identificamos sete *small indels* em sequências homopoliméricas ou homocopoliméricas, incluindo a nova c.1314_1315delCT. No total, as *small indels* nessas sequências compreendem 8,5% dos *small indels* no *AR*, o que suporta ainda mais o fato de que essas sequências são regiões propensas para ocorrências de *small indels* dentro de regiões exônicas.

Small indels complexas envolvem uma combinação de deleção e inserção (e às vezes uma substituição de base) de uma só vez (79). Devido à falta de ferramentas eficazes para mapear e detectar *small indels* complexas, elas têm sido um tanto negligenciadas. Portanto, a frequência, características fenotípicas e mecanismos de mutagênese dessas *small indels*

complexas são amplamente desconhecidos (80). Essas variantes complexas são atribuídas a sequências palindrômicas ou quasi-palindrômicas (81). Identificamos uma nova *small indel* complexa no *AR* (c.431_452delinsAAA). Ao examinar todas as *small indels* complexas reportadas no *AR* em indivíduos com AIS, identificamos outras 4 ocorrências de tais variantes entre indivíduos com AIS (n = 5; 6%). Apenas uma *small indel* complexa *frameshift* (c.157_165dup) (68) foi encontrada dentro de uma repetição em tandem, enquanto as outras 4 estavam localizadas em sequências palindrômicas, sugerindo que essas sequências são propensas a gerar *small indels* complexas.

7 CONCLUSÕES

7 CONCLUSÕES

- Variantes no *AR* afetando a região AF2 causam um fenótipo parcial na maioria dos casos afetados.
- A maioria das variantes identificadas no gene *AR* em indivíduos com fenótipo de AIS é classificada como patogênica ou provavelmente patogênica.
- CNVs são raras entre as alterações moleculares que causam AIS, mas são fortemente relacionadas ao fenótipo completo.
- Há uma correlação genótipo-fenótipo em indivíduos que possuem variantes do tipo *small indels* no gene do *AR* visto que a imensa maioria das *small indels* no gene *AR* causam fenótipo completo.
- As regiões exônicas repetitivas do *AR*, como as sequências de homopolímeros e as repetições de trinucleotídeos são regiões propensas a ocorrência de *small indels*.
- *Small indels* do tipo *frameshift* no *AR* são deletérias e relacionadas a fenótipo completo em todos os casos, enquanto que *small indels* do tipo não-*frameshift* podem levar ao fenótipo leve quando localizadas em repetições de trinucleotídeos dentro da região exônica do *AR*.

REFERÊNCIAS

- ¹ VIRTANEN, H. E.; TOPPARI, J. Embryology and physiology of testicular development and descent. *Pediatr Endocrinol Rev*, v. 11 Suppl 2, p. 206-13, Feb 2014. ISSN 1565-4753 (Print) 1565-4753 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/24683945> >.
- ² LUCAS-HERALD, A. K.; BASHAMBOO, A. Gonadal development. *Endocr Dev*, v. 27, p. 1-16, 2014. ISSN 1662-2979 (Electronic) 1421-7082 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/25247640> >.
- ³ KIM, K. S.; KIM, J. Disorders of sex development. *Korean J Urol*, v. 53, n. 1, p. 1-8, Jan 2012. ISSN 2005-6745. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22323966> >.
- ⁴ PASK, A. The Reproductive System. *Adv Exp Med Biol*, v. 886, p. 1-12, 2016. ISSN 0065-2598 (Print) 0065-2598 (Linking). Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/26659484> >.
- ⁵ MENDONCA, B. B. et al. 46,XY DSD due to impaired androgen production. *Best Pract Res Clin Endocrinol Metab*, v. 24, n. 2, p. 243-62, Apr 2010. ISSN 1878-1594. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20541150> >.
- ⁶ RUSSELL, D. W.; WILSON, J. D. Chapter 4A - Steroid 5 α -Reductase 2 Deficiency. In: HAMMER, M. I. N. L. P. T. Y. W. O. M. D. (Ed.). *Genetic Steroid Disorders*. San Diego: Academic Press, 2014. p.199-214. ISBN 978-0-12-416006-4.
- ⁷ HUGHES, I. A. et al. Consensus statement on management of intersex disorders. *J Pediatr Urol*, v. 2, n. 3, p. 148-62, Jun 2006. ISSN 1873-4898. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/18947601> >.
- ⁸ LEE, P. A. et al. Global Disorders of Sex Development Update since 2006: Perceptions, Approach and Care. *Horm Res Paediatr*, v. 85, n. 3, p. 158-80, 2016. ISSN 1663-2826. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/26820577> >.
- ⁹ BATISTA, R. L. et al. Androgen insensitivity syndrome: a review. *Arch Endocrinol Metab*, v. 62, n. 2, p. 227-235, 2018 Mar-Apr 2018. ISSN 2359-4292. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/29768628> >.
- ¹⁰ HUGHES, I. A. et al. Androgen insensitivity syndrome. *Semin Reprod Med*, v. 30, n. 5, p. 432-42, Oct 2012. ISSN 1526-4564. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/23044881> >.
- ¹¹ BATISTA, R. L. et al. Mild Androgen Insensitivity Syndrome: The Current Landscape. *Endocr Pract*, v. 28, n. 9, p. 911-917, Sep 2022. ISSN 1530-891X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/35660466> >.
- ¹² HORNIG, N. C.; HOLTERHUS, P. M. Molecular basis of androgen insensitivity syndromes. *Mol Cell Endocrinol*, v. 523, p. 111146, 03 2021. ISSN 1872-8057. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/33385475> >.

- ¹³ GOTTLIEB, B. et al. The androgen receptor gene mutations database: 2012 update. *Hum Mutat*, v. 33, n. 5, p. 887-94, May 2012. ISSN 1098-1004. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22334387> >.
- ¹⁴ GULÍA, C. et al. Androgen insensitivity syndrome. *Eur Rev Med Pharmacol Sci*, v. 22, n. 12, p. 3873-3887, Jun 2018. ISSN 2284-0729. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/29949163> >.
- ¹⁵ MONGAN, N. P. et al. Androgen insensitivity syndrome. *Best Pract Res Clin Endocrinol Metab*, v. 29, n. 4, p. 569-80, Aug 2015. ISSN 1878-1594. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/26303084> >.
- ¹⁶ BATISTA, R. L. et al. A Recurrent Synonymous Mutation in the Human Androgen Receptor Gene Causing Complete Androgen Insensitivity Syndrome. *J Steroid Biochem Mol Biol*, Jul 2017. ISSN 1879-1220. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/28743543> >.
- ¹⁷ HELLWINKEL, O. J. et al. A unique exonic splicing mutation in the human androgen receptor gene indicates a physiologic relevance of regular androgen receptor transcript variants. *J Clin Endocrinol Metab*, v. 86, n. 6, p. 2569-75, Jun 2001. ISSN 0021-972X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/11397856> >.
- ¹⁸ DAVEY, R. A.; GROSSMANN, M. Androgen Receptor Structure, Function and Biology: From Bench to Bedside. *Clin Biochem Rev*, v. 37, n. 1, p. 3-15, Feb 2016. ISSN 0159-8090. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/27057074> >.
- ¹⁹ TAN, M. H. et al. Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacol Sin*, v. 36, n. 1, p. 3-23, Jan 2015. ISSN 1745-7254. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/24909511> >.
- ²⁰ SCHLANGER, S.; HEEMERS, H. V. Functional Studies on Steroid Receptors. *Methods Mol Biol*, v. 1786, p. 117-130, 2018. ISSN 1940-6029. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/29786790> >.
- ²¹ BENNETT, N. C. et al. Molecular cell biology of androgen receptor signalling. *Int J Biochem Cell Biol*, v. 42, n. 6, p. 813-27, Jun 2010. ISSN 1878-5875. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19931639> >.
- ²² NADAL, M. et al. Structure of the homodimeric androgen receptor ligand-binding domain. *Nat Commun*, v. 8, p. 14388, Feb 2017. ISSN 2041-1723. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/28165461> >.
- ²³ CLINCKEMALIE, L. et al. The hinge region in androgen receptor control. *Mol Cell Endocrinol*, v. 358, n. 1, p. 1-8, Jul 2012. ISSN 1872-8057. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22406839> >.
- ²⁴ BUCHANAN, G. et al. Corepressor effect on androgen receptor activity varies with the length of the CAG encoded polyglutamine repeat and is dependent on receptor/corepressor ratio in prostate cancer cells. *Mol Cell Endocrinol*, v. 342, n. 1-2, p. 20-31, Aug 2011. ISSN 1872-8057. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/21664238> >.

- ²⁵ JIN, Y. et al. Communication between the Ligand-Binding Pocket and the Activation Function-2 Domain of Androgen Receptor Revealed by Molecular Dynamics Simulations. *J Chem Inf Model*, v. 59, n. 2, p. 842-857, 02 2019. ISSN 1549-960X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/30658039> >.
- ²⁶ LIU, Y. et al. Structural Based Screening of Antiandrogen Targeting Activation Function-2 Binding Site. *Front Pharmacol*, v. 9, p. 1419, 2018. ISSN 1663-9812. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/30555332> >.
- ²⁷ SILVA, J. M. et al. Androgen receptor mRNA analysis from whole blood: a low-cost strategy for detection of androgen receptor gene splicing defects. *Clin Genet*, Sep 2018. ISSN 1399-0004. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/30193409> >.
- ²⁸ BATISTA, R. L. et al. Mobile DNA in Endocrinology: LINE-1 retrotransposon causing Partial Androgen Insensitivity Syndrome. *J Clin Endocrinol Metab*, Aug 2019. ISSN 1945-7197. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/31393562> >.
- ²⁹ RICHARDS, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*, v. 17, n. 5, p. 405-24, May 2015. ISSN 1530-0366. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/25741868> >.
- ³⁰ IOANNIDIS, N. M. et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet*, v. 99, n. 4, p. 877-885, Oct 06 2016. ISSN 1537-6605. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/27666373> >.
- ³¹ GHOSH, R.; OAK, N.; PLON, S. E. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biol*, v. 18, n. 1, p. 225, Nov 28 2017. ISSN 1474-760X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/29179779> >.
- ³² MONTGOMERY, S. B. et al. The origin, evolution, and functional impact of short insertion-deletion variants identified in 179 human genomes. *Genome Res*, v. 23, n. 5, p. 749-61, May 2013. ISSN 1549-5469. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/23478400> >.
- ³³ HIORT, O. et al. The clinical and molecular spectrum of androgen insensitivity syndromes. *Am J Med Genet*, v. 63, n. 1, p. 218-22, May 03 1996. ISSN 0148-7299. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/8723113> >.
- ³⁴ RAJENDER, S. et al. L712V mutation in the androgen receptor gene causes complete androgen insensitivity syndrome due to severe loss of androgen function. *Steroids*, v. 78, n. 12-13, p. 1288-92, Dec 11 2013. ISSN 1878-5867. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/24055831> >.
- ³⁵ QUIGLEY, C. A. et al. Androgen receptor defects: historical, clinical, and molecular perspectives. *Endocr Rev*, v. 16, n. 3, p. 271-321, Jun 1995. ISSN 0163-769X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/7671849> >.

- ³⁶ EGGERS, S. et al. Disorders of sex development: insights from targeted gene sequencing of a large international patient cohort. *Genome Biol*, v. 17, n. 1, p. 243, 11 2016. ISSN 1474-760X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/27899157> >.
- ³⁷ ELFFERICH, P. et al. Functional analysis of novel androgen receptor mutations in a unique cohort of Indonesian patients with a disorder of sex development. *Sex Dev*, v. 3, n. 5, p. 237-44, 2009. ISSN 1661-5433. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19851057> >.
- ³⁸ MELO, K. F. et al. Clinical, hormonal, behavioral, and genetic characteristics of androgen insensitivity syndrome in a Brazilian cohort: five novel mutations in the androgen receptor gene. *J Clin Endocrinol Metab*, v. 88, n. 7, p. 3241-50, Jul 2003. ISSN 0021-972X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/12843171> >.
- ³⁹ ITTIWUT, C. et al. Novel mutations of the SRD5A2 and AR genes in Thai patients with 46, XY disorders of sex development. *J Pediatr Endocrinol Metab*, v. 30, n. 1, p. 19-26, Jan 2017. ISSN 2191-0251. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/27849622> >.
- ⁴⁰ HIORT, O. et al. Inherited and de novo androgen receptor gene mutations: investigation of single-case families. *J Pediatr*, v. 132, n. 6, p. 939-43, Jun 1998. ISSN 0022-3476. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/9627582> >.
- ⁴¹ BARBARO, M. et al. Complete androgen insensitivity without Wolffian duct development: the AR-A form of the androgen receptor is not sufficient for male genital development. *Clin Endocrinol (Oxf)*, v. 66, n. 6, p. 822-6, Jun 2007. ISSN 0300-0664. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/17408421> >.
- ⁴² LOTTRUP, G. et al. Identification of a novel androgen receptor mutation in a family with multiple components compatible with the testicular dysgenesis syndrome. *J Clin Endocrinol Metab*, v. 98, n. 6, p. 2223-9, Jun 2013. ISSN 1945-7197. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/23589523> >.
- ⁴³ LEDIG, S. et al. Novel and recurrent mutations in patients with androgen insensitivity syndromes. *Horm Res*, v. 63, n. 6, p. 263-9, 2005. ISSN 0301-0163. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15925895> >.
- ⁴⁴ LUNDBERG GIWERCMAN, Y. et al. Functional characterisation of mutations in the ligand-binding domain of the androgen receptor gene in patients with androgen insensitivity syndrome. *Hum Genet*, v. 103, n. 4, p. 529-31, Oct 1998. ISSN 0340-6717. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/9856504> >.
- ⁴⁵ WANG, Y. et al. AR mutations in 28 patients with androgen insensitivity syndrome (Prader grade 0-3). *Sci China Life Sci*, v. 60, n. 7, p. 700-706, Jul 2017. ISSN 1869-1889. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/28624954> >.
- ⁴⁶ HELLMANN, P. et al. Male patients with partial androgen insensitivity syndrome: a longitudinal follow-up of growth, reproductive hormones and the development of gynaecomastia. *Arch Dis Child*, v. 97, n. 5, p. 403-9, May 2012. ISSN 1468-2044. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22412043> >.

- 47 TANAKA, H. et al. One additional mutation at exon A amplifies thermolability of androgen receptor in a case with complete androgen insensitivity syndrome. *Gynecol Endocrinol*, v. 12, n. 2, p. 75-82, Apr 1998. ISSN 0951-3590. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/9610419> >.
- 48 AUDI, L. et al. Novel (60%) and recurrent (40%) androgen receptor gene mutations in a series of 59 patients with a 46,XY disorder of sex development. *J Clin Endocrinol Metab*, v. 95, n. 4, p. 1876-88, Apr 2010. ISSN 1945-7197. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20150575> >.
- 49 LOBACCARO, J. M. et al. Molecular modeling and in vitro investigations of the human androgen receptor DNA-binding domain: application for the study of two mutations. *Mol Cell Endocrinol*, v. 116, n. 2, p. 137-47, Feb 05 1996. ISSN 0303-7207. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/8647313> >.
- 50 MARCELLI, M. et al. A mutation in the DNA-binding domain of the androgen receptor gene causes complete testicular feminization in a patient with receptor-positive androgen resistance. *J Clin Invest*, v. 87, n. 3, p. 1123-6, Mar 1991. ISSN 0021-9738. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/1999491> >.
- 51 BELSHAM, D. D. et al. Leu-676-Pro mutation of the androgen receptor causes complete androgen insensitivity syndrome in a large Hutterite kindred. *Hum Mutat*, v. 5, n. 1, p. 28-33, 1995. ISSN 1059-7794. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/7537149> >.
- 52 ROSA, S. et al. Complete androgen insensitivity syndrome caused by a novel mutation in the ligand-binding domain of the androgen receptor: functional characterization. *J Clin Endocrinol Metab*, v. 87, n. 9, p. 4378-82, Sep 2002. ISSN 0021-972X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/12213902> >.
- 53 AHMED, S. F. et al. Phenotypic features, androgen receptor binding, and mutational analysis in 278 clinical cases reported as androgen insensitivity syndrome. *J Clin Endocrinol Metab*, v. 85, n. 2, p. 658-65, Feb 2000. ISSN 0021-972X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/10690872> >.
- 54 RAICU, F. et al. Novel mutation in the ligand-binding domain of the androgen receptor gene (1790p) associated with complete androgen insensitivity syndrome. *Asian J Androl*, v. 10, n. 4, p. 687-91, Jul 2008. ISSN 1008-682X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/18097502> >.
- 55 JÄÄSKELÄINEN, J. et al. Five novel androgen receptor gene mutations associated with complete androgen insensitivity syndrome. *Hum Mutat*, v. 27, n. 3, p. 291, Mar 2006. ISSN 1098-1004. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/16470553> >.
- 56 GALANI, A. et al. Sex-reversed phenotype in association with two novel mutations c.2494delA and c.T3004C in the ligand-binding domain of the androgen receptor gene. *Fertil Steril*, v. 90, n. 5, p. 2008.e1-4, Nov 2008. ISSN 1556-5653. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/18342859> >.

- ⁵⁷ BAJAJ, K. et al. Stereochemical criteria for prediction of the effects of proline mutations on protein stability. *PLoS Comput Biol*, v. 3, n. 12, p. e241, Dec 2007. ISSN 1553-7358. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/18069886> >.
- ⁵⁸ FOUREY, D. et al. Prevalence and Clinical Implication of Double Mutations in Hypertrophic Cardiomyopathy: Revisiting the Gene-Dose Effect. *Circ Cardiovasc Genet*, v. 10, n. 2, Apr 2017. ISSN 1942-3268. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/28420666> >.
- ⁵⁹ MILLER, T. E. et al. Whole blood RNA offers a rapid, comprehensive approach to genetic diagnosis of cardiovascular diseases. *Genet Med*, v. 9, n. 1, p. 23-33, Jan 2007. ISSN 1098-3600. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/17224687> >.
- ⁶⁰ POLIZZI, A. et al. Genotype-phenotype correlation in cystic fibrosis patients bearing [H939R;H949L] allele. *Genet Mol Biol*, v. 34, n. 3, p. 416-20, Jul 2011. ISSN 1678-4685. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/21931512> >.
- ⁶¹ MONGAN, N. P. et al. Two de novo mutations in the AR gene cause the complete androgen insensitivity syndrome in a pair of monozygotic twins. *J Clin Endocrinol Metab*, v. 87, n. 3, p. 1057-61, Mar 2002. ISSN 0021-972X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/11889162> >.
- ⁶² TILLEY, W. D. et al. Mutations in the androgen receptor gene are associated with progression of human prostate cancer to androgen independence. *Clin Cancer Res*, v. 2, n. 2, p. 277-85, Feb 1996. ISSN 1078-0432. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/9816170> >.
- ⁶³ PARADA-BUSTAMANTE, A. et al. Androgen receptor CAG and GGN polymorphisms in boys with isolated hypospadias. *J Pediatr Endocrinol Metab*, v. 25, n. 1-2, p. 157-62, 2012. ISSN 0334-018X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22570967> >.
- ⁶⁴ HUANG, G. et al. Androgen receptor gene CAG repeat polymorphism and risk of isolated hypospadias: results from a meta-analysis. *Genet Mol Res*, v. 14, n. 1, p. 1580-8, Mar 2015. ISSN 1676-5680. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/25867301> >.
- ⁶⁵ BATISTA, R. L.; MENDONÇA, B. B. Integrative and Analytical Review of the 5-Alpha-Reductase Type 2 Deficiency Worldwide. *Appl Clin Genet*, v. 13, p. 83-96, 2020. ISSN 1178-704X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/32346305> >.
- ⁶⁶ FERNÁNDEZ-CANCIO, M. et al. SRD5A2 gene mutations and polymorphisms in Spanish 46,XY patients with a disorder of sex differentiation. *Int J Androl*, v. 34, n. 6 Pt 2, p. e526-35, Dec 2011. ISSN 1365-2605. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/21631525> >.
- ⁶⁷ ANNA, A.; MONIKA, G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet*, Apr 2018. ISSN 2190-3883. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/29680930> >.

- ⁶⁸ BATISTA, R. L. et al. A recurrent synonymous mutation in the human androgen receptor gene causing complete androgen insensitivity syndrome. *J Steroid Biochem Mol Biol*, v. 174, p. 14-16, Nov 2017. ISSN 1879-1220. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/28743543> >.
- ⁶⁹ PÖS, O. et al. DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects. *Biomed J*, v. 44, n. 5, p. 548-559, Oct 2021. ISSN 2320-2890. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/34649833> >.
- ⁷⁰ ZHANG, F. et al. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet*, v. 10, p. 451-81, 2009. ISSN 1545-293X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19715442> >.
- ⁷¹ STANKIEWICZ, P.; LUPSKI, J. R. Structural variation in the human genome and its role in disease. *Annu Rev Med*, v. 61, p. 437-55, 2010. ISSN 1545-326X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20059347> >.
- ⁷² AKIN, J. W. et al. Evidence for a partial deletion in the androgen receptor gene in a phenotypic male with azoospermia. *Am J Obstet Gynecol*, v. 165, n. 6 Pt 1, p. 1891-4, Dec 1991. ISSN 0002-9378. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/1750490> >.
- ⁷³ LIN, M. et al. Effects of short indels on protein structure and function in human genomes. *Sci Rep*, v. 7, n. 1, p. 9313, Aug 24 2017. ISSN 2045-2322. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/28839204> >.
- ⁷⁴ DE LA CHAUX, N.; MESSER, P. W.; ARNDT, P. F. DNA indels in coding regions reveal selective constraints on protein evolution in the human lineage. *BMC Evol Biol*, v. 7, p. 191, Oct 2007. ISSN 1471-2148. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/17935613> >.
- ⁷⁵ MULLANEY, J. M. et al. Small insertions and deletions (INDELs) in human genomes. *Hum Mol Genet*, v. 19, n. R2, p. R131-6, Oct 15 2010. ISSN 1460-2083. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20858594> >.
- ⁷⁶ TYUTYUSHEVA, N. et al. Complete Androgen Insensitivity Syndrome: From Bench to Bed. *Int J Mol Sci*, v. 22, n. 3, Jan 27 2021. ISSN 1422-0067. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/33514065> >.
- ⁷⁷ LUCAS-HERALD, A. et al. The Long Term Outcome Of Boys With Partial Androgen Insensitivity Syndrome And A Mutation In The Androgen Receptor Gene. *J Clin Endocrinol Metab*, p. jc20161372, Jul 2016. ISSN 1945-7197. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/27403927> >.
- ⁷⁸ KÖHLER, B. et al. Androgen insensitivity syndrome: somatic mosaicism of the androgen receptor in seven families and consequences for sex assignment and genetic counseling. *J Clin Endocrinol Metab*, v. 90, n. 1, p. 106-11, Jan 2005. ISSN 0021-972X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15522944> >.

- ⁷⁹ BATISTA, R. L. et al. Partial androgen insensitivity syndrome due to somatic mosaicism of the androgen receptor. *J Pediatr Endocrinol Metab*, v. 31, n. 2, p. 223-228, Jan 2018. ISSN 2191-0251. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/29267169> >.
- ⁸⁰ XU, P. et al. Dynamics of strand slippage in DNA hairpins formed by CAG repeats: roles of sequence parity and trinucleotide interrupts. *Nucleic Acids Res*, v. 48, n. 5, p. 2232-2245, Mar 18 2020. ISSN 1362-4962. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/31974547> >.
- ⁸¹ TAYLOR, M. S.; PONTING, C. P.; COPLEY, R. R. Occurrence and consequences of coding sequence insertions and deletions in Mammalian genomes. *Genome Res*, v. 14, n. 4, p. 555-66, Apr 2004. ISSN 1088-9051. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15059996> >.
- ⁸² JUNG, S. C.; LEE, K. S.; KOO, S. K. Human gene mutations. Gene symbol: AR. Disease: Androgen insensitivity syndrome. *Hum Genet*, v. 114, n. 2, p. 222, Jan 2004. ISSN 0340-6717. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15046097> >.
- ⁸³ DECAESTECKER, K. et al. A novel mutation c.118delA in exon 1 of the androgen receptor gene resulting in complete androgen insensitivity syndrome within a large family. *Fertil Steril*, v. 89, n. 5, p. 1260.e3-1260.e7, May 2008. ISSN 1556-5653. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/17714709> >.
- ⁸⁴ BOEHMER, A. L. et al. Genotype versus phenotype in families with androgen insensitivity syndrome. *J Clin Endocrinol Metab*, v. 86, n. 9, p. 4151-60, Sep 2001. ISSN 0021-972X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/11549642> >.
- ⁸⁵ SINNECKER, G. H. et al. Phenotypic classification of male pseudohermaphroditism due to steroid 5 alpha-reductase 2 deficiency. *Am J Med Genet*, v. 63, n. 1, p. 223-30, May 1996. ISSN 0148-7299. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/8723114> >.
- ⁸⁶ CHENG, T. et al. Identification of three novel. *Asian J Androl*, Apr 2019. ISSN 1745-7262. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/31031332> >.
- ⁸⁷ PHILIBERT, P. et al. Complete androgen insensitivity syndrome is frequently due to premature stop codons in exon 1 of the androgen receptor gene: an international collaborative report of 13 new mutations. *Fertil Steril*, v. 94, n. 2, p. 472-6, Jul 2010. ISSN 1556-5653. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19463997> >.
- ⁸⁸ FERLIN, A. et al. Male infertility and androgen receptor gene mutations: clinical features and identification of seven novel mutations. *Clin Endocrinol (Oxf)*, v. 65, n. 5, p. 606-10, Nov 2006. ISSN 0300-0664. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/17054461> >.
- ⁸⁹ ZHU, Y. S. et al. A novel mutation in the CAG triplet region of exon 1 of androgen receptor gene causes complete androgen insensitivity syndrome in a large kindred. *J Clin Endocrinol Metab*, v. 84, n. 5, p. 1590-4, May 1999. ISSN 0021-972X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/10323385> >.

- ⁹⁰ GOTTLIEB, B. et al. Androgen insensitivity. *Am J Med Genet*, v. 89, n. 4, p. 210-7, Dec 29 1999. ISSN 0148-7299. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/10727996> >.
- ⁹¹ ROCCA, M. S. et al. Mutational screening of androgen receptor gene in 8224 men of infertile couples. *J Clin Endocrinol Metab*, Nov 17 2022. ISSN 1945-7197. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/36394509> >.
- ⁹² ZHU, H. et al. Phenotypic and biochemical characteristics and molecular basis in 36 Chinese patients with androgen receptor variants. *Orphanet J Rare Dis*, v. 16, n. 1, p. 122, Mar 09 2021. ISSN 1750-1172. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/33750429> >.
- ⁹³ BATCH, J. A. et al. Androgen receptor gene mutations identified by SSCP in fourteen subjects with androgen insensitivity syndrome. *Hum Mol Genet*, v. 1, n. 7, p. 497-503, Oct 1992. ISSN 0964-6906. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/1307250> >.
- ⁹⁴ JESKE, Y. W. et al. Androgen receptor genotyping in a large Australasian cohort with androgen insensitivity syndrome; identification of four novel mutations. *J Pediatr Endocrinol Metab*, v. 20, n. 8, p. 893-908, Aug 2007. ISSN 0334-018X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/17937062> >.
- ⁹⁵ HIORT, O. et al. Detection of point mutations in the androgen receptor gene using non-isotopic single strand conformation polymorphism analysis. German Collaborative Intersex Study Group. *Hum Mol Genet*, v. 3, n. 7, p. 1163-6, Jul 1994. ISSN 0964-6906. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/7981687> >.
- ⁹⁶ BRÜGGENWIRTH, H. T. et al. Molecular basis of androgen insensitivity. *J Steroid Biochem Mol Biol*, v. 58, n. 5-6, p. 569-75, Aug 1996. ISSN 0960-0760. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/8918984> >.
- ⁹⁷ LIU, Q.; YIN, X.; LI, P. Clinical characteristics, AR gene variants, and functional domains in 64 patients with androgen insensitivity syndrome. *J Endocrinol Invest*, v. 46, n. 1, p. 151-158, Jan 2023. ISSN 1720-8386. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/35974208> >.
- ⁹⁸ THIELE, B. et al. Complete androgen insensitivity caused by a new frameshift deletion of two base pairs in exon 1 of the human androgen receptor gene. *J Clin Endocrinol Metab*, v. 84, n. 5, p. 1751-3, May 1999. ISSN 0021-972X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/10323411> >.
- ⁹⁹ HE, X. et al. Case report: Identification of a frameshift mutation in GC enrichment and the GCC repeat region of the androgen insensitivity receptor (AR) gene in a patient with complete androgen insensitivity syndrome by whole-exome sequencing (WES) combined with specific PCR and deep sequencing. *Front Genet*, v. 13, p. 1038997, 2022. ISSN 1664-8021. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/36506311> >.
- ¹⁰⁰ IMAI, A. et al. A frame-shift mutation of the androgen receptor gene in a patient with receptor-negative complete testicular feminization: comparison with a single base substitution in

a receptor-reduced incomplete form. *Ann Clin Biochem*, v. 32 (Pt 5), p. 482-6, Sep 1995. ISSN 0004-5632. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/8830623> >.

¹⁰¹ BEITEL, L. K. et al. Complete androgen insensitivity due to mutations in the probable alpha-helical segments of the DNA-binding domain in the human androgen receptor. *Hum Mol Genet*, v. 3, n. 1, p. 21-7, Jan 1994. ISSN 0964-6906. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/8162033> >.

¹⁰² CHEN, C. P. et al. A frame shift mutation in the DNA-binding domain of the androgen receptor gene associated with complete androgen insensitivity, persistent müllerian structures, and germ cell tumors in dysgenetic gonads. *Fertil Steril*, v. 72, n. 1, p. 170-3, Jul 1999. ISSN 0015-0282. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/10428170> >.

¹⁰³ YUE, L. et al. [A novel deletion mutation in AR gene causes complete androgen insensitivity syndrome in a Chinese family]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi*, v. 27, n. 6, p. 631-3, Dec 2010. ISSN 1003-9406. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/21154321> >.

¹⁰⁴ RONG, H. L.; SUZUKI, N.; IMAI, A. A double nucleotide insertion-induced frame-shift mutation of the androgen receptor gene in a familial complete androgen insensitivity syndrome. *Eur J Obstet Gynecol Reprod Biol*, v. 148, n. 1, p. 53-5, Jan 2010. ISSN 1872-7654. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19815331> >.

¹⁰⁵ LOBACCARO, J. M. et al. Complete androgen insensitivity syndrome due to a new frameshift deletion in exon 4 of the androgen receptor gene: functional analysis of the mutant receptor. *Mol Cell Endocrinol*, v. 111, n. 1, p. 21-8, Apr 28 1995. ISSN 0303-7207. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/7649349> >.

¹⁰⁶ AVILA, D. M. et al. Immunoreactive AR and genetic alterations in subjects with androgen resistance and undetectable AR levels in genital skin fibroblast ligand-binding assays. *J Clin Endocrinol Metab*, v. 87, n. 1, p. 182-8, Jan 2002. ISSN 0021-972X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/11788645> >.

¹⁰⁷ HANNEMA, S. E. et al. Residual activity of mutant androgen receptors explains wolffian duct development in the complete androgen insensitivity syndrome. *J Clin Endocrinol Metab*, v. 89, n. 11, p. 5815-22, Nov 2004. ISSN 0021-972X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15531547> >.

¹⁰⁸ BALDAZZI, L. et al. Two mutations causing complete androgen insensitivity: a frame-shift in the steroid binding domain and a Cys-->Phe substitution in the second zinc finger of the androgen receptor. *Hum Mol Genet*, v. 3, n. 7, p. 1169-70, Jul 1994. ISSN 0964-6906. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/7981689> >.

¹⁰⁹ MACLEAN, H. E. et al. Novel androgen receptor gene mutations in Australian patients with complete androgen insensitivity syndrome. *Hum Mutat*, v. 23, n. 3, p. 287, Mar 2004. ISSN 1098-1004. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/14974091> >.

¹¹⁰ MUELLER, O. T.; ROOT, A. Gene symbol: AR. Disease: androgen insensitivity syndrome. *Hum Genet*, v. 119, n. 6, p. 672, Jul 2006. ISSN 0340-6717. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/17128455> >.

- ¹¹¹ TUREK-PLEWA, J.; JERZY, S.; WIESLAW, H. T. Novel human pathological mutations. Gene symbol: AR. Disease: androgen insensitivity syndrome. Hum Genet, v. 125, n. 3, p. 341, Apr 2009. ISSN 1432-1203. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/19309788>>.
- ¹¹² VILCHIS, F. et al. Extreme androgen resistance in a kindred with a novel insertion/deletion mutation in exon 5 of the androgen receptor gene. J Hum Genet, v. 48, n. 7, p. 346-51, 2003. ISSN 1434-5161. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/12908100>>.
- ¹¹³ WERNER, R. et al. In-vitro characterization of androgen receptor mutations associated with complete androgen insensitivity syndrome reveals distinct functional deficits. Sex Dev, v. 2, n. 2, p. 73-83, 2008. ISSN 1661-5433. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/18577874>>.
- ¹¹⁴ LIU, S. et al. Severe forms of complete androgen insensitivity syndrome caused by a p.Q65X novel mutation in androgen receptor: Clinical manifestations, imaging findings and molecular genetics. Steroids, v. 144, p. 47-51, Apr 2019. ISSN 1878-5867. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/30742848>>.

ANEXO 1 - Termo de Consentimento Livre e Esclarecido.

HOSPITAL DAS CLÍNICAS DA FACULDADE DE MEDICINA DA UNIVERSIDADE DE SÃO PAULO-HCFMUSP

DADOS SOBRE A PESQUISA

Expansão do Diagnóstico Molecular e das Possibilidades Terapêuticas da Síndrome de Insensibilidade aos Andrógenos

PESQUISADOR PRINCIPAL:

CARGO/FUNÇÃO:

INSCRIÇÃO CONSELHO REGIONAL

UNIDADE DO HCFMUSP: Endocrinologia

DEPARTAMENTO/INSTITUTO: Departamento de Clínica Médica, Disciplina de Endocrinologia, Unidade do Desenvolvimento; LIM (Laboratório de Investigações Médicas 42 – LIM 42)

Convidamos o(a) sr(a) para participar desta pesquisa que visa avaliar mutações no gene do Receptor Androgênico em pacientes com suspeita de Síndrome de Insensibilidade Androgênica. Considerando que a identificação das alterações moleculares podem aumentar o conhecimento sobre esta síndrome.

Para a realização deste estudo, os pacientes com este diagnóstico tratados no ambulatório de Desenvolvimento do Hospital das Clínicas da Universidade de São Paulo que aceitarem participar, serão submetidos a coletar amostra sanguínea para estudo molecular.

Os benefícios de um maior conhecimento sobre as alterações moleculares relacionadas a este gene estendem-se não apenas a indivíduos com Síndrome de Insensibilidade Androgênica mas também a indivíduos com Distúrbios da Diferenciação Sexual.

Ao participante é garantido completo SIGILO e PRIVACIDADE, sendo qualquer dado avaliado como um todo e não individualmente. A qualquer momento, o participante pode manifestar o desejo de retirar-se desta pesquisa, desejo ao qual será prontamente atendido. Caso concorde em participar, uma via deste termo do Consentimento Livre e Esclarecido será fornecida ao participante.

Esta pesquisa não prevê custos aos participantes, visto que a avaliação nos que foram participantes será feita em paralelo às suas consultas de rotina, sem nenhuma chamada adicional.

Em qualquer etapa do estudo, você terá acesso aos profissionais responsáveis pela pesquisa para esclarecimento de eventuais dúvidas no endereço: LIM 42 do Hospital das Clínicas da Universidade de São Paulo, localizado na rua Eneas de Carvalho Aguiar, 255 São Paulo/SP. Telefone(s) 11 26617512. Se você tiver alguma consideração ou dúvida sobre a ética da pesquisa, entre em contato com o Comitê de Ética em Pesquisa (CEP) – Rua Ovídio Pires de Campos, 225 – 5o andar – tel: (11) 2661-7585, (11) 2661-1548, (11) 2661-1549 – e-mail: cappesq.adm@hc.fm.usp.br

ANEXO 2 – Tabela com as variantes encontradas em pacientes com AIS em nossa coorte e reclassificação.

Tabela 8 - Variantes encontradas em pacientes com AIS em nossa coorte e reclassificação (continua).

Número de variantes	Fenótipo	Posição codon	Posição proteína	Tipo	AR exon	Domínio	Publicada	ACMG	Revel	Meta LR
1	CAIS	c.431_452delinsAAA	p.Gly144Glufs*25	Nonsense	1	1	nova	Provavelmente Patogênica	-	-
1	CAIS	c.387_388del	p.Gly130Leufs*27	Nonsense	1	1	nova	Provavelmente Patogênica	-	-
1	CAIS	c.1314_1315delCT	p.Phe439His*62	Nonsense	1	1	nova	Provavelmente Patogênica	-	-
1	CAIS	c.175C>T	p.Gln59Term	Nonsense	1	1	nova	Provavelmente Patogênica	-	-
1	CAIS	c.217 C>T	p.Gln73Term	Nonsense	1	1	nova	Provavelmente Patogênica	-	-
2	CAIS	c.262C>T	p.Gln88Term	Nonsense	1	1	nova	Provavelmente Patogênica	-	-
2	CAIS	c.362C>A	p.Ser121Term	Nonsense	1	1	nova	Provavelmente Patogênica	-	-
2	CAIS	c.382G>T	p.Glu128Term	Nonsense	1	1	nova	Provavelmente Patogênica	-	-
5	CAIS	c.1530C>T	p.Ser510Ser	Sinônima	1	1	nova	VUS	-	-
1	CAIS	Grande deleção	Exon 6,7,8 del	CNV		3	Nova	Patogênica		
1	CAIS	c.1644G>T	p.Leu548Phe	Missense	2	2	Publicada	Provavelmente Patogênica	Uncertain (0.58)	Benign (low) (0.38)
1	CAIS	c.1769-2A>G	Splice	Splice	3	2	Publicada	Provavelmente Patogênica	-	-
2	CAIS	c.1769-1 G>C	Splice	Splice	3	2	Publicada	Provavelmente Patogênica	-	-
5	CAIS	c.1847 G>A	p.Arg616His	Missense	3	2	Publicada	Provavelmente Patogênica	Deleterious (0.98)	(Strong) Deleterious (0.98)
2	CAIS	c.2117A>G	p.Asn706Ser	Missense	4	3	Publicada	Provavelmente Patogênica	Deleterious (0.85)	(Moderate) Deleterious (0.98)
1	CAIS	c.2155T>G	p.Trp719Gly	Missense	4	3	Publicada	Provavelmente Patogênica	Deleterious (0.98)	(Strong) Deleterious (1)
1	CAIS	c.2138 T>C	p.Leu713Pro	Missense	4	3	Publicada	Provavelmente Patogênica	Deleterious (0.98)	(Strong) Deleterious (0.96)
1	CAIS	c.2257C>T	p.Arg753Term	Nonsense	5	3	Publicada	Provavelmente Patogênica	-	-

Tabela 8 - Variantes encontradas em pacientes com AIS em nossa coorte e reclassificação (continuação).

Número de variantes	Fenótipo	Posição codon	Posição proteína	Tipo	AR exon	Domínio	Publicada	ACMG	Revel	Meta LR
4	CAIS	c.2301delT	p.Asp768Ilefs*21	Nonsense	5	3	Publicada	Patogênica	-	-
1	CAIS	c.2305C>G	p.Leu769Val	Missense	5	3	Publicada	Provavelmente Patogênica	Deleterious (0.91)	(Moderate) Deleterious (0.93)
1	CAIS	c.2319-1G>T	Splice	Splice	6	3	Publicada	Provavelmente Patogênica	-	-
1	CAIS	c.2338C>T	p.Arg780Trp	Missense	6	3	Publicada	Patogênica	Deleterious (0.84)	(Moderate) Deleterious (0.86)
2	CAIS	c.2336C>A	p. Ser779Tyr	Missense	6	3	Publicada	Provavelmente Patogênica	Deleterious (0.97)	(Strong) Deleterious (0.94)
1	CAIS	c.2450-1G>C	Splice	Splice	7	3	Publicada	Provavelmente Patogênica	-	-
1	CAIS	c.2595C>G	p.Asp865Glu	Missense	7	3	Publicada	Provavelmente Patogênica	Deleterious (0.86)	(Moderate) Deleterious (0.93)
2	CAIS	c.2521C>G	p.Arg841Gly	Missense	7	3	Publicada	Provavelmente Patogênica	Deleterious (0.91)	(Moderate) Deleterious (0.92)
2	CAIS	c.2566 C>T	p.Arg856Cys	Missense	7	3	Publicada	Patogênica	Deleterious (0.95)	(Strong) Deleterious (0.97)
1	CAIS	c.2674T>G	p.Phe892Val	Missense	8	3	Publicada	Provavelmente Patogênica	Deleterious (0.94)	(Strong) Deleterious (0.99)
1	CAIS	c.2695A>T	p.Ile899Phe	Missense	8	3	Publicada	Provavelmente Patogênica	Deleterious (0.96)	(Strong) Deleterious (0.99)
1	CAIS	c.2714C>G	p.Pro905Arg	Missense	8	3	Publicada	Provavelmente Patogênica	Deleterious (0.97)	(Strong) Deleterious (0.99)
1	PAIS	c.1215G>T	p.Gln405His	Missense	1	1	Publicada	VUS	Uncertain (0.62)	Deleterious (0.9)
1	PAIS	c.1727C>T	p.Thr576Ile	Missense	2	2	Publicada	Provavelmente Patogênica	Deleterious (0.97)	(Strong) Deleterious (0.96)
3	PAIS	c.1789 G>A	p.Ala597Thr	Missense	3	1	Publicada	Provavelmente Patogênica	Deleterious (0.8)	(Moderate) Deleterious (0.91)
1	PAIS	c.1885G>A	p.Ala629Thr	Missense	3	2	Publicada	VUS	Deleterious (0.9)	(Moderate) Deleterious (0.77)
2	PAIS	c.1807A>C	p.Thr603Pro	Missense	3	2	Publicada	Provavelmente Patogênica	Deleterious (0.92)	(Moderate) Deleterious (0.88)
1	PAIS	c.2084 C>T	p.Pro695Leu	Missense	4	3	Publicada	Provavelmente Patogênica	Deleterious (0.79)	(Moderate) Deleterious (0.99)

Tabela 8 - Variantes encontradas em pacientes com AIS em nossa coorte e reclassificação (continuação).

Número de variantes	Fenótipo	Posição codon	Posição proteína	Tipo	AR exon	Domínio	Publicada	ACMG	Revel	Meta LR	
1	PAIS	c.2144 A>C	p.His715Pro	Missense	4	3	Publicada	Provavelmente Patogênica	Deleterious (0.8)	(Moderate)	Deleterious (0.85)
1	PAIS	c.2110A>G	p.Ser704Gly	Missense	4	3	Publicada	Provavelmente Patogênica	Deleterious (0.94)	(Strong)	Deleterious (0.93)
1	PAIS	c.2108C>A	p.Ser703Tyr	Missense	4	3	Publicada	Provavelmente Patogênica	Uncertain (0.63)		Deleterious (0.89)
1	PAIS	c.2227A>G	p.Met743Val	Missense	5	3	Publicada	Provavelmente Patogênica	Deleterious (0.95)	(Strong)	Deleterious (0.94)
1	PAIS	c.2226G>C	p.Trp742Cys	Missense	5	3	Publicada	Provavelmente Patogênica	Deleterious (0.97)	(Strong)	Deleterious (0.96)
1	PAIS	c.2342 T>G	p.Met781Arg	Missense	6	3	Publicada	Provavelmente Patogênica	Deleterious (0.97)	(Strong)	Deleterious (0.99)
2	PAIS	c.170_172dup	p.Leu57dup	Nonsense	7	3	Publicada	Provavelmente Patogênica	-		-
3	PAIS	c.2521 C>A	p.Arg841Ser	Missense	7	3	Publicada	Provavelmente Patogênica	Deleterious (0.88)	(Moderate)	Deleterious (0.91)
2	PAIS	c.2522 G>A	p. Arg841His	Missense	7	3	Publicada	Patogênica	Deleterious (0.91)	(Moderate)	Deleterious (0.91)
1	PAIS	c.2515C>T	p.Leu839Phe	Missense	7	3	Publicada	Provavelmente Patogênica	Deleterious (0.93)	(Strong)	Deleterious (1)
2	PAIS	c.2567 G>A	p.Arg856His	Missense	7	3	Publicada	Patogênica	Deleterious (0.97)	(Strong)	Deleterious (0.97)
3	PAIS	c.2599G>C	p.Val867Leu	Missense	7	3	Publicada	Provavelmente Patogênica	Uncertain (0.56)		Deleterious (0.75)
2	PAIS	c.2612C>T	p.Ala871Val	Missense	8	3	Publicada	Patogênica	"		
Uncertain (0.62)"	Deleterious (0.87)	Benign (0)									
1	PAIS	c.2741C>G	p.Pro914Leu	Missense	8	3	Publicada	Provavelmente Patogênica	Deleterious (0.95)	(Strong)	Deleterious (0.99)
1	PAIS	c.2629T>G	p.Phe887Val	Missense	8	3	Publicada	Provavelmente Patogênica	Deleterious (0.96)	(Strong)	Deleterious (0.8)
2	PAIS	c.2612C>G	p.Ala871Gly	Missense	8	3	Publicada	Provavelmente Patogênica	Deleterious (0.96)	(Strong)	Deleterious (0.8)
1	PAIS	c.2677C>T	p.Pro893Ser	Missense	8	3	Publicada	Provavelmente Patogênica	Deleterious (0.97)	(Strong)	Deleterious (0.99)

Tabela 8 - Variantes encontradas em pacientes com AIS em nossa coorte e reclassificação (conclusão).

Número de variantes	Fenótipo	Posição codon	Posição proteína	Tipo	AR exon	Domínio	Publicada	ACMG	Revel	Meta LR
1	PAIS	c.2667C>T	p.Ser889Ser	Sinônima	8	3	Publicada	Provavelmente Patogênica		

ANEXO 3 – Tabela com as variantes do tipo *small indels* identificadas em pacientes com AIS.

Tabela 9 - Variantes do tipo *small indels* identificadas em pacientes com AIS (continua).

Fenótipo	Tipo de variante	Exon	Domínio	Posição do codon	Número do nucleotídeo	Tipo de indel	Classificação da indel	Posição proteína	Frame	Reference
CAIS	Del	1	NTD	c.19del	1	NR,non-CCC	Non-CCC	p.Leu7Trpfs*27	FS	41
CAIS	Ins	1	NTD	c.39_42dup	4	NR,non-CCC	Non-CCC	p.Pro15Alafs*70	FS	81
CAIS	Del	1	NTD	c.115_118del	4	NR,non-CCC	Non-CCC	p.Pro39del	NFS	82
CAIS	Del	1	NTD	c.118del	1	NR,non-CCC	Non-CCC	p.Arg40Glyfs*135	FS	83
CAIS	Del	1	NTD	c.125del	1	NR,non-CCC	CCC	p.Pro42Glnfs*133	FS	84
CAIS	Del	1	NTD	c.141_149del	9	NR,non-CCC	Non-CCC	p.Ala48_Pro50del	NFS	85
MAIS	Ins	1	NTD	c.157_165dup	9	TR	CCC	p.Ser53_Leu55dup	NFS	86
CAIS	Ins	1	NTD	c.163_164insC	1	TR	CCC	p.Leu55Profs*29	FS	87
MAIS	Ins	1	NTD	NR	3	TR	CCC	p.Leu57dup	NFS	88
CAIS	Ins	1	NTD	c.179_180ins	1	TR	CCC	p.Gln61Alafs*23	FS	89
CAIS	Del	1	NTD	c.227del	1	NR,non-CCC	Non-CCC	p.Gln76Arg*99	FS	87
CAIS	Ins	1	NTD	c.240del	1	NR,non-CCC	CCC	p.Glu81Argfs*3	FS	48
CAIS	Del	1	NTD	c.244_248del	5	NR,non-CCC	Non-CCC	p.Thr82profs*10	FS	87
CAIS	Ins	1	NTD	c.257insA	1	TR	CCC	p.Gln87Alafs*7	FS	90
MAIS	Ins	1	NTD	c.256_257insCAG	3	TR	CCC	p.Gln86dup	NFS	91
MAIS	Del	1	NTD	c.256_257delCAG	3	TR	CCC	p.Gln86del	NFS	91
CAIS	Ins	1	NTD	c.311_312insA	1	NR,non-CCC	Non-CCC	p.Thr105Hisfs*8	FS	87
CAIS	Del	1	NTD	c.312del	1	NR,non-CCC	CCC	p.Thr105Glnfs*70	FS	90
CAIS	Ins	1	NTD	c.368dup	1	NR,non-CCC	CCC	p.Glu124Glyfs*34	FS	92
CAIS	Del	1	NTD	c.381del	1	NR,non-CCC	CCC	p.Glu128Argfs*47	FS	90
CAIS	Del	1	NTD	c.387del	1	NR,non-CCC	Non-CCC	p.Gly130Valfs*45	FS	93
CAIS	Del	1	NTD	c.502del	1	NR,non-CCC	CCC	p.Gly168Alafs*7	FS	94
CAIS	Ins	1	NTD	c.606_609dup	4	NR,non-CCC	CCC	p.Glu204Ilefs*32	FS	93
CAIS	Del	1	NTD	c.619del	1	NR,non-CCC	Non-CCC	p.Gly228Alafs*30	FS	53
CAIS	Ins	1	NTD	c.649dup	1	NR,non-CCC	CCC	p.Ala217Glyfs*18	FS	95

Tabela 9 - Variantes do tipo *small indels* identificadas em pacientes com AIS (continuação).

Fenótipo	Tipo de variante	Exon	Domínio	Posição do codon	Número do nucleotídeo	Tipo de indel	Classificação da indel	Posição proteína	Frame	Reference
CAIS	Del	1	NTD	c.663del	1	NR,non-CCC	Non-CCC	p.Lys222Argfs*5	FS	48
CAIS	Ins	1	NTD	c.720dup	1	NR,non-CCC	CCC	p.Lys241*	FS	48
CAIS	Ins	1	NTD	c.796dup	1	NR,non-CCC	CCC	p.Asp266Gly*39	FS	13
CAIS	Del	1	NTD	c.797del	1	NR,non-CCC	Non-CCC	Asp266Valfs*30	FS	96
CAIS	Del	1	NTD	c.1029_1032del	4	NR,non-CCC	Non-CCC	p.Thr344Cysfs*134	FS	87
CAIS	Del	1	NTD	c.1035delinsAG	2	NR,non-CCC	Non-CCC	p.Ser346Glyfs*148	FS	87
CAIS	Del	1	NTD	c.1041_1042del	2	NR,non-CCC	Non-CCC	p.Tyr348Glnfs*153	FS	48
CAIS	Del	1	NTD	c.1041del	1	NR,non-CCC	Non-CCC	p.Tyr348Thrfs*131	FS	53
CAIS	Ins	1	NTD	c.1082_1083insC	1	NR,non-CCC	Non-CCC	p.Arg362Ser*140	FS	97
CAIS	Ins	1	NTD	c.1133_1141dup	9	TR	CCC	p.Pro379_Pro381dup	NFS	87
CAIS	Del	1	NTD	c.1389_1390del	2	TR	Non-CCC	p.Gly464Argfs*37	FS	98
CAIS	Ins	1	NTD	c.1396_1397insAG	2	TR	Non-CCC	p.Gly466Glu*14	FS	99
CAIS	Del	1	NTD	c.1421_1422del	2	TR	Non-CCC	p.Glu474Gly*27	FS	90
CAIS	Del	1	NTD	c.1462del	1	NR,non-CCC	Non-CCC	p.Gln488Argfs*22	FS	84
CAIS	Indel	1	NTD	c.1504_1509delins CGGCATGGTGA GCAGAGTGGCA		NR,non-CCC	Non-CCC	p.Val502Argfs*34	FS	87
CAIS	Ins	1	NTD	c.1589_1590insGCTA	4	NR,non-CCC	Non-CCC	p.Ser530Argfs*23	FS	87
CAIS	Ins	1	NTD	c.1314_1315del	2	NR,non-CCC	CCC	p.Phe439Hisfs*62	FS	Este estudo
CAIS	Ins	Intron	NTD	c.1616+1dup	1	NR,non-CCC	CCC	-	-	97
CAIS	Ins	1	NTD	c.228_229insTAG	3	TR	Non-CCC	p.Q77Term	NFS	Este estudo
CAIS	Indel	1	NTD	c.431_452delinsAA		NR,non-CCC	Non-CCC	p.Gly144Glufs*25	FS	Este estudo
CAIS	Ins	2	DBD	c.1686_1689dup	4	NR,non-CCC	CCC	p.Cys563dup	NFS	94
CAIS	Del	2	DBD	c.1740del	1	NR,non-CCC	Non-CCC	p.Cys580*	FS	100

Tabela 9 - Variantes do tipo *small indels* identificadas em pacientes com AIS (continuação).

Fenótipo	Tipo de variante	Exon	Domínio	Posição do codon	Número do nucleotídeo	Tipo de indel	Classificação da indel	Posição proteína	Frame	Reference
CAIS	Del	2	DBD	c.1750_1752del	3	NR,non-CCC	Non-CCC	p.Phe584del	NFS	101
CAIS	Del	3	DBD	c.1799del	1	NR,non-CCC	CCC	p.Asn600Metfs*26	FS	43
CAIS	Del	3	DBD	c.1825_1829del	5	NR,non-CCC	Non-CCC	p.Arg609Lysfs*12	FS	102
CAIS	Ins	3	DBD	c.1831insA	1	NR,non-CCC	CCC	p.Asn611Ilefs*15	FS	43
CAIS	Del	3	DBD	c.1846_1048del	3	NR,non-CCC	Non-CCC	p.Arg616del	NFS	101
CAIS	Del	4	Hinge	c.1910del	1	NR,non-CCC	Non-CCC	p.Asn637Ilefs*2	FS	103
CAIS	Ins	4	LDB	c.2055_2056dup	2	NR,non-CCC	CCC	p.Val686Argfs*104	FS	104
CAIS	Del	4	LDB	c.2074_2076del	3	NR,non-CCC	CCC	p.Asn693del	NFS	53
CAIS	Del	4	LDB	c.2101del	1	NR,non-CCC	CCC	p.Leu701Cysfs*88	FS	43
CAIS	Del	4	LDB	c.2123del	1	NR,non-CCC	Non-CCC	p.Leu708Argfs*81	FS	105
CAIS	Del	4	LDB	c.2135del	1	NR,non-CCC	Non-CCC	p.Gln712Argfs*77	FS	56
CAIS	Del	4	LDB	c.2164del	1	NR,non-CCC	CCC	p.Ala722Profs*67	FS	106
CAIS	Del	4	LDB	c.2166del	1	NR,non-CCC	CCC	p.Leu723Cysfs*66	FS	106
CAIS	Del	4	LDB	c.2174del	1	NR,non-CCC	CCC	Gly725Alafs*64	FS	107
CAIS	Del	5	LDB	c.2300del	1	NR,non-CCC	CCC	Pro767Leufs*22	FS	106
CAIS	Del	5	LDB	c.2301del	1	NR,non-CCC	Non-CCC	Asp768Ilefs*21	FS	108
CAIS	Del	5	LDB	c.2660del	1	NR,non-CCC	Non-CCC	p.Met887Argfs*2	FS	107
CAIS	Indel	5	LDB	c.2312_2313delTCinsAA	2	NR,non-CCC	Non-CCC	p.Phe771*	NFS	78
CAIS	Del	5	LDB	c.2317_2321del	5	NR,non-CCC	Non-CCC	p.Glu773Thrfs*15	FS	94
CAIS	Del	6	LDB	c.2378_2380del	2	NR,non-CCC	CCC	p.Gln793Argfs*36	FS	48
CAIS	Del	6	LDB	c.2356_2357del	2	NR,non-CCC	CCC	p.Ser783Argfs*46	FS	78
CAIS	Ins	6	LDB	c.2365_2368dup	4	NR,non-CCC	CCC	p.Arg789Lysfs*41	FS	43
CAIS	Del	6	LDB	c.2378_2379del	2	NR,non-CCC	Non-CCC	p.Gln793Argfs*36	FS	48
CAIS	Del	6	LDB	c.2403del	1		CCC	Gln803Argfs*6	FS	84
CAIS	Ins	7	LDB	c.2478_2479ins	2	NR,non-CCC	Non-CCC	p.Phe827Hisfs*8	FS	109

Tabela 9 - Variantes do tipo *small indels* identificadas em pacientes com AIS (conclusão).

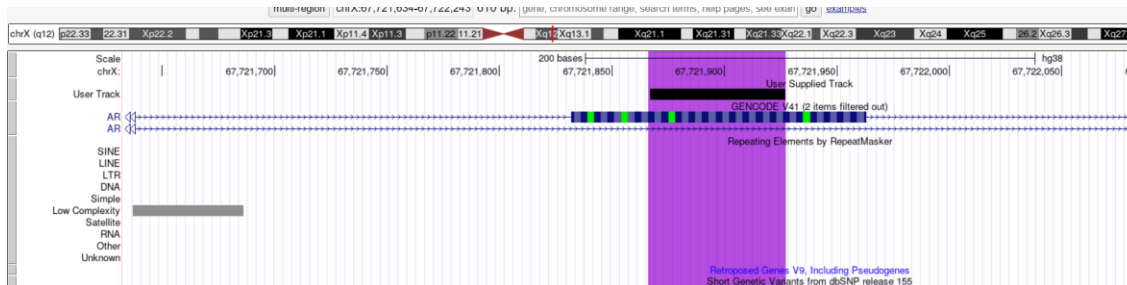
Fenótipo	Tipo de variante	Exon	Domínio	Posição do codon	Número do nucleotídeo	Tipo de indel	Classificação da indel	Posição proteína	Frame	Reference
CAIS	Del	7	LDB	c.2500_2502del	3		Non-CCC	p.Asn834del	NFS	48
CAIS	Ins	7	LDB	c.2546insA	1	HR	CCC	p.Asn849Lysfs*32	FS	53
CAIS	Del	7	LDB	c.2562_2563del	2	NR,non-CCC	Non-CCC	p.Arg855Thrfs*25	FS	110
CAIS	Ins	7	LDB	c.2573dup	1	NR,non-CCC	CCC	p.Tyr858*	FS	111
CAIS	Indel	8	LDB	c.2640_2646delins TTCGCCCTGA		NR,non-CCC	Non-CCC	p.Leu881Serfs*80	FS	112
CAIS	Del	8	LDB	c.2683_2686del	4	NR,non-CCC	CCC	p.Met896del	NFS	110
CAIS	Del	8	LDB	c.2715del	1	NR,non-CCC	CCC	Lys906Argfs*38	FS	48
CAIS	Del	8	LDB	c.2749_2763del	15	NR,non-CCC	Non-CCC	**	FS	113
PAIS	Del	4	LDB	c.2077_2079del	3	NR,non-CCC		p.Asn693del	NFS	84
CAIS	Ins	1	NTD	c.192_193insTAGC AG	6	CCC	CCC	p.Gln65*	NFS	114

ANEXO 4 - Sondas usadas no MLPA da grande deleção.

Primeira sonda deletada no MLPA:

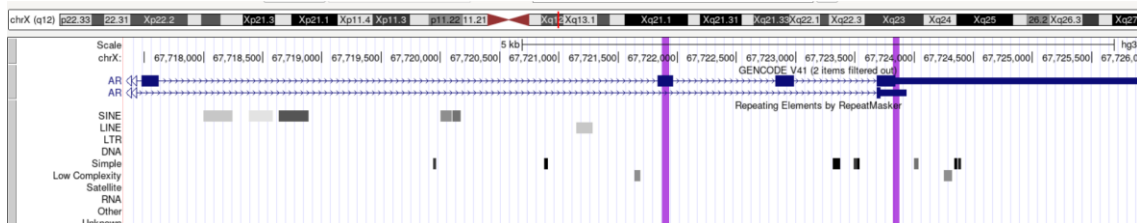
TGTGTCCGAATGAGGCACCTCTCTCAAGAGTTTGGATGGCTCCAAATCACCCCCCAG
GAAT

chrX: 67721867 67721927 61pb exon 6



Última sonda deletada no MLPA:

chrX 67723821 67723877



ANEXO 5 - Critérios ACMG/AMP para classificação das variantes alélicas quanto à patogenicidade.

Critério muito forte para patogenicidade	
PVS1	Variante <i>nonsense</i> , <i>frameshift</i> , em sítios de <i>splice</i> canônicos ± 1 ou 2, códon de iniciação ou deleção de 1 ou vários exons (<i>null variant</i>) em um gene no qual variantes com perda de função (<i>LoF</i>) são mecanismos conhecidos de doença
Critérios fortes para patogenicidade	
PS1	Troca de aminoácido já estabelecida como patogênica, independente da troca do nucleotídeo
PS2	Variante <i>de novo</i> em um paciente com doença sem história familiar (com paternidade confirmada)
PS3	Presença de estudos funcionais (<i>in vitro</i> ou <i>in vivo</i>) bem estabelecidos que suportem o efeito prejudicial no gene ou sua proteína
PS4	A prevalência da variante nos indivíduos afetados é significativamente maior quando comparada com a prevalência em controles
Critérios moderados para patogenicidade	
PM1	Variante localizada em uma região <i>hotspot</i> para mutações e/ou região importante já estabelecida com um domínio funcional da proteína, sem variantes descritas como benignas
PM2	Variante ausente nos bancos de dados populacionais (<i>1000Genomes</i> , ExAC) (ou presente em frequência extremamente baixa, para os casos recessivos)
PM3	Variante detectada em <i>trans</i> com uma variante patogênica para as doenças recessivas
PM4	Modificação no tamanho da proteína causada por deleções ou inserções <i>in-frame</i> em região não repetitiva ou por variantes que perdem o códon de parada (tipo <i>stop-loss</i>)
PM5	Nova variante <i>missense</i> em um aminoácido em que uma variante <i>missense</i> diferente já foi estabelecida como patogênica
PM6	Variante considerada <i>de novo</i> , sem confirmação da paternidade
Critérios fracos para patogenicidade	
PP1	Segregação da variante (em um gene conhecido como causador de doença) com a doença em uma família com vários membros afetados
PP2	Variante <i>missense</i> em um gene com uma baixa taxa de mutações <i>missense</i> benignas e no qual esse tipo de variante é considerado um mecanismo comum de doença
PP3	Várias ferramentas computacionais suportam a evidência de efeito deletério no gene ou proteína (conservação evolutiva, impacto no <i>splicing</i> etc)
PP4	O fenótipo do paciente ou a história da família são bastante específicos de uma doença monogênica
PP5	Variante recém-publicada como patogênica, porém o laboratório não apresenta condições de fazer um estudo funcional independente.
Critérios fracos para benignidade	
BP1	Variante <i>missense</i> em um gene no qual variantes com perda de função (<i>LoF</i>) são mecanismos conhecidos de doença

- BP2** Variante detectada em *trans* com uma variante patogênica dominante com penetrância completa ou observada em *cis* com uma variante patogênica em qualquer tipo de herança
- BP3** Deleções ou inserções *in-frame* em região repetitiva sem função conhecida

(conclusão) - Critérios do ACMG/AMP para classificação das variantes alélicas quanto à patogenicidade

Critérios fracos para benignidade

- BP4** Várias ferramentas computacionais sugerindo ausência de efeito deletério no gene ou proteína (conservação evolutiva, impacto no splicing etc)
- BP5** Variante encontrada em um caso com uma base molecular alternativa para doença
- BP6** Variante recém-publicada como benigna, porém o laboratório não apresenta condições de fazer um estudo funcional independente.
- BP7** Variante sinônima cujas ferramentas de predição de splicing não predizem impacto na sequência consenso de splice nem a criação de um novo sítio, e o nucleotídeo não é conservado

Critérios fortes para benignidade

- BS1** A frequência do alelo é maior que a esperada para a doença
- BS2** Variante para doença recessiva (em homozigose), dominante (em heterozigose) ou ligada ao X (hemizigose), com penetrância completa e de início precoce, observada em indivíduo adulto saudável
- BS3** Presença de estudos funcionais (*in vitro* ou *in vivo*) bem estabelecidos mostrando ausência de efeito prejudicial para a proteína ou *splicing*
- BS4** Ausência de segregação nos indivíduos afetados da família

Critério benigno por si só

- BA1** Frequência alélica maior que 5% nos bancos de dados populacionais (*1000Genomes*, *ExAC*)

Fonte: Adaptado de Richards e colaboradores, 2015 (61). PVS: critério muito forte para patogenicidade; PS: critério forte para patogenicidade; PM: critério moderado para patogenicidade; PP: critério fraco para patogenicidade; BP: critério fraco para benignidade; BS: critério forte para benignidade; BA: critério benigno por si só. *Lof*: perda de função (de *loss of function*); *1000Genomes*: *1000Genomes Project*; *ExAC*: *Exome Aggregation Consortium*.

3 _Classificação	Combinação dos critérios
Variante patogênica	<p><i>(I)</i> 1 critério muito forte para patogenicidade (PVS1) <i>e</i></p> <p><i>(a)</i> ≥ 1 critérios fortes para patogenicidade (PS1-PS4) <i>ou</i></p> <p><i>(b)</i> ≥ 2 critérios moderados para patogenicidade (PM1-PM6) <i>ou</i></p> <p><i>(c)</i> 1 critério moderado para patogenicidade (PM1-PM6) <i>e</i> 1 critério fraco para patogenicidade (PP1-PP5) <i>ou</i></p> <p><i>(d)</i> ≥ 2 critérios fracos para patogenicidade (PP1-PP5) <i>ou</i></p> <p><i>(II)</i> ≥ 2 critérios fortes para patogenicidade (PS1-PS4) <i>ou</i></p> <p><i>(III)</i> 1 critério forte para patogenicidade (PS1-PS4) <i>e</i></p> <p><i>(a)</i> ≥ 3 critérios moderados para patogenicidade (PM1-PM6) <i>ou</i></p> <p><i>(b)</i> 2 critérios moderados para patogenicidade (PM1-PM6) <i>e</i> ≥ 2 critérios fracos para patogenicidade (PP1-PP5) <i>ou</i></p> <p><i>(c)</i> 1 critérios moderados para patogenicidade (PM1-PM6) <i>e</i> ≥ 4 critérios fracos para patogenicidade (PP1-PP5)</p>
Variante provavelmente patogênica	<p><i>(I)</i> 1 critério muito forte para patogenicidade (PVS1) <i>e</i> 1 critério moderado para patogenicidade (PM1-PM6) <i>ou</i></p> <p><i>(II)</i> 1 critério forte para patogenicidade (PS1-PS4) <i>e</i> 1 ou 2 critérios moderado para patogenicidade (PM1-PM6) <i>ou</i></p> <p><i>(III)</i> 1 critério forte para patogenicidade (PS1-PS4) <i>e</i> ≥ 2 critérios fracos para patogenicidade (PP1-PP5) <i>ou</i></p> <p><i>(IV)</i> ≥ 3 critérios moderados para patogenicidade (PM1-PM6) <i>ou</i></p> <p><i>(V)</i> 2 critérios moderados para patogenicidade (PM1-PM6) <i>e</i> ≥ 2 critérios fracos para patogenicidade (PP1-PP5) <i>ou</i></p> <p><i>(VI)</i> 1 critério moderado para patogenicidade (PM1-PM6) <i>e</i> ≥ 4 critérios fracos para patogenicidade (PP1-PP5)</p>
Variante benigna	<p><i>(I)</i> 1 critério benigno por si só (BA1) <i>ou</i></p> <p><i>(II)</i> ≥ 2 critérios fortes para benignidade (BS1-BS4)</p>
Variante provavelmente benigna	<p><i>(I)</i> 1 critério forte para benignidade (BS1-BS4) <i>e</i> 1 critério fraco para benignidade (BP1-BP7) <i>ou</i></p> <p><i>(II)</i> ≥ 2 critérios fracos para benignidade (BP1-BP7)</p>

Variante de significado incerto	(I)	outra combinação de critérios não mencionada acima <i>ou</i>
	(II)	critérios para benignidade e patogenicidade contraditórios

Fonte: Adaptado de *Richards* e colaboradores, 2015 (61). PVS: critério muito forte para patogenicidade; PS: critério forte para patogenicidade; PM: critério moderado para patogenicidade; PP: critério fraco para patogenicidade; BP: critério fraco para benignidade; BS: critério forte para benignidade; BA: critério benigno por si só.

APÊNDICES

APÊNDICE 1- Técnica MLPA AR

Kit: P074AR-100R

Dia 1 - Quantificar DNA

Desnaturação

Adicionar 5uL de DNA (50-100ng) e completar com H₂O

Desnaturar a amostra por 5 min a 98°C

Esfriar amostras a 25°C – ∞

Hibridização

Vortex MLPA buffer e MLPA probemix

Preparar mix para cada amostra

1,5uL MLPA buffer (yellow cap)

1,5uL MLPA pobemix (black cap)

Vortex

Add 3uL em cada amostra

Misturar pipetando (up and down)

Continuar no termociclador: incubação por 1 min a 95°C e 16-20h a 60°C

Dia 2

Vortex nos buffers

Preparar o mix para cada amostra

Adicionar 2,5 de H₂O

Adicionar 3uL de Buffer A

Adicionar 3uL de Buffer B

Adicionar 1uL de enzima ligase

Misturar pipetando gentilmente

Colocar no termociclador até 54°C, quando atingir a temperatura, adicionar 32uL de ligase em cada tubo

Continuar no termociclador a 54°C por 15 minutos e a 98°C por 5 minutos, pausar a 20°C

PCR

Vortex SALSA PCR primers

Preparação mix para cada reação

7,5ul de H2O

2uL SALSA PCR primer mix

0,5uL SALSA polymerase

Misture pipetando

Adicionar 10uL mix polimerase e misture

Termociclador

35 ciclos {95°C - 30 segundos

{60°C- 30 segundos


{72°C - 1 minuto

72°C - 20 minutos

15°C – ∞

APÊNDICE 2- Artigo “Small Indels in the Androgen Receptor Gene: Phenotype Implications and Mechanisms of Mutagenesis”

Small Indels in the Androgen Receptor Gene: Phenotype Implications and Mechanisms of Mutagenesis

Raquel Martinez Ramos,¹ Reginaldo José Petrolí,² Nathália Da Roz D'Alessandre,³ Gabriela Der Agopian Guardia,³ Ana Caroline de Freitas Afonso,¹ Mirian Yumie Nishi,¹ Sorahia Domenice,¹ Pedro Alexandre Favoretto Galante,³ Berenice Bilharinho Mendonça,¹ and Rafael Loch Batista^{1,4} 

¹Developmental Endocrinology Unit, Hormone and Molecular Genetics Laboratory (LIM/42), Endocrinology Division, Internal Medicine Department, Medical School, University of São Paulo (USP), São Paulo, SP, 05403-000, Brazil

²Faculdade de Medicina da Universidade Federal de Alagoas (UFAL), Programa de Pós-Graduação em Ciências Médicas—UFAL, Maceió, AL, 57072-900, Brazil

³Molecular Oncology Center, Hospital Sírio Libanês, São Paulo, SP, 01308-060, Brazil

⁴Instituto do Câncer do Estado de São Paulo da Faculdade, de Medicina da Universidade de São Paulo (ICESP), São Paulo, SP, 01246-000, Brazil

Correspondence: Rafael Loch Batista, MD, PhD, Unidade de Endocrinologia do Desenvolvimento, Laboratório de Hormônios e Genética Molecular/LIM42, Hospital das Clínicas da Universidade de São Paulo, Disciplina de Endocrinologia, do Departamento de Clínica Médica, Faculdade de Medicina da Universidade de São Paulo-USP, Eneas Carvalho de Aguiar, 255, São Paulo, SP, 05403-000, Brazil. Email: rafael.loch@hc.fm.usp.br.

Abstract

Context: Despite high abundance of small indels in human genomes, their precise roles and underlying mechanisms of mutagenesis in Mendelian disorders require further investigation.

Objective: To profile the distribution, functional implications, and mechanisms of small indels in the androgen receptor (AR) gene in individuals with androgen insensitivity syndrome (AIS).

Methods: We conducted a systematic review of previously reported indels within the coding region of the AR gene, including 3 novel indels. Distribution throughout the AR coding region was examined and compared with genomic population data. Additionally, we assessed their impact on the AIS phenotype and investigated potential mechanisms driving their occurrence.

Results: A total of 82 indels in AIS were included. Notably, all frameshift indels exhibited complete AIS. The distribution of indels across the AR gene showed a predominance in the N-terminal domain, most leading to frameshift mutations. Small deletions accounted for 59.7%. Most indels occurred in nonrepetitive sequences, with 15.8% situated within triplet regions. Gene burden analysis demonstrated significant enrichment of frameshift indels in AIS compared with controls ($P < .00001$), and deletions were overrepresented in AIS ($P < .00001$).

Conclusion: Our findings underscore a robust genotype-phenotype relationship regarding small indels in the AR gene in AIS, with a vast majority presenting complete AIS. Triplet regions and homopolymeric runs emerged as prone loci for small indels within the AR. Most were frameshift indels, with polymerase slippage potentially explaining half of AR indel occurrences. Complex frameshift indels exhibited association with palindromic runs. These discoveries advance understanding of the genetic basis of AIS and shed light on potential mechanisms underlying pathogenic small indel events.

Key Words: androgen receptor, androgen insensitivity syndrome, human genetics, disorders of sex development, sex development, differences in sex development

Abbreviations: ACMG, American College of Medical Genetics and Genomics; AIS, androgen insensitivity syndrome; AMH, anti-Müllerian hormone; AR, androgen receptor; CAIS, complete AIS; CCC, change in copy count; DSD, differences in sexual development; FS, frameshift; FSH, follicle-stimulating hormone; HGMD, Human Gene Mutation Database; LH, luteinizing hormone; MAIS, mild AIS; NFS, non-frameshift; PAIS, partial AIS; PCR, polymerase chain reaction.

Indels, which are insertions or deletions of one or more nucleotides in the DNA sequence (1), are abundant in human genomes, constituting 15% to 21% of human polymorphisms, second only to single nucleotide polymorphisms (SNPs) (2). These alterations can disrupt proteins through various mechanisms, such as modifying amino acid sequences and gene expression (2). Depending on their location within coding regions, indels can result in frameshift (FS) or non-frameshift (NFS) changes. NFS indels involve multiples of 3 base pairs,

leading to the insertion or deletion of one or more amino acids while maintaining the remaining protein sequence unchanged. In contrast, FS indels alter the reading frame from the site of the indel, resulting in an aberrant protein sequence or a premature stop codon. Studies have demonstrated the role of indels in causing Mendelian diseases (3, 4), with more than 4000 disease-causing in-frame indels reported, accounting for 2.2% of all mutations in the Human Gene Mutation Database (HGMD) (3).

Androgen insensitivity syndrome (AIS; OMIM 300068) represents the most common etiology of differences in sexual development (DSD) among 46,XY individuals (5-7). The molecular basis of AIS lies in allelic variants of the androgen receptor gene (AR; OMIM 313700), located on the X chromosome (8), making it an X-linked inherited disorder. AIS manifests as 3 distinct phenotypes: complete (CAIS) with typically female external genitalia, partial (PAIS) with partially virilized external genitalia, and mild (MAIS) presenting with typical male external genitalia but accompanied by male infertility and gynecomastia (5). The literature describes more than 800 allelic variants in the AR gene (6). However, the pathogenic role of small indels in the genome requires further investigation due to certain challenges. Next-generation sequencing (NGS) methods may introduce artifacts, complicating accurate mapping, and coding regions have a low density of indels (9). Consequently, while evidence grows regarding the role of indels in genome evolution, our understanding of their involvement in genetic disorders remains limited (10). In this study, we collected all small indels previously reported in the AR gene among individuals with AIS, including 3 novel small indels found in 3 individuals with CAIS. Our objectives were to construct the small indels profile in AIS by analyzing their distribution throughout the AR coding region, comparing them with genomic population data, and assessing their impact on the AIS phenotype and their probable mutagenesis mechanisms.

Methods

This study presents a comprehensive systematic review focusing on small indels in the AR gene associated with AIS. The primary objective is to thoroughly investigate the presence and functional implications of exonic small insertions and deletions (indels) occurring within the AR gene's coding region in individuals diagnosed with AIS. The study incorporates an analysis of previously reported indels and introduces 3 novel indels, aiming to assess their distribution within the AR coding region, compare them with genomic data from the general population, and evaluate their potential impact on the AIS phenotype and the underlying mutagenic mechanisms.

Inclusion criteria: This study encompasses all previously reported indels occurring within the coding region of the AR gene in individuals diagnosed with AIS. Included indels must be classified as pathogenic or likely pathogenic, following the American College of Medical Genetics and Genomics (ACMG) criteria. Additionally, variants of uncertain significance were considered if they exhibited a classical AIS phenotype, demonstrated familial segregation, or were supported by supplementary functional studies.

Exclusion criteria: Small indels reported without any clinical evidence of AIS phenotype were excluded from the study.

Literature review: For the literature review, we performed a comprehensive search on PubMed from its inception to April 15, 2023, without any restrictions on language or geographical regions, using the search terms “androgen receptor” AND “androgen insensitivity syndrome” OR “disorders of sex development” OR “DSD” for research articles.

Database review: For the database review, we conducted a search for androgen receptor allelic variants in the HGMD (<https://www.hgmd.cf.ac.uk/>) and the androgen receptor mutations database (<http://androgendb.mcgill.ca/>) (8).

All small indels identified within the AR gene were subjected to classification according to the ACMG standards,

which include 5 categories of pathogenicity: pathogenic, likely pathogenic, variant of uncertain significance, likely benign, and benign (11).

Population data were drawn from 2 public genomic databases: the Genome Aggregation Database (gnomAD, <https://gnomad.broadinstitute.org>) and the Androgen Receptor Database (<https://abraom.ib.usp.br>) (8). To predict the pathogenicity of small indels in the AR gene, various computational tools were utilized, including PolyPhen2, SIFT, REVEL, MetaLR, MetaRNN, MetaSVM, Mutation Assessor, Mutation Taster, FATHMM, FATHMM-MKL, LRT, M-CAP, and DAAN (<https://franklin.genoox.com/clinical-db/home>). To ensure consistent variant description, the mutalyzer web tool (<https://mutalyzer.nl/>) was applied.

In-depth analysis of the genomic context of these indels was conducted, involving repetitive sequences, homopolymers, and trinucleotide repeats. Indels related to the polymerase slippage mechanism were classified as “change in copy count” (CCC) indels, reflecting expansion or contraction of the adjacent reference base or bases. The proportion of indels likely formed through polymerase slippage events was estimated by categorizing each indel as either CCC or non-CCC (12, 13).

For population data analysis, the Genome Aggregation Database (gnomAD) served as a comprehensive resource for studying variants impacting the AR gene. To provide additional genetic evidence of the association between small indels in the AR gene and the androgen insensitivity syndrome phenotype, rare small indels across the entire gene were compared between AIS cases and the gnomAD public database (as controls) using the Fisher exact test.

Furthermore, an enrichment analysis of frameshift (FS) indels was conducted in both the AR and gnomAD databases. Additionally, the occurrence and frequency of small insertions and deletions, as well as short coding indels spanning 1 to 3 nucleotides, were examined in both datasets. Statistical significance was defined at a *P* value of less than .05.

For the 3 novel cases of small indels in the AR gene, karyotyping analysis using G-banding was performed on at least 30 peripheral blood lymphocytes to obtain the karyotype. Genomic DNA was extracted from peripheral blood leukocytes using the proteinase K-SDS salting-out method (14). The entire coding region and the 5' UTR region of the AR gene was polymerase chain reaction (PCR)-amplified using specific primers, as previously reported (15). Additionally, the SRD5A2 gene was sequenced by exon-specific PCR in all 3 cases. For SRD5A2 sequencing, exons 1 to 5 of the SRD5A2 gene were amplified by PCR using pairs of specific primers for each exon, as previously described (16, 17).

Written informed consent for the genetic and laboratory study was obtained from patients or their relatives. This study was approved by the local ethical committee (under the number CAAE: 66367717.6.0000.0068).

Results

We identified 3 novel indels in patients with CAIS (Table 1).

Subject 1:

The first patient came to our attention at the age of 2 years after undergoing bilateral gonadectomy during the correction of a bilateral inguinal hernia. The external genitalia exhibited female-like characteristics. Histological examination of both testes revealed hypotrophic testis with atrophic Wolff derivatives. Hormonal analysis performed before gonadectomy,

Table 1. Identification of novel small indels within the androgen receptor gene in 3 individuals with complete androgen insensitivity syndrome

Subjects	AR	c.DNA	AR protein	Frame	ACMG criteria
1	exon 1	c.431_452delinsAAA	p.Gly144Glufs*25	FS	Pathogenic/PVS1
2	exon 1	c.228_229insTAG	p.Gln77*	NFS	Pathogenic/PVS1
3	exon 1	c.1314_1315delCT	p.Phe439His*62	FS	Pathogenic/PVS1

following a 72-hour hCG stimulus, indicated high testosterone levels of 37 ng/dL, along with an anti-Müllerian hormone (AMH) level of 236 ng/mL. The karyotype was determined to be 46,XY, and AR sequencing showed a small indel (c.431_452delinsAAA) in exon 1 of the AR gene (Table 1). Transabdominal ultrasound revealed the absence of the uterus. Notably, the AR analysis of her mother (46,XX) revealed the same indel in the AR gene in heterozygosity.

Subject 2:

The second patient came to our observation at the age of 16 years for an investigation of primary amenorrhea. She was referred to our service by her aunt, who had been diagnosed with CAIS during a primary amenorrhea investigation. Chromosome analysis confirmed a 46,XY karyotype. The patient displayed female-like external genitalia, feminine adipose tissue body distribution, typically developed breasts (Tanner stage 5, bilateral), and minimal axillary and pubic hair. Hormonal analysis revealed elevated LH (30.6 UI/L) and testosterone (743 ng/dL) levels, alongside normal follicle-stimulating hormone (FSH) levels (2.6 IU/L) and AMH levels of 172.5 ng/mL. Transabdominal ultrasound and pelvic nuclear magnetic resonance did not identify uterus or Müllerian derivatives. AR exonic sequencing identified a short indel (c.387_388delGA) in exon 1 of the AR gene (Table 1).

Subject 3:

The third patient came to our observation at the age of 2 years after bilateral inguinal hernia was detected at 2 months of age. Transabdominal ultrasound revealed the absence of the uterus and the presence of bilateral gonads in the inguinal region. Chromosome analysis confirmed a 46,XY karyotype. Hormonal study showed elevated LH (16 IU/L) and testosterone levels (725 ng/dL), along with normal FSH levels (5 IU/L). AR sequencing revealed a short indel (c.1314_1315delCT) in exon 1 of the AR gene (Table 1). Similar to Subject 1, the AR analysis of her mother (46,XX) showed the same indel in the AR gene in heterozygosity.

Importantly, no allelic variants were identified in the SRD5A2 gene in any of the described cases.

AR Indels in AIS

We included a total of 82 distinct indels (Fig. 1), including the 3 novel ones, in the AR gene reported among individuals with AIS, and the details are presented in Table 2. Most of these cases (n = 78; 95.1%) exhibited CAIS phenotype, while the remaining 4 cases were associated with MAIS. The indel c.2074_2076del was reported in 2 individuals from the same family, with one presenting PAIS and the other CAIS. The distribution of small indels across the AR gene showed a predominance in the N-terminal domain (n = 45; 54.9%), followed by the ligand-binding domain (LBD) (n = 28; 34.1%), DNA-binding domain (DBD) (n = 8; 9.7%), and Hinge region (n = 1). The majority of these indels were frameshift mutations (n = 64; 79%), with small deletions being

more prevalent (n = 49; 59.7%) than small insertions (n = 29; 35.3%) and complex frameshift indels (n = 4; 4.8%). Most of the indels were short, involving fewer than 10 base pairs (n = 81 = 98.8%), with an average of 2 to 4 deleted or inserted nucleotides. These indels were primarily found in nonrepetitive sequences (n = 68; 82.9%), while 13 were identified within triplet regions (15.8%), and 7 were located in homopolymers runs (8.5%).

Thirty-nine (47.5%) were CCC indels regarding the change in copy counting. Indels reported in MAIS were all located at tandem repeats, while all at nonrepetitive sequences caused CAIS.

In the comparative gene burden analysis, we analyzed a total of 79 and 118 indels reported in the AR gene and gnomAD database (consisting of 46,XY individuals as controls), respectively. Frameshift (FS) indels showed a significant enrichment in AIS individuals (63/79, 79.9%) compared with the male control population from gnomAD (14/118, 11.9%) ($P < .00001$). Additionally, deletions were significantly over-represented in AIS (64/79, 81%) compared with male controls (36/118, 30.5%) ($P < .00001$). In the context of short coding indels spanning 1 to 3 nucleotides, we observed a notable enrichment in individuals with AIS (61/79, 77.2%) relative to 46,XY controls (40/118, 33.9%) ($P < .00001$). These findings emphasize the association between specific types of indels and the development of AIS, providing valuable insights into the genetic basis of this condition.

Discussion

Indels are defined as additions or deletions of one or more nucleotides within a DNA sequence (1). Typically, small indels refer to insertions or deletions that are shorter than 50 base pairs, while longer indels are categorized as structural variants. In the realm of human genomes, indels are highly prevalent, constituting approximately 15% to 21% of human polymorphisms, second only to single nucleotide polymorphisms (2).

The significance of short insertion and deletion polymorphisms in shaping human genetic diversity and influencing various human phenotypes has been increasingly acknowledged (58, 59). Beyond their contribution to genetic variations in healthy populations, deleterious indels have been increasingly implicated in the pathogenesis of numerous genetic disorders, affecting both coding and noncoding regions (60, 61). A classic example of such a deleterious indel is observed in cystic fibrosis, a genetic condition caused by a 3-base-pair deletion within the CFTR (Cystic Fibrosis Transmembrane Conductance Regulator) gene (62).

In AIS, indels represent the second most prevalent type of genetic variants identified in the AR gene (5, 63). Tyutyusheva et al reported on 66 CAIS cases and compared the types of genetic variants identified with 2 genomic databases (64). Insertions and deletions were found to be the

Table 2. Indels in the androgen receptor gene among individuals with androgen insensitivity syndrome

Phenotype	Mutation type	Exon	Domain	cDNA	Size (nt)	Indel type	Indel classification	Predicted protein	Frame	ACMG	Ref
CAIS	Del	1	NTD	c.19del	1	NR, non-CCC	Non-CCC	p.Leu77Trpfs*27	FS	LP	(18)
CAIS	Ins	1	NTD	c.39_42dup	4	NR, non-CCC	Non-CCC	p.Pro15Alafs*70	FS	LP	(19)
CAIS	Del	1	NTD	c.115_118del	4	NR, non-CCC	Non-CCC	p.Pro39del	NFS	LP	(20)
CAIS	Del	1	NTD	c.118del	1	NR, non-CCC	Non-CCC	p.Arg40Glyfs*135	FS	LP	(21)
CAIS	Del	1	NTD	c.125del	1	NR, non-CCC	CCC	p.Pro42Glnfs*133	FS	LP	(22)
CAIS	Del	1	NTD	c.141_149del	9	NR, non-CCC	Non-CCC	p.Ala48_Pro50del	NFS	VUS	(23)
MAIS	Ins	1	NTD	c.157_165dup	9	TR	CCC	p.Ser53_Leu55dup	NFS	VUS	(24)
CAIS	Ins	1	NTD	c.163_164insC	1	TR	CCC	p.Leu55Profs*29	FS	LP	(25)
MAIS	Ins	1	NTD	NR	3	TR	CCC	p.Leu57dup	NFS	LP	(26)
CAIS	Ins	1	NTD	c.179_180ins	1	TR	CCC	p.Gln61Alafs*23	FS	LP	(27)
CAIS	Del	1	NTD	c.227del	1	NR, non-CCC	Non-CCC	p.Gln76Arg*99	FS	LP	(25)
CAIS	Ins	1	NTD	c.240del	1	NR, non-CCC	CCC	p.Glu81Argfs*3	FS	VUS	(28)
CAIS	Del	1	NTD	c.244_248del	5	NR, non-CCC	Non-CCC	p.Thr82profs*10	FS	LP	(25)
CAIS	Ins	1	NTD	c.257insA	1	TR	CCC	p.Gln87Alafs*7	FS	LP	(29)
MAIS	Ins	1	NTD	c.256_257insCAG	3	TR	CCC	p.Gln86dup	NFS	VUS	(30)
MAIS	Del	1	NTD	c.256_257delCAG	3	TR	CCC	p.Gln86del	NFS	VUS	(30)
CAIS	Ins	1	NTD	c.311_312insA	1	NR, non-CCC	Non-CCC	p.Thr105Hisfs*8	FS	LP	(25)
CAIS	Del	1	NTD	c.312del	1	NR, non-CCC	CCC	p.Thr105Glnfs*70	FS	LP	(29)
CAIS	Ins	1	NTD	c.368dup	1	NR, non-CCC	CCC	p.Glu124Glyfs*34	FS	LP	(31)
CAIS	Del	1	NTD	c.381del	1	NR, non-CCC	CCC	p.Glu128Argfs*47	FS	LP	(29)
CAIS	Del	1	NTD	c.387del	1	NR, non-CCC	Non-CCC	p.Gly130Valfs*45	FS	LP	(32)
CAIS	Del	1	NTD	c.502del	1	NR, non-CCC	CCC	p.Gly168Alafs*7	FS	LP	(33)
CAIS	Ins	1	NTD	c.606_609dup	4	NR, non-CCC	CCC	p.Glu204Ilefs*32	FS	LP	(32)

(continued)

Table 2. Continued

Phenotype	Mutation type	Exon	Domain	cDNA	Size (nt)	Indel type	Indel classification	Predicted protein	Frame	ACMG	Ref
CAIS	Del	1	NTD	c.619del	1	NR, non-CCC	Non-CCC	p.Gly228Alafs*30	FS	LP	(34)
CAIS	Ins	1	NTD	c.649dup	1	NR, non-CCC	CCC	p.Ala217Glyfs*18	FS	LP	(35)
CAIS	Del	1	NTD	c.663del	1	NR, non-CCC	Non-CCC	p.Lys222Argfs*5	FS	LP	(28)
CAIS	Ins	1	NTD	c.720dup	1	NR, non-CCC	CCC	p.Lys241*	FS	LP	(28)
CAIS	Ins	1	NTD	c.796dup	1	NR, non-CCC	CCC	p.Asp266Gly*39	FS	LP	(8)
CAIS	Del	1	NTD	c.797del	1	NR, non-CCC	Non-CCC	Asp266Valfs*30	FS	LP	(36)
CAIS	Del	1	NTD	c.1029_1032del	4	NR, non-CCC	Non-CCC	p.Thr344Cysfs*134	FS	LP	(25)
CAIS	Del	1	NTD	c.1035delinsAG	2	NR, non-CCC	Non-CCC	p.Ser346Glyfs*148	FS	LP	(25)
CAIS	Del	1	NTD	c.1041_1042del	2	NR, non-CCC	Non-CCC	p.Tyr348Glnfs*153	FS	LP	(28)
CAIS	Del	1	NTD	c.1041del	1	NR, non-CCC	Non-CCC	p.Tyr348Thrfs*131	FS	LP	(34)
CAIS	Ins	1	NTD	c.1082_1083insC	1	NR, non-CCC	Non-CCC	p.Arg362Ser*140	FS	LP	(37)
CAIS	Ins	1	NTD	c.1133_1141dup	9	TR	CCC	p.Pro379_Pro381dup	NFS	VUS	(25)
CAIS	Del	1	NTD	c.1389_1390del	2	TR	Non-CCC	p.(Gly464Argfs*37)	FS	LP	(38)
CAIS	Ins	1	NTD	c.1396_1397insAG	2	TR	Non-CCC	p.Gly466Glu*14	FS	LP	(39)
CAIS	Del	1	NTD	c.1421_1422del	2	TR	Non-CCC	p.Glu474Gly*27	FS	VUS	(29)
CAIS	Del	1	NTD	c.1462del	1	NR, non-CCC	Non-CCC	p.Gln488Argfs*22	FS	LP	(22)
CAIS	Indel	1	NTD	c.1504_1509delinsCGGCATGGTGAGCAGAGTGGCA	27	NR, non-CCC	Non-CCC	p.Val502Argfs*34	FS	LP	(25)
CAIS	Ins	1	NTD	c.1589_1590insGCTA	4	NR, non-CCC	Non-CCC	p.Ser330Argfs*23	FS	LP	(25)
CAIS	Ins	1	NTD	c.1314_1315del	2	NR, non-CCC	CCC	p.Phe439Hisfs*62	FS	LP	This study (37)
CAIS	Ins	splice	NTD	c.1616+1dup	1	NR, non-CCC	CCC	—	—	LP	(37)
CAIS	Ins	1	NTD	c.228_229insTAG	3	TR	Non-CCC	p.Q77Term	NFS	LP	This study
CAIS	Indel	1	NTD	c.431_452delinsAAA	24	NR, non-CCC	Non-CCC	p.Gly144Glufs*25	FS	LP	This study

(continued)

Table 2. Continued

Phenotype	Mutation type	Exon	Domain	cDNA	Size (nt)	Indel type	Indel classification	Predicted protein	Frame	ACMG	Ref
CAIS	Ins	2	DBD	c.1686_1689dup	4	NR, non-CCC	CCC	p.Cys563dup	NFS	LP	(33)
CAIS	Del	2	DBD	c.1740del	1	NR, non-CCC	Non-CCC	p.Cys580*	FS	LP	(40)
CAIS	Del	2	DBD	c.1750_1752del	3	NR, non-CCC	Non-CCC	p.Phe584del	NFS	LP	(41)
CAIS	Del	3	DBD	c.1799del	1	NR, non-CCC	CCC	p.Asn600Metfs*26	FS	LP	(42)
CAIS	Del	3	DBD	c.1825_1829del	5	NR, non-CCC	Non-CCC	p.Arg609Lysfs*12	FS	LP	(43)
CAIS	Ins	3	DBD	c.1831insA	1	NR, non-CCC	CCC	p.Asn611Ilefs*15	FS	LP	(42)
CAIS	Del	3	DBD	c.1846_1048del	3	NR, non-CCC	Non-CCC	p.Arg616del	NFS	LP	(41)
CAIS	Del	4	Hinge	c.1910del	1	NR, non-CCC	Non-CCC	p.Asn637Ilefs*2	FS	LP	(44)
CAIS	Ins	4	LBD	c.2055_2056dup	2	NR, non-CCC	CCC	p.Val686Argfs*104	FS	LP	(45)
CAIS	Del	4	LBD	c.2074_2076del	3	NR, non-CCC	CCC	p.Asn693del	NFS	LP	(34)
CAIS	Del	4	LBD	c.2101del	1	NR, non-CCC	CCC	p.Leu701Cysfs*88	FS	LP	(42)
CAIS	Del	4	LBD	c.2123del	1	NR, non-CCC	Non-CCC	p.Leu708Argfs*81	FS	LP	(46)
CAIS	Del	4	LBD	c.2135del	1	NR, non-CCC	Non-CCC	p.Gln712Argfs*77	FS	LP	(47)
CAIS	Del	4	LBD	c.2164del	1	NR, non-CCC	CCC	p.Ala722Profs*67	FS	LP	(48)
CAIS	Del	4	LBD	c.2166del	1	NR, non-CCC	CCC	p.Leu723Cysfs*66	FS	LP	(48)
CAIS	Del	4	LBD	c.2174del	1	NR, non-CCC	CCC	Gly725Alafs*64	FS	LP	(49)
CAIS	Del	5	LBD	c.2300del	1	NR, non-CCC	CCC	Pro767Leufs*22	FS	LP	(48)
CAIS	Del	5	LBD	c.2301del	1	NR, non-CCC	Non-CCC	Asp768Ilefs*21	FS	LP	(50)
CAIS	Del	5	LBD	c.2660del	1	NR, non-CCC	Non-CCC	p.Met887Argfs*2	FS	LP	(49)
CAIS	Indel	5	LBD	c.2312_2313del/TCinsAA	2	NR, non-CCC	Non-CCC	p.Phe771*	NFS	LP	(51)

(continued)

Table 2. Continued

Phenotype	Mutation type	Exon	Domain	cDNA	Size (nt)	Indel type	Indel classification	Predicted protein	Frame	ACMG	Ref
CAIS	Del	5	LBD	c.2317_2321del	5	NR, non-CCC	Non-CCC	p.Glu773Thrfs*15	FS	LP	(33)
CAIS	Del	6	LBD	c.2378_2380del	2	NR, non-CCC	CCC	p.Gln793Argfs*36	FS	LP	(28)
CAIS	Del	6	LBD	c.2356_2357del	2	NR, non-CCC	CCC	p.Ser783Argfs*46	FS	LP	(51)
CAIS	Ins	6	LBD	c.2365_2368dup	4	NR, non-CCC	CCC	p.Arg789Lysfs*41	FS	LP	(42)
CAIS	Del	6	LBD	c.2378_2379del	2	NR, non-CCC	Non-CCC	p.Gln793Argfs*36	FS	VUS	(28)
CAIS	Del	6	LBD	c.2403del	1	NR, non-CCC	CCC	Gln803Argfs*6	FS	LP	(22)
CAIS	Ins	7	LBD	c.2478_2479ins	2	NR, non-CCC	Non-CCC	p.Phe827Hisfs*8	FS	LP	(52)
CAIS	Del	7	LBD	c.2500_2502del	3	NR, non-CCC	Non-CCC	p.Asn834del	NFS	LP	(28)
CAIS	Ins	7	LBD	c.2546insA	1	HR	CCC	p.Asn849Lysfs*32	FS	P	(34)
CAIS	Del	7	LBD	c.2562_2563del	2	NR, non-CCC	Non-CCC	p.Arg855Thrfs*25	FS	LP	(53)
CAIS	Ins	7	LBD	c.2573dup	1	NR, non-CCC	CCC	p.Tyr858*	FS	LP	(54)
CAIS	Indel	8	LBD	c.2640_2646delinsTTCCGCCCTGA	17	NR, non-CCC	Non-CCC	p.Leu881Serfs*80	FS	LP	(55)
CAIS	Del	8	LBD	c.2683_2686del	4	NR, non-CCC	CCC	p.Met896del	NFS	LP	(53)
CAIS	Del	8	LBD	c.2715del	1	NR, non-CCC	CCC	Lys906Argfs*38	FS	LP	(28)
CAIS	Del	8	LBD	c.2749_2763del	15	NR, non-CCC	Non-CCC	**	FS	LP	(56)
PAIS	Del	4	LBD	c.2077_2079del	3	NR, non-CCC	Non-CCC	p.Asn693del	NFS	LP	(22)
CAIS	Ins	1	NTD	c.192_193insTAGCAG	6	CCC	CCC	p.Gln65*	NFS	LP	(57)

Abbreviations: ACMG, American College of Medical Genetics and Genomics; CCC, change in copy count; DBD, DNA-binding domain; FS, frameshift; LBD, ligand-binding domain; LP, likely pathogenic; non-CCC, no change in copy; NFS, non-frameshift; NTD, N-terminal domain; NR, nonrepetitive sequence; P, pathogenic; VUS, variants of uncertain significance

revealed a notable overrepresentation of deletions in individuals with AIS, suggesting that small deletions may be more commonly associated with pathogenicity compared to small insertions. Furthermore, we noticed a remarkable enrichment of short coding indels spanning 1 to 3 nucleotides in individuals with AIS, indicating that most pathogenic indels in the AR gene are caused by small variations. The prevalence of small indels affecting the AR gene in individuals with AIS, particularly those ranging from 1 to 3 nucleotides, is noteworthy and implies potential functional implications.

Coding small indels exhibit an enrichment at the N or C terminal regions in the human genome (64). In the context of AIS, we also observed a similar pattern, as most short indels are located at the N-terminal domain (NTD) and ligand-binding domain (LBD), with a particular preponderance at the NTD domain. This preponderance may be attributed to the fact that the NTD contains 2 long trinucleotide repeats (poly-Q and poly-G tracts).

Trinucleotide repeats are sequences of 3 nucleotides that are found in varying copy numbers throughout the human genome (72). Like other repetitive regions in the genome, trinucleotide repeats are prone to slippage events (73). Given that slippage has been implicated in the generation of indels, trinucleotide repeats within exonic regions may represent hotspots for indel formation. In the AR gene, 2 well-known polymorphic trinucleotide repeats exist. The first one is the CAG repeat (poly-Q tract) located within exon 1 (spanning from amino acid 57 to 80; NM_000044.6), and the second one is the CCG repeat (poly-G tract) also within exon 1 (spanning from amino acid 451 to 473). The AR gene has a predicted individual mutation rate exceeding 10^{-5} per generation, making it an excellent gene for analyzing mutational patterns across generations. Based on this, Montgomery et al identified that long trinucleotide repeats within the AR coding region are hotspots for small indels in a healthy population, explaining the polymorphic nature of these poly-repeats in humans (12).

However, intriguingly, we found that only 15.8% of the pathogenic small indels in AIS were within the AR poly-repeats (Fig. 2; D). This suggests that the hotspots for indels within the AR gene observed in the healthy general population do not fully represent the mechanisms underlying the pathogenic indels identified in individuals with AIS.

In addition to the classical poly-Q and poly-G tracts, the AR gene contains 2 other fixed trinucleotide repeats, a second poly-glutamine stretch (amino acids 86 to 91) and a stretch of 8 proline residues (amino acids 374 to 381). Although smaller in size, these 2 poly-repeats have not been extensively studied for their individual roles in androgen receptor activity. However, our analysis revealed the presence of 3 AR small indels within the poly-proline region and 6 within the second poly-glutamine stretch. While trinucleotide repeat regions do represent risk loci for small indel occurrences, they contribute to only a limited proportion of the indels observed in AIS. Nevertheless, the risk of pathogenic indel events within these poly-repeats appears to be independent of their size. Additionally, these small poly-repeats may exert a more significant influence on the androgen receptor's activity beyond their role in the receptor's amino acid sequence. Given the prevalence of triplet repeat poly-repeats in various human genes, understanding the potential impact of these small poly-repeats could be crucial in comprehending the pathogenicity of variants identified in these regions and advancing

our understanding of mutagenesis mechanisms linked to small indels.

Regarding the mechanism of mutagenesis, polymerase slippage has emerged as the predominant force driving short indel generation in the genome (12, 74). This phenomenon occurs during replication and typically results in the contraction or expansion of tandem repeat regions. In the human genome, polymerase slippage explains approximately 75% of events in repetitive hotspot regions and around 50% of indels in non-hotspot areas (19). Notably, about 45% of indels are concentrated in only about 4% of the genome and can largely be attributed to polymerase slippage (75). Sequences prone to polymerase slippage, such as tandem repeat regions and homopolymeric runs, harbor around 70% of indels (12).

Since polymerase slippage frequently occurs in repetitive sequences, CCC events are likely to be caused by this mechanism. In our analysis, we classified each indel event as either CCC or non-CCC events to estimate the proportion of indels in the coding region of the AR gene. Interestingly, we found that only 52% of the AR small indels are non-CCC events, suggesting that other molecular mechanisms may play a role in the generation of these small indels. While these mechanisms are still largely unknown, disturbances in DNA cleavage might be involved in indel generation, as DNA cleavage intensities close to the start and end points of small indels were significantly lower than in other genomic regions. Further research is warranted to elucidate these molecular mechanisms and their contribution to pathogenic small indel events in the AR gene.

In their search for a genomic consensus sequence prone to indel generation, Krawczak and Cooper identified a 6-bp consensus sequence (TG A/G A/G G/T A/C) that frequently occurs at sites of germline deletions in several genetic disorders (76). This consensus sequence has also been identified as a hotspot for indels in the TP53 gene within the codons 173-175 (77). To determine if pathogenic small indels in the AR gene occur within this consensus sequence, we analyzed the genomic context of each AR indel and compared the nucleotide context with the consensus sequence. We identified 3 small indels within the consensus sequence (c.2317_2321del, c.2356_2357del, and c.2478_2479ins). In total, indels at this consensus sequence comprised 3.65% of all AR indels, which is similar to the frequency of indels at this consensus sequence in the TP53 gene.

Homopolymers or homocopolymers are runs of purines or pyrimidines (eg, AAAAAA, CCTTCC, AAAGGG) (77). These regions have been reported as hotspot regions for indels (75). For instance, in the TP53 gene, homopolymer runs of 4 or more nucleotides are sites for indels, suggesting that homopolymer runs are risk regions for indel events, even within coding areas (78). In the AR gene, 7 indels occur at homopolymer or homocopolymer runs, including the novel c.1314_1315delCT (subject 3). In total, indels at these runs comprise 8.5% of the small indels in the AR gene, further supporting the idea that these runs are prone regions for exonic small indel occurrences.

Complex indels involve a combination of deletion and insertion (and sometimes a base substitution) all at once (79). Due to the lack of effective tools for mapping and detecting complex indels with short sequence reads, they have been somewhat neglected since the introduction of next-generation sequencing. Therefore, the frequency, phenotypic features, and mechanisms of mutagenesis of these indels are largely

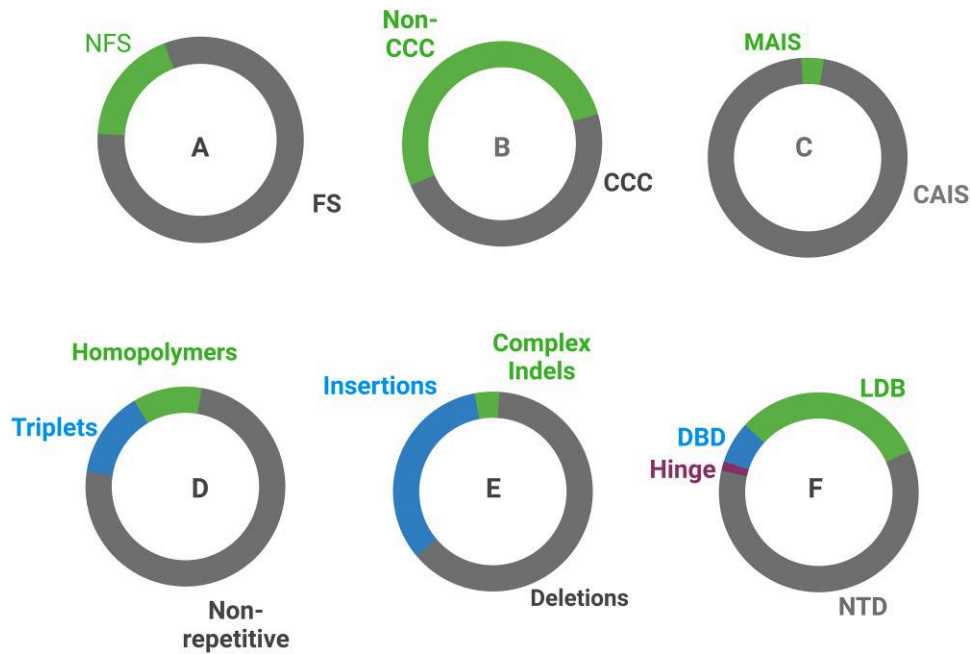


Figure 2. The comprehensive landscape of indels within the androgen receptor gene in androgen insensitivity syndrome (AIS). A) The frequency of non-frameshift (NFS) indels ($n = 18$; 21%) and frameshift (FS) indels ($n = 79$; 79%). B) Indel locations—non-CCC region ($n = 43$, 52%) and CCC region ($n = 39$, 48%). C) AIS phenotypes—complete androgen insensitivity syndrome (CAIS) ($n = 78$; 95%) and mild androgen insensitivity syndrome (MAIS) ($n = 4$, 5%). D) Exonic features of indels—triplet sequences ($n = 13$; 16%), homopolymer runs ($n = 7$; 8.5%), and nonrepetitive sequences ($n = 62$, 70.5%). E) Indel nature—insertions ($n = 29$; 35%), Deletions ($n = 49$, 60%), and complex indels ($n = 4$, 5%). F) Indels distribution—N-terminal domain (NTD) ($n = 45$, 55%), Ligand-binding domain (LBD) ($n = 28$, 34%), Hinge region ($n = 1$), and DNA-binding domain (DBD) ($n = 8$, 10%).

unknown (80). Complex frameshift variants are attributed to palindromic or quasi-palindromic loops, which function as double-stranded DNA (81). Alternatively, complex frameshift indels can occur at sites of tandem repeats (77). In this study, we present a new case of complex frameshift variants in the AR (c.431_452delinsAAA - case 1). Upon examination of all complex frameshift indels in the AR, we identified 4 other occurrences of such variants among individuals with AIS ($n = 5$; 6%). Only one complex frameshift indel (c.157_165dup) (68) was found at tandem repeats, while the remaining 4 were located at palindromic runs, suggesting that these runs are prone to complex indels. Further research is needed to better understand the significance and underlying mechanisms of complex indels in the context of genetic disorders such as AIS.

In conclusion, our study reveals a robust genotype-phenotype relationship concerning small indels in the AR gene and AIS, with the vast majority of cases presenting with CAIS. Trinucleotide repeat regions and homopolymers or homocopolymers in the AR gene represent risk loci for small indel occurrences in individuals with AIS, although they contribute to only a limited proportion of the observed indels. Predominantly, small indels in the AR coding region among AIS involve FS indels, consistent with the general enrichment of FS indels in AR coding regions throughout the human genome. While polymerase slippage may explain approximately half of the AR indels' occurrence, our findings suggest that other molecular mechanisms may also play a role in their generation. Complex frameshift indels in the AR gene are rare but tend to be associated with palindromic runs. These insights contribute to our understanding of the genetic basis of AIS and shed light on potential mechanisms underlying small indel events in the AR gene. Overall, our study underscores the genetic and molecular complexity of this

condition and provides valuable information for further AR research.

Funding

This work was supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), grant 2019/26780-9 to B.B.M.

Disclosures

The authors have nothing to disclose.

Data Availability

The datasets analyzed during the current study are publicly available on HGMD (<https://www.hgmd.cf.ac.uk/>), the androgen receptor mutations database (<http://androgendb.mcgill.ca/>), and the Genome Aggregation Database (gnomAD, <https://gnomad.broadinstitute.org>).

References

- Mullaney JM, Mills RE, Pittard WS, Devine SE. Small insertions and deletions (INDELs) in human genomes. *Hum Mol Genet.* 2010;19(R2):R131-R136.
- Lin M, Whitmire S, Chen J, Farrel A, Shi X, Guo JT. Effects of short indels on protein structure and function in human genomes. *Sci Rep.* 2017;7(1):9313.
- Sergouniotis PI, Barton SJ, Waller S, *et al.* The role of small in-frame insertions/deletions in inherited eye disorders and how structural modelling can help estimate their pathogenicity. *Orphanet J Rare Dis.* 2016;11(1):125.
- Chuzhanova NA, Anassis EJ, Ball EV, Krawczak M, Cooper DN. Meta-analysis of indels causing human genetic disease: mechanisms

- of mutagenesis and the role of local DNA sequence complexity. *Hum Mutat.* 2003;21(1):28-44.
5. Batista RL, Costa EMF, Rodrigues AS, *et al.* Androgen insensitivity syndrome: a review. *Arch Endocrinol Metab.* 2018;62(2):227-235.
 6. Wisniewski AB, Batista RL, Costa EMF, *et al.* Management of 46, XY differences/disorders of sex development (DSD) throughout life. *Endocr Rev.* 2019;40(6):1547-1572.
 7. Ahmed SF, Alimusina M, Batista RL, *et al.* The use of genetics for reaching a diagnosis in XY DSD. *Sex Dev.* 2022;16(2-3):207-224.
 8. Gottlieb B, Beitel LK, Nadarajah A, Paliouras M, Trifiro M. The androgen receptor gene mutations database: 2012 update. *Hum Mutat.* 2012;33(5):887-894.
 9. Wang N, Lysenkov V, Orte K, *et al.* Tool evaluation for the detection of variably sized indels from next generation whole genome and targeted sequencing data. *PLoS Comput Biol.* 2022;18(2): e1009269.
 10. Gagliano SA, Sengupta S, Sidore C, *et al.* Relative impact of indels versus SNPs on complex disease. *Genet Epidemiol.* 2019;43(1): 112-117.
 11. Richards S, Aziz N, Bale S, *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17(5):405-424.
 12. Montgomery SB, Goode DL, Kvikstad E, *et al.* The origin, evolution, and functional impact of short insertion-deletion variants identified in 179 human genomes. *Genome Res.* 2013;23(5): 749-761.
 13. Challis D, Antunes L, Garrison E, *et al.* The distribution and mutagenesis of short coding INDELS from 1,128 whole exomes. *BMC Genomics.* 2015;16(1):143.
 14. Silva JM, Batista RL, De Santi Rodrigues A, *et al.* Androgen receptor mRNA analysis from whole blood: a low-cost strategy for detection of androgen receptor gene splicing defects. *Clin Genet.* 2018;94(5):489-490.
 15. Batista RL, Yamaguchi K, di Santi Rodrigues A, *et al.* Mobile DNA in endocrinology: LINE-1 retrotransposon causing partial androgen insensitivity syndrome. *J Clin Endocrinol Metab.* 2019;104(12):6385-6390.
 16. Mendonca BB, Inacio M, Costa EM, *et al.* Male pseudohermaphroditism due to steroid 5 α -reductase 2 deficiency. Diagnosis, psychological evaluation, and management. *Medicine (Baltimore).* 1996;75(2):64-76.
 17. Thigpen AE, Davis DL, Milatovich A, *et al.* Molecular genetics of steroid 5 α -reductase 2 deficiency. *J Clin Invest.* 1992;90(3): 799-809.
 18. Barbaro M, Oscarson M, Almskog I, Hamberg H, Wedell A. Complete androgen insensitivity without Wolffian duct development: the AR-A form of the androgen receptor is not sufficient for male genital development. *Clin Endocrinol (Oxf).* 2007;66(6): 822-826.
 19. Taylor MS, Ponting CP, Copley RR. Occurrence and consequences of coding sequence insertions and deletions in Mammalian genomes. *Genome Res.* 2004;14(4):555-566.
 20. Jung SC, Lee KS, Koo SK. Human gene mutations. Gene symbol: AR. Disease: androgen insensitivity syndrome. *Hum Genet.* 2004;114(2):222.
 21. Decaestecker K, Philibert P, De Baere E, *et al.* A novel mutation c.118delA in exon 1 of the androgen receptor gene resulting in complete androgen insensitivity syndrome within a large family. *Fertil Steril.* 2008;89(5):1260.e3-e7.
 22. Boehmer AL, Brinkmann O, Brüggewirth H, *et al.* Genotype versus phenotype in families with androgen insensitivity syndrome. *J Clin Endocrinol Metab.* 2001;86(9):4151-4160.
 23. Sinnecker GH, Hiort O, Dibbelt L, *et al.* Phenotypic classification of male pseudohermaphroditism due to steroid 5 α -reductase 2 deficiency. *Am J Med Genet.* 1996;63(1):223-230.
 24. Cheng T, Wang H, Han B, *et al.* Identification of three novel. *Asian J Androl.* 2019;21(6):577.
 25. Philibert P, Audran F, Pienkowski C, *et al.* Complete androgen insensitivity syndrome is frequently due to premature stop codons in exon 1 of the androgen receptor gene: an international collaborative report of 13 new mutations. *Fertil Steril.* 2010;94(2):472-476.
 26. Ferlin A, Vinanzi C, Garolla A, *et al.* Male infertility and androgen receptor gene mutations: clinical features and identification of seven novel mutations. *Clin Endocrinol (Oxf).* 2006;65(5):606-610.
 27. Zhu YS, Cai LQ, Cordero JJ, Canovatchel WJ, Katz MD, Imperato-McGinley J. A novel mutation in the CAG triplet region of exon 1 of androgen receptor gene causes complete androgen insensitivity syndrome in a large kindred. *J Clin Endocrinol Metab.* 1999;84(5):1590-1594.
 28. Audi L, Fernández-Cancio M, Carrascosa A, *et al.* Novel (60%) and recurrent (40%) androgen receptor gene mutations in a series of 59 patients with a 46,XY disorder of sex development. *J Clin Endocrinol Metab.* 2010;95(4):1876-1888.
 29. Gottlieb B, Pinsky L, Beitel LK, Trifiro M. Androgen insensitivity. *Am J Med Genet.* 1999;89(4):210-217.
 30. Rocca MS, Minervini G, Vinanzi C, *et al.* Mutational screening of androgen receptor gene in 8222 men of infertile couples. *J Clin Endocrinol Metab.* 2022;108(5):1181-1191.
 31. Zhu H, Yao H, Xu Y, *et al.* Phenotypic and biochemical characteristics and molecular basis in 36 Chinese patients with androgen receptor variants. *Orphanet J Rare Dis.* 2021;16(1):122.
 32. Batch JA, Williams DM, Davies HR, *et al.* Androgen receptor gene mutations identified by SSCP in fourteen subjects with androgen insensitivity syndrome. *Hum Mol Genet.* 1992;1(7):497-503.
 33. Jeske YW, McGown IN, Cowley DM, *et al.* Androgen receptor genotyping in a large Australasian cohort with androgen insensitivity syndrome; identification of four novel mutations. *J Pediatr Endocrinol Metab.* 2007;20(8):893-908.
 34. Ahmed SF, Cheng A, Dovey L, *et al.* Phenotypic features, androgen receptor binding, and mutational analysis in 278 clinical cases reported as androgen insensitivity syndrome. *J Clin Endocrinol Metab.* 2000;85(2):658-665.
 35. Hiort O, Wodtke A, Struve D, Zöllner A, Sinnecker GH. Detection of point mutations in the androgen receptor gene using non-isotopic single strand conformation polymorphism analysis. German Collaborative Intersex Study Group. *Hum Mol Genet.* 1994;3(7): 1163-1166.
 36. Brüggewirth HT, Boehmer AL, Verleun-Mooijman MC, *et al.* Molecular basis of androgen insensitivity. *J Steroid Biochem Mol Biol.* 1996;58(5-6):569-575.
 37. Liu Q, Yin X, Li P. Clinical characteristics, AR gene variants, and functional domains in 64 patients with androgen insensitivity syndrome. *J Endocrinol Invest.* 2023;46(1):151-158.
 38. Thiele B, Weidemann W, Schnabel D, Romalo G, Schweikert HU, Spindler KD. Complete androgen insensitivity caused by a new frameshift deletion of two base pairs in exon 1 of the human androgen receptor gene. *J Clin Endocrinol Metab.* 1999;84(5): 1753-1753.
 39. He X, Ma Q, Zhang Q, *et al.* Case report: identification of a frameshift mutation in GC enrichment and the GCC repeat region of the androgen insensitivity receptor (AR) gene in a patient with complete androgen insensitivity syndrome by whole-exome sequencing (WES) combined with specific PCR and deep sequencing. *Front Genet.* 2022;13:1038997.
 40. Imai A, Ohno T, Iida K, Ohsuye K, Okano Y, Tamaya T. A frameshift mutation of the androgen receptor gene in a patient with receptor-negative complete testicular feminization: comparison with a single base substitution in a receptor-reduced incomplete form. *Ann Clin Biochem.* 1995;32(5):482-486.
 41. Beitel LK, Prior L, Vasilou DM, *et al.* Complete androgen insensitivity due to mutations in the probable alpha-helical segments of the DNA-binding domain in the human androgen receptor. *Hum Mol Genet.* 1994;3(1):21-27.
 42. Ledig S, Jakubiczka S, Neulen J, *et al.* Novel and recurrent mutations in patients with androgen insensitivity syndromes. *Horm Res.* 2005;63(6):263-269.

43. Chen CP, Chen SR, Wang TY, Wang W, Hwu YM. A frame shift mutation in the DNA-binding domain of the androgen receptor gene associated with complete androgen insensitivity, persistent Müllerian structures, and germ cell tumors in dysgenetic gonads. *Fertil Steril.* 1999;72(1):170-173.
44. Yue L, Wu P, Xia Z, Fan C, Xia Q. [A novel deletion mutation in AR gene causes complete androgen insensitivity syndrome in a Chinese family]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi.* 2010;27(6):631-633.
45. Rong HL, Suzuki N, Imai A. A double nucleotide insertion-induced frame-shift mutation of the androgen receptor gene in a familial complete androgen insensitivity syndrome. *Eur J Obstet Gynecol Reprod Biol.* 2010;148(1):53-55.
46. Lobaccaro JM, Lumbroso S, Poujol N, et al. Complete androgen insensitivity syndrome due to a new frameshift deletion in exon 4 of the androgen receptor gene: functional analysis of the mutant receptor. *Mol Cell Endocrinol.* 1995;111(1):21-28.
47. Galani A, Sofocleous C, Karahaliou F, Papathanasiou A, Kitsiou-Tzeli S, Kalpini-Mavrou A. Sex-reversed phenotype in association with two novel mutations c.2494delA and c.T3004C in the ligand-binding domain of the androgen receptor gene. *Fertil Steril.* 2008;90(5):2008.e1-4.
48. Avila DM, Wilson CM, Nandi N, Griffin JE, McPhaul MJ. Immunoreactive AR and genetic alterations in subjects with androgen resistance and undetectable AR levels in genital skin fibroblast ligand-binding assays. *J Clin Endocrinol Metab.* 2002;87(1):182-188.
49. Hannema SE, Scott IS, Hodapp J, et al. Residual activity of mutant androgen receptors explains wolffian duct development in the complete androgen insensitivity syndrome. *J Clin Endocrinol Metab.* 2004;89(11):5815-5822.
50. Baldazzi L, Baroncini C, Pirazzoli P, et al. Two mutations causing complete androgen insensitivity: a frame-shift in the steroid binding domain and a Cys→Phe substitution in the second zinc finger of the androgen receptor. *Hum Mol Genet.* 1994;3(7):1169-1170.
51. Köhler B, Lumbroso S, Leger J, et al. Androgen insensitivity syndrome: somatic mosaicism of the androgen receptor in seven families and consequences for sex assignment and genetic counseling. *J Clin Endocrinol Metab.* 2005;90(1):106-111.
52. MacLean HE, Ball EM, Rekaris G, Warne GL, Zajac JD. Novel androgen receptor gene mutations in Australian patients with complete androgen insensitivity syndrome. *Hum Mutat.* 2004;23(3):287.
53. Mueller OT, Root A. Gene symbol: AR. Disease: androgen insensitivity syndrome. *Hum Genet.* 2006;119(6):672.
54. Turek-Plewa J, Jerzy S, Wieslaw HT. Novel human pathological mutations. Gene symbol: AR. Disease: androgen insensitivity syndrome. *Hum Genet.* 2009;125(3):341.
55. Vilchis F, Ramos L, Kofman-Alfaro S, Zenteno JC, Méndez JP, Chávez B. Extreme androgen resistance in a kindred with a novel insertion/deletion mutation in exon 5 of the androgen receptor gene. *J Hum Genet.* 2003;48(7):346-351.
56. Werner R, Zhan J, Gesing J, Struve D, Hiort O. In-vitro characterization of androgen receptor mutations associated with complete androgen insensitivity syndrome reveals distinct functional deficits. *Sex Dev.* 2008;2(2):73-83.
57. Liu S, Wang Z, Jiang J, et al. Severe forms of complete androgen insensitivity syndrome caused by a p.Q65X novel mutation in androgen receptor: clinical manifestations, imaging findings and molecular genetics. *Steroids.* 2019;144:47-51.
58. Zhang F, Gu W, Hurler ME, Lupski JR. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet.* 2009;10(1):451-481.
59. Mills RE, Pittard WS, Mullaney JM, et al. Natural genetic variation caused by small insertions and deletions in the human genome. *Genome Res.* 2011;21(6):830-839.
60. Ueki M, Kimura-Kataoka K, Fujihara J, et al. Evaluation of the functional effects of genetic variants—missense and nonsense SNPs, indels and copy number variations—in the gene encoding human deoxyribonuclease I potentially implicated in autoimmunity. *Sci Rep.* 2019;9(1):13660.
61. Zhen Q, Yang Z, Wang W, et al. Genetic study on small insertions and deletions in psoriasis reveals a role in Complex human diseases. *J Invest Dermatol.* 2019;139(11):2302-12.e14.
62. Banjar H, Moghrabi N, Alotaibi T, Alotaibi S, Gamalmaz H. The first case report of double homozygous of 2 different mutations in the. *Int J Pediatr Adolesc Med.* 2017;4(1):41-46.
63. Hornig NC, Holterhus PM. Molecular basis of androgen insensitivity syndromes. *Mol Cell Endocrinol.* 2021;523:111146.
64. Tyutyusheva N, Mancini I, Baroncelli GI, et al. Complete androgen insensitivity syndrome: from bench to bed. *Int J Mol Sci.* 2021;22(3):1264.
65. Gulía C, Baldassarra S, Zangari A, et al. Androgen insensitivity syndrome. *Eur Rev Med Pharmacol Sci.* 2018;22(12):3873-3887.
66. Hughes IA, Davies JD, Bunch TL, Pasterski V, Mastroyannopoulou K, MacDougall J. Androgen insensitivity syndrome. *Lancet.* 2012;380(9851):1419-1428.
67. Batista RL, Rodrigues AS, Nishi MY, et al. Heterozygous nonsense mutation in the androgen receptor gene associated with partial androgen insensitivity syndrome in an individual with 47, XXY karyotype. *Sex Dev.* 2017;11(2):78-81.
68. Chen YH, Xu HY, Wang ZY, et al. An insertion mutation in the androgen receptor gene in a patient with azoospermia. *Asian J Androl.* 2015;17(5):857-858.
69. Lund A, Juvonen V, Lähdetie J, Aittomäki K, Tapanainen JS, Savontaus ML. A novel sequence variation in the transactivation regulating domain of the androgen receptor in two infertile Finnish men. *Fertil Steril.* 2003;79 (Suppl 3):1647-1648.
70. Nguyen LS, Wilkinson MF, Geck J. Nonsense-mediated mRNA decay: inter-individual variability and human disease. *Neurosci Biobehav Rev.* 2014;46 (Pt 2):175-186.
71. Bermejo-Das-Neves C, Nguyen HN, Poch O, Thompson JD. A comprehensive study of small non-frameshift insertions/deletions in proteins and prediction of their phenotypic effects by a machine learning method (KD4i). *BMC Bioinformatics.* 2014;15(1):111.
72. Kozłowski P, de Mezer M, Krzyżosiak WJ. Trinucleotide repeats in human genome and exome. *Nucleic Acids Res.* 2010;38(12):4027-4039.
73. Xu P, Pan F, Roland C, Sagui C, Weninger K. Dynamics of strand slippage in DNA hairpins formed by CAG repeats: roles of sequence parity and trinucleotide interrupts. *Nucleic Acids Res.* 2020;48(5):2232-2245.
74. Ellegren H. Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet.* 2000;16(12):551-558.
75. Nesta AV, Tafur D, Beck CR. Hotspots of human mutation. *Trends Genet.* 2021;37(8):717-729.
76. Krawczak M, Reiss J, Cooper DN. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet.* 1992;90(1-2):41-54.
77. Greenblatt MS, Grollman AP, Harris CC. Deletions and insertions in the p53 tumor suppressor gene in human cancers: confirmation of the DNA polymerase slippage/misalignment model. *Cancer Res.* 1996;56(9):2130-2136.
78. Redston MS, Caldas C, Seymour AB, et al. P53 mutations in pancreatic carcinoma and evidence of common involvement of homopolymer tracts in DNA microdeletions. *Cancer Res.* 1994;54(11):3025-3033.
79. Ripley LS. Frameshift mutation: determinants of specificity. *Annu Rev Genet.* 1990;24(1):189-213.
80. Ye K, Wang J, Jayasinghe R, et al. Systematic discovery of complex insertions and deletions in human cancers. *Nat Med.* 2016;22(1):97-104.
81. Levine JG, Schaaper RM, DeMarini DM. Complex frameshift mutations mediated by plasmid pKM101: mutational mechanisms deduced from 4-aminobiphenyl-induced mutation spectra in Salmonella. *Genetics.* 1994;136(3):731-746.