MAPEAMENTO DE POLIMORFISMOS (SNPs) ASSOCIADOS À INFERTILIDADE MASCULINA IDIOPÁTICA: UMA ABORDAGEM BIOINFORMÁTICA

Larissa de Gouvêa Santos Rufino; Leonardo Agostinho de Sales; Bruno Henrique Godoi

Universidade de Taubaté, Avenida Tiradentes, 500 - Bom Conselho - 12030-180 - Taubaté-SP, Brasil, larissa.gsrufino@unitau.br

Resumo
A infertilidade masculina idiopática apresenta forte componente genético ainda pouco elucidado. Este estudo mapeou, in silico, polimorfismos missense em nove genes ligados à espermatogênese e à motilidade espermática (CATSPER1-4, DNAH1, DNAH17, SPEF2, TEX11 e DAZ2), minerados do NCBI dbSNP, do Ensembl e do UniProt. A triagem por SIFT e PolyPhen-2 classificou 4.728 das 10.837 variantes analisadas como Tier 1, concentradas majoritariamente em SPEF2 e CATSPER2. Após filtragem adicional por conservação filogenética (≥ 75%) e confiança estrutural do AlphaFold (pLDDT ≥ 70), 17 variantes-âncora dos genes CATSPER1, CATSPER2, CATSPER3 e DNAH17 foram submetidas à modelagem 3D e à mutagênese in silico no PyMOL, revelando alterações estruturais compatíveis com os fenótipos de astenozoospermia associados. Por utilizar exclusivamente dados públicos, o trabalho dispensa submissão a Comitê de Ética em Pesquisa (CEP).

Palavras-chave: Reprodução humana. Infertilidade masculina. SNPs. Bioinformática genômica. Genética.

Área do Conhecimento: Ciências Biológicas e da Saúde.

Introdução

A infertilidade é um problema de saúde pública que afeta aproximadamente 15% dos casais em idade reprodutiva em todo o mundo, sendo o fator masculino responsável por cerca de 40 a 50% dos casos (Agarwal et al., 2015). Apesar dos avanços nos métodos diagnósticos, uma parcela significativa dos casos ainda é classificada como idiopática, ou seja, sem causa definida, o que evidencia a necessidade de investigar os fatores genéticos envolvidos nesse processo. Nesse contexto, estudos têm demonstrado que alterações genéticas, especialmente os polimorfismos de nucleotídeo único (Single Nucleotide Polymorphisms – SNPs), podem influenciar diretamente a fertilidade masculina, afetando genes envolvidos na espermatogênese, na motilidade e na função espermática (Krausz; Riera-Escamilla, 2018).

Entre os genes mais estudados nesse contexto, destacam-se os membros da família CATSPER (CATSPER1, CATSPER2 e CATSPER3). Eles formam o canal de cálcio dependente de voltagem específico do espermatozoide, essencial para a hiperativação flagelar e para a motilidade progressiva (Qi et al., 2019). Destaca-se, ainda, o gene DNAH17, que codifica uma cadeia pesada de dineína axonemal do braço externo, estrutura motora responsável pelo batimento do flagelo (Whitfield et al., 2019). Variantes patogênicas nesses genes já foram associadas a quadros de astenozoospermia e à desorganização estrutural do axonema. No caso específico do gene CATSPER2, deleções contíguas que o acometem em conjunto com o gene STRC foram associadas, ainda, à síndrome de surdez-infertilidade (Avidan et al., 2003).

Com o avanço das ferramentas de bioinformática, tornou-se possível analisar essas variações genéticas a partir de dados disponíveis em bancos públicos, como o NCBI dbSNP, o Ensembl e o UniProt. Essas plataformas permitem a identificação de SNPs já descritos, enquanto ferramentas computacionais como o SIFT e o PolyPhen-2 possibilitam prever o impacto funcional dessas alterações sobre a estrutura das proteínas, contribuindo para a compreensão de suas possíveis consequências biológicas (Ng; Henikoff, 2003; Adzhubei et al., 2010). Entretanto, muitos estudos ainda se limitam à identificação dessas variantes, sem explorar de forma aprofundada seu impacto funcional e estrutural. Essas evidências reforçam a importância de uma análise integrada dos SNPs, uma vez que nem todas as variações genéticas resultam em alterações fenotípicas relevantes. Dessa forma, o presente estudo tem como objetivo identificar e caracterizar, por meio de ferramentas de bioinformática estrutural, polimorfismos missense associados à infertilidade masculina nos genes CATSPER1, CATSPER2, CATSPER3 e DNAH17, avaliando sua conservação evolutiva, seu impacto funcional previsto e suas consequências sobre a estrutura tridimensional das respectivas proteínas.

Metodologia

Este estudo caracterizou-se como uma pesquisa descritiva e analítica de abordagem computacional (in silico), desenvolvida a partir de dados genômicos e estruturais secundários de acesso público. Inicialmente, realizou-se uma busca direcionada na literatura para selecionar genes fundamentais à espermatogênese, à hiperativação flagelar e à motilidade espermática, contemplando nove genes-alvo: a família dos canais de cálcio espermáticos (CATSPER1, CATSPER2, CATSPER3 e CATSPER4), motores moleculares axonemais e flagelares (DNAH1, DNAH17, SPEF2 e TEX11) e o gene associado à maturação espermática DAZ2. Em seguida, mineraram-se variantes não sinônimas (missense) nesses nove alvos a partir de bases de dados integradas (NCBI dbSNP, Ensembl e UniProt), estruturando-se uma matriz mestra de variantes.

Para a triagem funcional e a priorização patogênica, aplicou-se uma avaliação computacional combinada, utilizando os algoritmos SIFT (ponto de corte ≤ 0,05 para predição de efeito deletério) e PolyPhen-2 (ponto de corte ≥ 0,85 para dano proteico provável) (Ng; Henikoff, 2003; Adzhubei et al., 2010). A partir da concordância entre os dois preditores, o conjunto total de variantes minerado foi estratificado em três níveis de evidência: Tier 1, com concordância deletéria entre os dois preditores; Tier 2, de impacto moderado a alto; e Tier 3, de impacto indeterminado ou baixo.

Analisou-se, ainda, a conservação filogenética dos sítios mutados por meio de alinhamentos ortólogos comparativos entre vertebrados, incluindo, conforme a disponibilidade de sequência para cada gene, *Homo sapiens*, *Pan troglodytes*, *Macaca mulatta, Bos taurus*, *Rattus norvegicus*, *Mus musculus*, *Canis lupus familiaris*, *Danio rerio* e *Gallus gallus*. Categorizaram-se como altamente conservadas as posições com taxa de identidade ≥ 75%. Parâmetros biofísicos de variação de massa molecular (ΔMassa) e de coeficiente de hidropatia de Kyte-Doolittle (ΔHidropatia) foram quantificados para cada substituição de resíduo. Dentro do conjunto Tier 1, selecionou-se um subconjunto de 17 variantes-âncora, concentradas nos genes CATSPER1, CATSPER2, CATSPER3 e DNAH17 por sua relevância mecanística já estabelecida na sinalização de cálcio flagelar e na motilidade axonemal, exigindo-se cumulativamente conservação filogenética ≥ 75% e confiança estrutural pLDDT ≥ 70 para essa etapa de modelagem aprofundada.

A modelagem estrutural tridimensional foi realizada com coordenadas de predição de alta confiança obtidas do AlphaFold Database (EMBL-EBI), com base nos identificadores UniProt e filtradas pelo escore local de qualidade preditiva (pLDDT ≥ 70) (Jumper et al., 2021). A mutagênese in silico foi conduzida de forma automatizada em linguagem Python integrada à biblioteca computacional do PyMOL (Schrödinger, LLC, 2020). O microambiente do sítio mutado foi delimitado em um raio de 5 Å, avaliando-se o impedimento estérico (strain) entre rotâmeros, a alteração na rede de pontes de hidrogênio e a perturbação estérica local. Por fim, construíram-se pranchas integradas de alta resolução (300 DPI), reunindo alinhamento evolutivo, renderização 3D (ray tracing) e ficha biofísica, por meio de scripts personalizados com as bibliotecas Pandas e Matplotlib. Por utilizar exclusivamente dados públicos pré-existentes e desprovidos de identificação humana, a pesquisa prescindiu de submissão a Comitê de Ética em Pesquisa (CEP), em conformidade com a Resolução CNS nº 510/2016.

Resultados

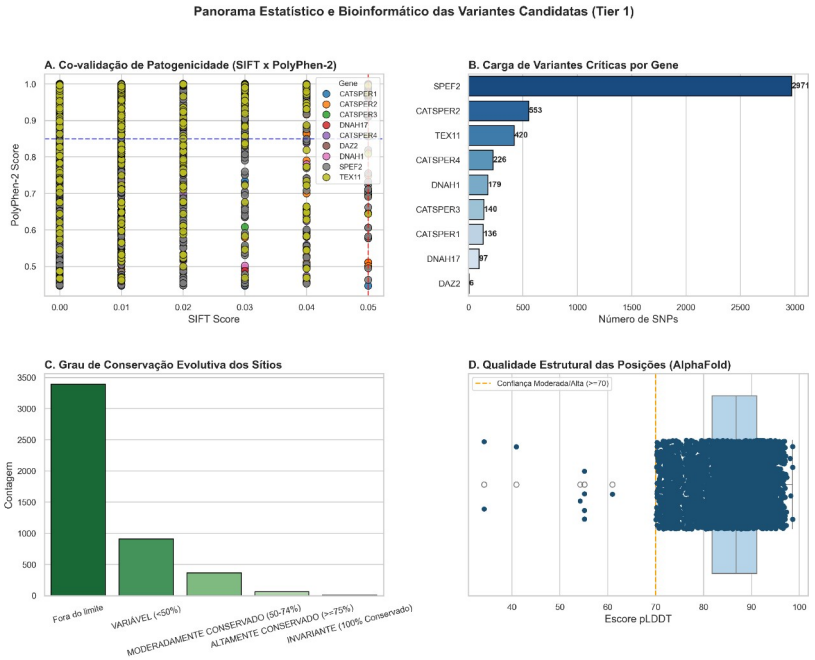
A mineração inicial de dados genômicos nos nove genes-alvo (CATSPER1, CATSPER2, CATSPER3, CATSPER4, DNAH1, DNAH17, SPEF2, TEX11 e DAZ2) resultou em uma matriz mestra de 10.837 variantes missense anotadas no NCBI dbSNP. A aplicação do filtro combinado de patogenicidade (SIFT e PolyPhen-2) estratificou esse conjunto em três níveis de evidência: Tier 1, com concordância deletéria entre os dois preditores (n = 4.728; 43,6%); Tier 2, de impacto moderado a alto (n = 3.233; 29,8%); e Tier 3, de impacto indeterminado ou baixo (n = 2.876; 26,5%).

A distribuição das variantes Tier 1 por gene-alvo mostrou-se marcadamente desigual: SPEF2 concentrou o maior número de candidatas (n = 2.971; 62,8% do Tier 1), seguido por CATSPER2 (n = 553), TEX11 (n = 420), CATSPER4 (n = 226), DNAH1 (n = 179), CATSPER3 (n = 140), CATSPER1 (n = 136), DNAH17 (n = 97) e DAZ2 (n = 6). Essa disparidade reflete, em parte, o tamanho e o grau de polimorfismo de cada gene, não devendo ser interpretada isoladamente como indicador de relevância biológica (Figura 1).

Quanto à conservação evolutiva dos sítios classificados como Tier 1, a maior parte das posições não pôde ser integralmente avaliada por ausência de sequências ortólogas suficientes (categoria “fora do limite”, predominante) ou apresentou conservação variável (< 50%), enquanto apenas uma fração restrita atingiu os patamares de conservação moderada (50–74%) ou alta (≥ 75%). Em contrapartida, a distribuição dos escores de confiança estrutural do AlphaFold (pLDDT) mostrou-se favorável: a maioria das posições mutadas situou-se em regiões de

alta a altíssima confiança (pLDDT > 80), indicando que, mesmo quando a conservação entre espécies era limitada ou não computável, as variantes recaíam predominantemente sobre domínios estruturados e bem resolvidos pelo modelo, e não sobre regiões desordenadas.

Figura 1 - Panorama estatístico e bioinformático das variantes Tier 1.



Fonte: elaborado pelo próprio autor. (A) Correlação SIFT vs. PolyPhen-2 para as variantes deletérias pelo SIFT, coloridas por gene; (B) carga de variantes Tier 1 por gene-alvo; (C) grau de conservação evolutiva dos sítios; (D) distribuição dos escores pLDDT do AlphaFold.

Aplicando-se, sobre o conjunto Tier 1, os critérios adicionais de conservação e confiança estrutural descritos na Metodologia, obteve-se o subconjunto final de 17 variantes-âncora (Tabela 1), todas com predição SIFT deletéria (0,00–0,04) e conservação entre 75,0% e 87,5% nas espécies analisadas.

Tabela 1 - Variantes de nucleotídeo único (SNPs) selecionadas para caracterização estrutural aprofundada.

Gene	rsID	Variante	SIFT	PolyPhen-2	pLDDT	Conservação	ΔHidrop.
CATSPER1	rs777914787	p.Phe756Val	0,00	0,917	82,44	75,0%	+1,4
CATSPER2	rs745482778	p.Asn121Tyr	0,00	0,992	91,62	83,3%	+2,2
CATSPER2	rs1326039812	p.Asn121Thr	0,00	0,954	91,62	83,3%	+2,8
CATSPER2	rs1468568874	p.Phe243Cys	0,00	0,996	89,94	83,3%	−0,3
CATSPER3	rs1196929204	p.Leu183Arg	0,00	0,999	92,88	85,7%	−8,3
CATSPER3	rs766050220	p.Phe185Leu	0,04	0,995	95,94	85,7%	+1,0
CATSPER3	rs1371719244	p.Phe218Ile	0,00	0,998	95,62	85,7%	+1,7
CATSPER3	rs1446392869	p.Thr248Asn	0,00	0,999	94,44	85,7%	−2,8
CATSPER3	rs1554069590	p.Leu254Pro	0,00	0,968	96,12	85,7%	−5,4
CATSPER3	rs1349612854	p.Val266Ala	0,00	0,997	86,00	85,7%	−2,4
DNAH17	rs752654814	p.Lys23Thr	0,01	0,471	91,06	75,0%	+3,2
DNAH17	rs760469852	p.Leu440Met	0,00	0,738	80,69	75,0%	−1,9
DNAH17	rs1448797465	p.Glu441Lys	0,00	0,829	84,75	87,5%	−0,4
DNAH17	rs778904420	p.Gly447Asp	0,00	0,670	90,62	87,5%	−3,1
DNAH17	rs778904420	p.Gly447Val	0,00	0,711	90,62	87,5%	+4,6
DNAH17	rs1243598044	p.Gly447Ser	0,01	0,488	90,62	87,5%	−0,4
DNAH17	rs764686433	p.Leu453Arg	0,00	0,973	91,50	87,5%	−8,3

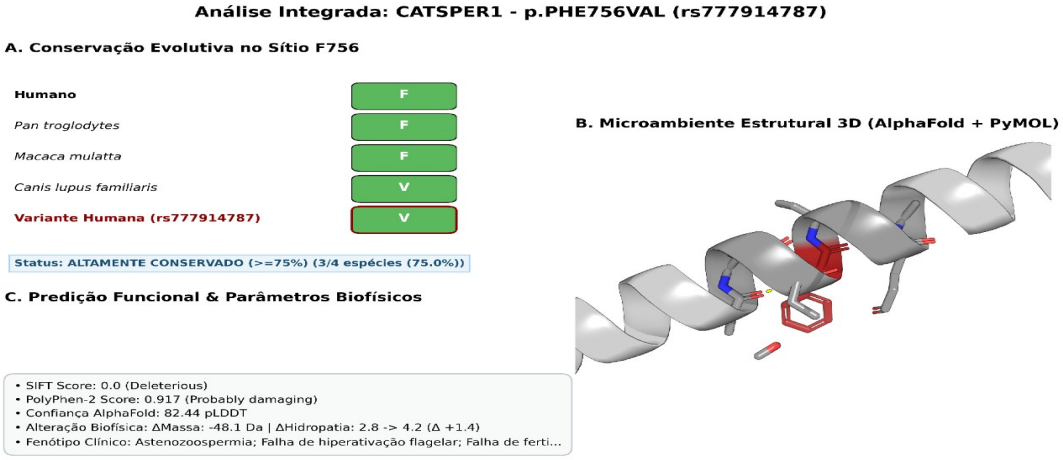
Fonte: elaborado pelo próprio autor.

Dentro desse subconjunto final (Tabela 1), a concordância entre os dois preditores de patogenicidade foi completa para as variantes dos genes CATSPER1, CATSPER2 e CATSPER3, todas classificadas como

"provavelmente danosas" pelo PolyPhen-2 (escore ≥ 0,92). Já em DNAH17, seis das sete variantes apresentaram escores de PolyPhen-2 na faixa de "possivelmente danoso" (0,47–0,83), a despeito do escore SIFT deletério e da alta conservação evolutiva, evidenciando divergência pontual entre os preditores nesse gene e reforçando o valor complementar da validação estrutural adotada neste estudo.

A fim de ilustrar o nível de detalhe estrutural obtido para cada uma das 17 variantes-âncora, a Figura 2 apresenta a análise integrada da variante mais crítica do canal CatSper, p.Phe756Val (CATSPER1, rs777914787).

Figura 2 - Análise integrada da variante p.Phe756Val (CATSPER1, rs777914787).



Fonte: elaborado pelo próprio autor. (A) Conservação evolutiva do sítio F756 entre quatro espécies; (B) microambiente estrutural 3D no PyMOL (AlphaFold), com o resíduo mutado em vermelho e os contatos em um raio de 5 Å; (C) predição funcional e parâmetros biofísicos.

Discussão

Os resultados obtidos com o SIFT e PolyPhen-2 concordam que as mutações identificadas em CATSPER1, CATSPER2 e CATSPER3 são prejudiciais, conclusão reforçada pela alta conservação evolutiva dessas regiões até 85,7% entre as espécies. Na variante p.Phe756Val de CATSPER1, a substituição de um aminoácido volumoso, fenilalanina, por outro menor, valina, ocorre diretamente no poro do canal (Figura 2B). Essa troca aumenta a hidropatia local e desfaz interações importantes com resíduos vizinhos. Como os canais CatSper controlam a entrada de cálcio necessária para a cauda do espermatozoide se mover com vigor, hiperativação, danos nessa região explicam a baixa motilidade, astenozoospermia, e a falha de hiperativação relatadas por Qi et al. (2019). Do ponto de vista clínico, esse resultado corrobora os achados de Avenarius et al. (2009) em homens com infertilidade. Contudo, os mecanismos propostos são distintos: enquanto Avenarius observou a ausência total da proteína, este estudo identifica um defeito pontual e mais sutil na arquitetura do poro. Esse modelo é respaldado pela estrutura tridimensional descrita por criomicroscopia eletrônica, validando as previsões estruturais feitas com o AlphaFold (Lin et al., 2021).

A ausência de ensaios funcionais voltados às variantes identificadas em CATSPER2 e CATSPER3 limita comparações diretas. Ainda assim, nossos achados alinham-se à revisão de Hildebrand et al. (2010), que demonstrou como variantes de perda de função nos genes CATSPER1–4 convergem para o comprometimento da hiperativação flagelar e da fertilidade masculina, independentemente do tipo específico de mutação. A importância clínica de CATSPER2 já havia sido evidenciada por Avidan et al. (2003) na síndrome de surdez-infertilidade, decorrente de deleções gênicas contíguas. Contudo, diferentemente da perda estrutural ampla descrita por esses autores, as variantes missense aqui priorizadas (p.Asn121Tyr, p.Asn121Thr e p.Phe243Cys) apontam para um mecanismo distinto, baseado em perturbações conformacionais pontuais na proteína.

Entre as variantes de DNAH17, destaca-se a p.Leu453Arg. A substituição do resíduo hidrofóbico (leucina) por um resíduo positivamente carregado (arginina) inverte o caráter eletrostático local (com variação de hidropatia de +3,8 para −4,5) e desestabiliza a rede de ligações de hidrogênio observada no PyMOL. Essa alteração tende a comprometer o empacotamento da cadeia pesada de dineína axonemal do braço externo, maquinário motor indispensável ao batimento flagelar coordenado (Whitfield et al., 2019). Esse achado alinha-se aos dados de Zhang

et al. (2020) em modelo murino, no qual uma variante missense homozigótica em DNAH17 desestabilizou os dupletos de microtúbulos 4–7 e resultou em astenozoospermia isolada — sem as múltiplas anormalidades morfológicas do flagelo (MMAF) comuns a outras dineínas axonemais. Essa consonância fenotípica fundamenta a classificação clínica proposta para as variantes de DNAH17 (Tabela 1). Ademais, seis das sete variantes avaliadas em DNAH17, incluindo as três no sítio Gly447, apresentaram escores de PolyPhen-2 mais moderados (0,47–0,83) em relação às alterações em CATSPER1–3, a despeito de apresentarem nível equivalente de conservação evolutiva. Esse perfil sugere que a região de Gly447 tolera certas acomodações estruturais mesmo sob forte restrição evolutiva. O achado evidencia a importância da modelagem estrutural individualizada, capaz de revelar nuances biofísicas que preditores baseados unicamente em alinhamento de sequência não conseguem capturar de forma isolada.

Em termos quantitativos, destaca-se o fato de os genes com maior volume de variantes Tier 1 na triagem inicial (SPEF2 e TEX11) não terem sido submetidos à modelagem estrutural aprofundada. Essa escolha metodológica priorizou alvos cujo papel na motilidade flagelar já se encontra bem estabelecido, a exemplo do complexo CatSper e da dineína axonemal DNAH17. Essa estratégia encontra respaldo em estudos populacionais: Behvarz et al. (2022), ao investigarem polimorfismos em CATSPER1 e TEX11 em homens inférteis, confirmaram associação significativa apenas para o primeiro. Esse contraste evidencia que um elevado número de variantes sinalizadas por preditores computacionais nem sempre se traduz em relevância clínica imediata. Tal constatação não invalida os 420 candidatos Tier 1 identificados em TEX11 nem os 2.971 em SPEF2, mas reforça a necessidade de validações populacionais e funcionais complementares antes de uma triagem estrutural individualizada. Essa etapa pendente configura, portanto, uma limitação do estudo e traça uma perspectiva clara para investigações futuras.

Conclusão

Este estudo demonstrou a viabilidade de uma abordagem bioinformática integrada, combinando triagem por patogenicidade (SIFT e PolyPhen-2), conservação filogenética e modelagem estrutural tridimensional (AlphaFold e PyMOL), para priorizar, dentre milhares de polimorfismos catalogados em bancos públicos, um conjunto reduzido e biologicamente plausível de variantes candidatas à infertilidade masculina idiopática. As 17 variantes-âncora identificadas nos genes CATSPER1, CATSPER2, CATSPER3 e DNAH17 apresentam evidência convergente de dano funcional e estrutural, compatível com os fenótipos de astenozoospermia relatados na literatura, e constituem candidatas promissoras a biomarcadores diagnósticos. Estudos futuros deverão estender a modelagem estrutural aos demais genes triados (SPEF2, TEX11, CATSPER4, DNAH1 e DAZ2) e buscar validação experimental funcional das variantes aqui priorizadas.

Referências

ADZHUBEI, I. A. et al. A method and server for predicting damaging missense mutations. **Nature Methods**, v. 7, n. 4, p. 248–249, 2010.

AGARWAL, A.; MULGUND, A.; HAMADA, A.; CHYATTE, M. R. A unique view on male infertility around the globe. **Reproductive Biology and Endocrinology**, v. 13, p. 37, 2015.

AVENARIUS, M. R. et al. Human male infertility caused by mutations in the CATSPER1 channel protein. **American Journal of Human Genetics**, v. 84, n. 4, p. 505–510, 2009.

AVIDAN, N. et al. CATSPER2, a human autosomal nonsyndromic male infertility gene. **European Journal of Human Genetics**, v. 11, p. 497–502, 2003.

BEHVARZ, M.; RAHMANI, S. A.; SIASI TORBATI, E.; DANAEI MEHRABAD, S.; BIKHOF TORBATI, M. Association of CATSPER1, SPATA16 and TEX11 genes polymorphism with idiopathic azoospermia and oligospermia risk in Iranian population. **BMC Medical Genomics**, v. 15, artigo 47, 2022.

CONSELHO NACIONAL DE SAÚDE (Brasil). Resolução nº 510, de 7 de abril de 2016. Dispõe sobre as normas aplicáveis a pesquisas em Ciências Humanas e Sociais. Diário Oficial da União: seção 1, Brasília, DF, 24 maio 2016.

HILDEBRAND, M. S. et al. Genetic male infertility and mutation of CATSPER ion channels. **European Journal of Human Genetics**, v. 18, n. 11, p. 1178–1184, 2010.

JUMPER, J. et al. Highly accurate protein structure prediction with AlphaFold. **Nature**, v. 596, n. 7873, p. 583–589, 2021.

KRAUSZ, C.; RIERA-ESCAMILLA, A. Genetics of male infertility. **Nature Reviews Urology**, v. 15, p. 369–384, 2018.

LIN, S.; KE, M.; ZHANG, Y.; YAN, Z.; WU, J. Structure of a mammalian sperm cation channel complex. **Nature**, v. 595, n. 7869, p. 746–750, 2021.

NG, P. C.; HENIKOFF, S. SIFT: Predicting amino acid changes that affect protein function. **Nucleic Acids Research**, v. 31, n. 13, p. 3812–3814, 2003.

QI, H. et al. CatSper channels in sperm function and male fertility. **Cell Calcium**, v. 80, p. 72–80, 2019.

SCHRÖDINGER, LLC. The PyMOL Molecular Graphics System, Version 2.5. New York, 2020.

WHITFIELD, M. et al. Mutations in DNAH17, encoding a sperm-specific axonemal outer dynein arm heavy chain, cause isolated male infertility due to asthenozoospermia. **American Journal of Human Genetics**, v. 105, n. 1, p. 198–212, 2019.

ZHANG, B. et al. A DNAH17 missense variant causes flagella destabilization and asthenozoospermia. **Journal of Experimental Medicine**, v. 217, e20182365, 2020.