X-linked Adrenoleukodystrophy: Phenotype-genotype correlation

Andreia Filipa Pereira Amaro
Master's in Forensic Genetics
Department of Biology
2023

Supervisor
Luísa Azevedo, Doutorado Equiparado a Investigador Auxiliar, ICBAS

Co-supervisor
Dulce Quelhas, Farmacêutica- Assessor de Genética Humana, Centro Hospitalar Universitário de Santo António


U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Sworn Statement

I, Andreia Filipa Pereira Amaro, enrolled in the Master Degree of Forensic Genetics at the Faculty of Sciences of the University of Porto hereby declare, in accordance with the provisions of paragraph a) of Article 14 of the Code of Ethical Conduct of the University of Porto, that the content of this dissertation reflects perspectives, research work and my own interpretations at the time of its submission.

By submitting this dissertation, I also declare that it contains the results of my own research work and contributions that have not been previously submitted to this or any other institution.

I further declare that all references to other authors fully comply with the rules of attribution and are referenced in the text by citation and identified in the bibliographic references section. This dissertation does not include any content whose reproduction is protected by copyright laws.

I am aware that the practice of plagiarism and self-plagiarism constitute a form of academic offense.

Andreia Amaro

28/06/2023

Agradecimentos

Todos os fins trazem novos inícios. Na vida todos sofremos uma espécie de metamorfose continua e encontramos nos em constante evolução. Todas as etapas da vida são recheadas de bons momentos e momentos menos bons. E é com a certeza que em cada impossível há uma possibilidade que conclui mais um capítulo da minha jornada.

Este foi um capítulo, com percalços e com alegrias. Foi um capítulo que me enriqueceu mais enquanto pessoa e enquanto futura profissional. Aprendi, cresci, mas não o fiz sozinha, como tal quero aqui deixar os meus agradecimentos.

Agradeço à professora Luísa, por toda a sua dedicação, por cada ensinamento que levarei para toda a vida, pela oportunidade e pela paciência. Pelas suas palavras de encorajamento e incentivo que nos dias mais difíceis sempre me ajudaram. Tal como a professora disse no início deste projeto vai correr tudo bem.

Quero agradecer também à Drª Dulce, pela oportunidade e tempo dedicado.

Agradeço à minha Mãe, por tudo o que sempre fez e faz por mim. Por ser sempre a primeira acreditar que eu sou capaz, por me aturar em cada crise existencial, por ter sempre uma palavra amiga, um colo, e por nunca deixar de acreditar. É uma honra ser tua filha, e ter uma mulher como tu na vida. Nem sempre foi fácil, mas tudo que é difícil compensa no fim. Obrigada pelos incontáveis sacrifícios que fizeste.

Quero também agradecer ao meu namorado, pela paciência, pelo apoio, por também ele acreditar, pelo carinho, pelo companheirismo. Por me fazer sorrir quando as coisas estavam menos bem. Por apesar de não ser da área de ciências, sempre me querer ouvir e sempre ajudar. Por entender que nem sempre pude dar toda a minha atenção, por me ter dado força e motivação para ultrapassar todos os obstáculos.

Apesar de não saberem ler, quero deixar uma menção de agradecimento às minhas companhias diárias, o meu cão e gato. Fazem parte de todo este percurso comigo, sempre a dar a sua presença e um aconchego.

Quero agradecer, aos meus animais, à minha mãe e ao meu namorado por me ensinarem o que o amor significa, por me fazerem sentir amada e apoiada em cada passo da minha vida. Obrigada!

E por último, mas não menos importante agradeço a mim mesma, por me ter permitido rodear de pessoas que me ajudaram a superar os obstáculos, que me ensinaram que a vida pode ser difícil, mas no fim tudo fica bem.

A todos mencionados, o meu muito obrigada por me terem ensinado e ajudado a crescer de formas diferentes. Por me ajudarem a sobrepor obstáculos, a perseverar. Vou continuar a crescer e aprender, mas levarei para sempre comigo o que cada pessoa me ensinou.

O meu mais sincero obrigada!

Resumo

Loci ligados ao X apresentam um padrão único de hereditariedade e por isso é são recursos valiosos em genética populacional. A Adrenoleucodistrofia ligada ao cromossoma X (X-ALD) é uma doença genética rara embora seja a forma mais comum de todas as leucodistrofias. É uma doença progressiva que afeta o córtex adrenal e os sistemas central e periféricos do sistema nervoso. Embora esteja ligada ao cromossoma X e os pacientes masculinos apresentem as manifestações clínicas mais severas da doença, 80% das pacientes femininas apresentam uma progressão que se assemelha a adrenomieloneuropatia (AMN), uma forma de X-ALD.

Pacientes com ALD apresentam variantes deletérios no gene *ABCD1*. Este gene codifica uma *ATP-binding cassette sub-family D member 1* (ABCD1), é uma proteína transmembranar que transporta ácidos gordos de cadeia muito longa (VLCFA), como ésteres-CoA, para o peroxissoma onde ocorre β-oxidação. Variantes deletérios neste gene perturbam o normal metabolismo e levam à acumulação de VLCFA no plasma, na medula espinhal, na matéria branca do cérebro e no córtex adrenal. O conhecimento atual é ainda insuficiente para prever com precisão a severidade da doença, uma vez que dentro da mesma família podem ocorrer diferentes manifestações da doença. O largo espectro fenotípico pode ser observado, até mesmo em gémeos homozigóticos portadores do mesmo variante patogénico. Isto indica que modificadores genéticos (intragénicos e/ou intergénicos) podem estar a modelar o fenótipo de forma semelhante ao que já foi observado em outros genes humanos.

Para avaliar a possibilidade de que variantes polimórficas intragénicas podem agir como modificadores genéticos da doença, neste trabalho foram examinados polimorfismos do gene *ABCD1*. As análises foram realizadas ao nível populacional para avaliar diferenças nas frequências alélicas. Estudos de conservação foram também utilizados como medida para prever o potencial impacto de cada polimorfismo. Os nossos resultados indicam que entre os polimorfismos mais comuns do gene *ABCD1*, Val583Ala e Val604Ile podem ter um efeito modelador e que necessitam de investigação mais aprofundada, incluindo análises experimentais para determinar a possibilidade de existência de diferenças na atividade de cada alelo e a sua influência na proteína quando combinadas com outros alelos no mesmo fundo genético.

Palavras-chave: X-ALD, Leucodistrofia, AMN, Bioinformática, ABCD1, ALDP

Abstract

X-linked loci have a unique pattern of inheritance and are therefore a valuable resource in population genetics. X-linked adrenoleukodystrophy (ALD) is a rare genetic disease and the most common leukodystrophy. It is a progressive disease that affects the adrenal cortex and the central and peripheral nervous system. Although is X-linked and male patients show the most severe clinical manifestations of the disease, 80% of female patients have a progression that resembles adrenomyeloneuropathy (AMN) that is a form of X-ALD.

Patients affected by ALD have deleterious variants in the *ABCD1* gene. The *ABCD1* gene codes an ATP-binding cassette sub-family D member 1 (ABCD1), a transmembrane protein that carries very long chain fatty acids (VLCFA), such are CoA-esters, to the peroxisome where β-oxidation occurs. Deleterious variants in this gene disrupt normal metabolism and lead to the accumulation of very long-chain fatty acids (VLCFA) in the plasma, spinal cord, the white matter of the brain, and adrenal cortex. Current knowledge is still insufficient to accurately predict the severity of the disease, because even within the same family different manifestations of the disease can occur. This wide range of clinical phenotyopes has been observed, even in monozygotic twins carrying the same deleterious variant. This indicates that genetic modifiers (intragenic and/or intergenic) exist and can modulate the phenotype, similarly to what was observed in other human genes.

To evaluate the possibility that intragenic polymorphic variants can act as genetic modifiers of the disease the polymorphic variants of the *ABCD1* gene were examined in this work. The analyses were conducted at the population level to assess differences in allelic frequencies. Conservation scores were also utilized as a measure to predict the potential impact of each polymorphism. Our findings indicate that among the most common polymorphism of the *ABCD1* gene, Val583Ala and Val604Ile, may have a modulating effect, therefore, future investigation should include experimental analyses to determine if there are differences in the activity of each allele and the influence when co-occurring with other alleles in the same genetic background.

Keywords: X-ALD, Leukodystrophy, AMN, Bioinformatics, ABCD1, ALDP

Table of contents

List of tables

List of Figures

List of Abbreviations

ABCD1	ATP-BINDING CASSETTE SUB-FAMILY D MEMBER 1
ALD-X	ADRENOLEUKODYSTROPHY X-LINKED
ALDP	ADRENOLEUKODYSTROPHY PROTEIN
AMN	ADRENOMYELONEUROPATHY
ARDA	ANTI-RESTRICTION PROTEIN
ASPA	ASPARTOCYCLASE
B/L.B	BENIGN/LIKELY BENIGN
CADD	COMBINED ANNOTATION-DEPENDENT DEPLETION
FATHMM	FUNCTIONAL ANALYSIS THROUGH HIDDEN MARKOV MODELS
GALC	GALACTOSYLCERAMIDASE
GWAS	GENOME-WIDE ASSOCIATION STUDIES
MLD	METACHROMATIC LEUKODYSTROPHY
NBD	NUCLEOTIDE-BINDING DOMAIN
PDB	PROTEIN DATA BANK
PLP1	PROTEOLIPID PROTEIN 1
POLYPHEN	POLYMORPHISM PHENOTYPING
SNPS&GO	SINGLE NUCLEOTIDE POLYMORPHIMS AND GENE ONTOLOGY
TMD	TRANSMEMBRANE DOMAIN
VLFCA	NUCLEOTIDE-BINDING DOMAIN
VUS	VARIANT OF UNCERTAIN SIGNIFICANCE

1. Introduction

1.1. Characteristics of the X-chromosome

Studies on the X-chromosome can help to improve our understanding of sex differences in health and disease. At the population level, research on the X-chromosome can provide valuable insights into the prevalence and distribution of X-linked phenotypes, by the identification of patterns and how these phenotypes vary across different populations ((Schaffner 2004); (Gomes, Pinto et al. 2020)). Because males are invariably hemizygous, linkage disequilibrium analyses direct the access to demographic events (Slatkin 2008) and the identification of casual variants in Genome-Wide Association Studies (GWAS) ((König, Loley et al. 2014); (Uffelmann, Huang et al. 2021); (Keur, Ricaño-Ponce et al. 2022); (Gorlov and Amos 2023))

According to OMIM there are more than 500 X-linked diseases. Most of the X-linked diseases are inherited in a recessive mode, which indicates that male individuals will more frequently manifest these diseases.(Sun, Fan and Wang 2022). Females having two X-chromosomes are not expected to manifest recessive diseases in cases of heterozygosity, which means that diseases such as Duchenne muscular dystrophy (Duan, Goemans et al. 2021), haemophilia (Berntorp, Fischer et al. 2021), among others, are rare in females. However, in some instances, females can also reveal the clinical symptoms of X-linked recessive diseases depending on the pattern of X-chromosome inactivation in their cells (Migeon 2020).

1.2. Leukodystrophies

Leukodystrophies were defined as progressive genetic disorders that affect the myelin of the central nervous system in the 1980s (Kevelam, Steenweg et al. 2016). During this time, the first genes associated with leukodystrophies were identified (Kevelam, Steenweg et al. 2016): *PLP1* for Pelizaeus-Merzbacher disease (Hudson, Puckett et al. 1989); *ASPA* for Canavan disease (Kaul, Ping Gao et al. 1993); *ARSA* for MLD (Polten, Fluharty et al. 1991) and *GALC* for Krabbe disease (Sakai, Inui et al. 1994)

X-linked Adrenoleukodystrophy (X-ALD) is a genetic disease caused by deleterious variants in the *ABCD1* gene, located in Xq28. X-ALD is a rare genetic condition with an

incidence of 1 in 14700 births (Honey, Jaspers et al. 2021). Despite its rarity, it is still the most common of all the leukodystrophies (Volmrich, Cuénant et al. 2022). In 1993 the ALD gene was identified by Hugo Moser ((Kemp and Wanders 2010); (Mosser, Douar et al. 1993)), and since 1999 the allelic diversity associated with this gene have been reported in the *ABCD1* variant database (https://adrenoleukodystrophy.info/) (Mallack, Gao et al. 2022).

The *ABCD1* gene encodes for the ATP-binding cassette sub-family D member 1 protein, which is responsible for transporting very long chain fatty acids (VLCFA) from the cytosol into the peroxisome. Inside the peroxisome, VLFCA are degraded by β-oxidation. A disease-causing variants in this gene can result in the accumulation of VLCFA, which interferes with the normal metabolism. This accumulation can be found in the plasma, in the spinal cord, adrenal cortex, white matter and in cultured skin fibroblasts. It is common in clinical practice to analyze the levels of VLFCA in patients with X-ALD, as these levels are typically elevated and can aid in diagnosis (Fujitani, Saito et al. 2022).

ABCD1 is part of a family of 4 transporters, ABCD 1, 2, and 3 are involved in transporting fatty acyl-CoAs from the cytosol into the peroxisome. Although they transport the same molecules, they differ in substrate specificity. ABCD1 has an affinity for transporting saturated and monosaturated VLCFA-CoAs, such are C22:0-CoA, C24:0-CoA, C26:0-CoA, and C26:1-CoA. ABC transporters are expressed as single polypeptides with a transmembrane domain (TMD) and a nucleotide-binding domain (NBD). The ABCD1 protein is structured as a two-fold symmetric homodimer (Chen, Xu et al. 2022) where each subunit contains a TMD consisting of six transmembrane helices (TMs) that are tightly packed against each other in the peroxisomal leaflet. These helices split in the cytosolic membrane leaflet and extend into the cytosol to form two diverged wings (Chen, Xu et al. 2022). The NBD has a canonical NBD fold as well as a pair of helices that form a crossover at the C-terminus. Sequence comparisons have revealed that these C-terminus helices are highly conserved across species.

1.3. Mutational spectrum of *ABCD1*

Although there are three other transporters, only the *ABCD1* gene is known to cause X-ALD (Le, Thompson et al. 2022).

The *ABCD1* gene is responsible for encoding a protein consisting of 745 amino acids, which are distributed across 10 exons. Exon 1 contains the majority of deleterious variants and is a hotspot for missense replacements, accounting for 40% of them (Mallack, Gao et al. 2022). Approximately 30% of missense replacements occur in the region that codes for the ATP-binding domain, which is located in exons 6 to 9. According to Jia, Zhang et al. (2022), there are at least 43 deleterious variants that can affect the conformational coupling between the NBD and TMD. Additionally, some alterations in the C-terminus can difficult the recognition of PEX19 that targets ABCD1 to the membrane of peroxisome (Jia, Zhang et al. 2022).

Currently, more than 1300 variants are annotated in the ALD database (https://adrenoleukodystrophy.info/), of which 890 (65%) are tagged as “pathogenic” and “likely pathogenic” (Figure 1A). The second most common category is VUS, representing one quarter of the total variants. Benign (3%) and synonymous (7%) and unlikely to have any deleterious effect. Similarly to other disease-associated proteins the mutational spectrum includes different types of replacements like missense, indels and frameshift (small insertions and deletions-indels) variants ((Zemanova, Chrastina et al. 2021);(Kim, Kim et al. 2021); (Nogales-Gadea, Brull et al. 2015)) (Figure 1B). The most common type of deleterious substitutions is missense (45%) followed by frameshifts which reach 37% of the total according to the ALD database.

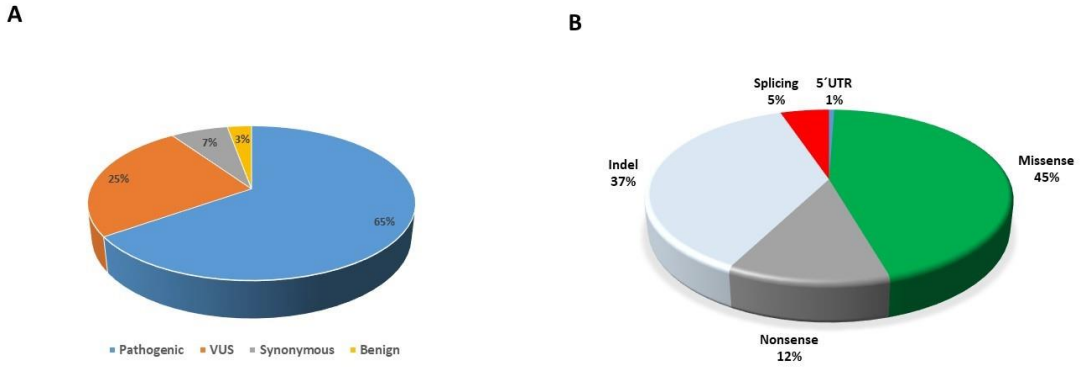

Figure 1 Distribution of variants as annotated in ALD database (A) and categories of pathogenic variants (B)

Phenotypic spectrum

X-ALD is a challenging disease from clinical perspective as it present four main types of manifestation: (1) childhood cerebral ALD (ccALD), (2) adult cerebral ALD (acALD), (3) adrenomyeloneuropathy (Karczewski, Francioli et al.), and (4) Addison's disease. Despite these being the most common manifestations of the disease, it is still not

possible to determine which one a person with a specific disease-associated variant in the gene will manifest.

Female patients typically do not manifest the symptoms early on. In fact, 80% only start to manifest much later in life and are commonly diagnosed with AMN (Huffnagel, Dijkgraaf et al. (2019)). Male patients are the ones who experience the most severe forms of ALD, such as cerebral ALD. Symptoms usually begin to manifest between the ages of 3 and 12 years old and death normally occurs within 2 years of diagnose.

Currently, there is no cure for ALD. The only treatment available is hematopoietic stem cells transplant, which can halt the progression of neurological symptoms. However, this treatment is only effective, in children and must be performed before the symptoms begin to manifest (Kemp, Huffnagel et al. 2016).

Establishing a correlation between phenotype and genotype in X-ALD has proven to be difficult because different forms of the disease can be identified even in the same family. This variability is one of the most intriguing characteristics of the disease.

For example, in 1996, a set of monozygotic twins sharing the same deleterious variant manifested different clinical phenotypes of the disease (Korenke, Fuchs et al. 1996). In another study, a mother and her four children, were all found to carry the same variant, but despite sharing the same genotype, they all showed different phenotypes. One of the male brothers did not manifest any symptoms, while the other two only had one symptom in common (Mohn, Polidori et al. 2021).

In this study, we evaluate the possibility that intragenic polymorphic variants can act as genetic modifiers of the *ABCD1* gene and can therefore explain, at least partially, the lack of genotype-phenotype correlation, given the current gap in knowledge related to this topic.

2. Methods

2.1. Selection of polymorphism

The GnomAD database (Karczewski, Francioli et al. 2020) was used to retrieve information on the position, variant ID, amino acids changes, protein and transcript consequences, allele frequency, and homo- and hemizygote count for *ABCD1* variants. Only missense replacements were selected. The search resulted in 305 results that were further filtered as follows: all the variants that displayed a count of hemizygotes below 10 were eliminated to maximize the probability of these being benign variants. This resulted eight missense variants and selected for further analyses. For each variant, the allelic frequencies were obtained from GnomAD for eight populations: European (Finnish), European (non-Finnish), East Asian, South Asian, Latin/Admixed American, African/African American, Ashkenazi Jewish and other.

2.2. Computational analyses

The eight variants were analysed using different online software such as, FATHMM, SNP&GO (which shows protein structure/function), CADD, and Polyphen. All of analyses were made using the default parameters.

The ConSurf server, a tool that estimates the evolutionary conservation of amino/nucleic acid positions in a protein based on phylogenetic relations between homologous sequences, was used to access the conservation of the eight polymorphisms. This was a useful tool for accessing the conservation of the protein in the positions that were included in the study (Landau, Mayrose et al. 2005).

The FATHMM software predicts the functional effects of missense mutations in a protein by combining sequence conservation with hidden Markov models (HMMs). This provides, an overall measure of the protein/domain's tolerance to mutations. Hidden Markov models provide a toolkit for building complex models, they make a probabilistic model for linear sequence "labeling" problems. This software was used to retrieve information on how missense mutations impact the function of the protein. Predictions are given as p-value, with values above 0.5 predicted to be deleterious and those below 0.5 predicted to be neutral or benign (Shihab, Gough et al. (2013)).

The Combined Annotation Dependent Depletion (CADD), software it provides two types of scores: RAW-score and PHRED (scaled C-score). For the analyses, the PHRED score was collected. This score ranks a variant in relation to all possible substitutions in the human genome. A score greater than 10 indicates that the variant is predicted to be in the 10% most deleterious substitutions in the genome, while a score greater than 20 indicates that it is predicted to be in the 1% most deleterious (Rentzsch, Schubach et al. (2021)).

Polyphen-2 is a tool that predicts the impact of amino acid substitutions on the structure and function. It provides two results: HumDiv and HumVar. HumDiv is the default classifier model, used for evaluating rare alleles, dense mapping regions, and analysing natural selection. HumVar is more appropriated for Mendelian diseases, that require distinguishing mutations with drastic effects. The Polyphen score values range from 0.0 (tolerable) to 1.0 (deleterious) (Adzhubei, Jordan and Sunyaev 2013)

SNPs&GO is a database that provides information derived from protein sequence and function, as well as 3D structure. It combines information from various sources, including gene ontology, to predict whether a given variation is disease-related or neutral.

2.3. Structural analyses

PDB code for ABCD1 protein was obtained from Protein Data Bank (Le, Thompson et al. 2022). All structures were visualized in PyMOL (academic version) (PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC). Missense3D was used to predict the structural changes introduced by missense variants at the 3D structure (Ittisoponpisan, Islam et al. 2019). The data was input at the “Position on Protein Sequence” option using the following parameters: UnitProt ID P33897, PDB code 7RRA and PDB chain ID A. This 3D structure lacks information for amino acid 13 and 452.

3. Results and Discussion

3.1. ABCD1 polymorphic variants

To achieve the aim of this study, we started the analyses by searching GnomAD database to find putative functional polymorphisms that might be genetic modifiers involved in the observed reduced penetrance found in X-ALD. For this purpose, missense polymorphisms that were detected in at least 10 hemizygous individuals were selected. A reasonable number of homozygous individuals would indicate that these polymorphisms might be *per se* neutral in the general population. The set of polymorphisms are shown in Table 1.

Table 1 Polymorphisms of the ABCD1 protein obtained from gnomAD database

Variant ID	Position	Protein consequence	Transcript consequence	Allele frequency	Homozygotes (total) / Hemizygotes (total)
rs183021839	152990759	p.Asn13Thr	c.38A>C	2.68E-3	2/ 79
rs139415350	152991322	p.Val201Met	c.601G>A	1.67E-4	0/ 13
rs781932570	152991412	p.Arg231Trp	c.691C>T	3.03E-4	0/ 16
rs201455322	152991428	p.Arg236His	c.707G>A	1.29E-3	2/ 85
rs150151955	152991478	p.Leu253Val	c.757C>G	8.01E-4	0/ 39
rs368061976	153001928	p.Arg452Trp	c.1354C>T	4.65E-5	0/ 27
rs79383557	153006141	p.Val583Ala	c.1748T>C	1.49E-4	0/ 10
rs151201945	153008470	p.Val604Ile	c.1810G>A	6.29E-4	1/ 24

For all the eight variants, we evaluated the biochemical properties of the wild-type allele and the derived allele, considering that a conservative replacement implies that a given amino acid is replaced by another that, has similarly charged. The non-conservative replacement occurs when the change implicates an amino acid with different charge properties from the original. The results demonstrate that out of the eight replacements,

six were conservative, and only two represent non-conservative replacements: p.Arg231Trp and p.Arg452Trp.

Next, we located these variants in order to evaluate a possible clustering of polymorphic sites within the gene. Figure 1 shows the exons where the eight variants are located. Previous studies showed that exon 1 is a hotspot for missense deleterious variants (Mallack, Gao et al. 2022) and in terms of common polymorphic variation the scenario is similar. In fact, Val583Ala and Val604Ile, which are located in exons, 7 and 8 respectively, all the others are in exon 1.

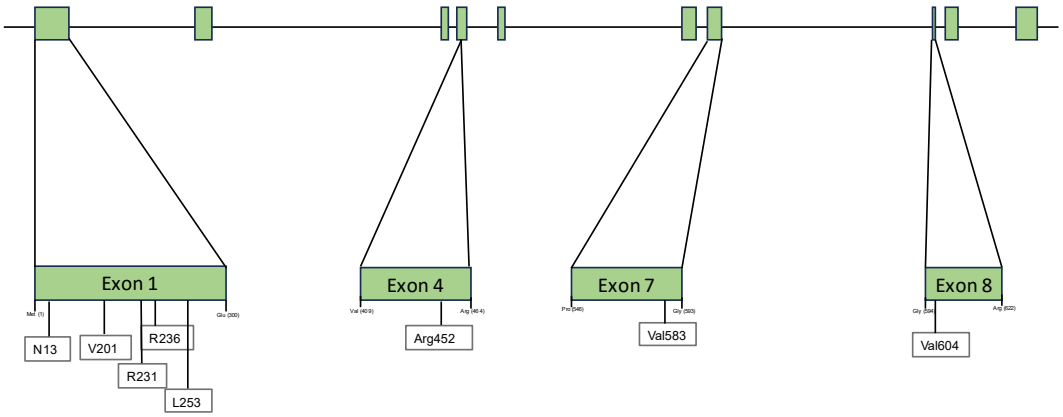

Figure 2 structure of the ABCD1 gene, showing the location of exonic variants.

3.2. In silico predictions of putative functional impact

Table 2 presents the scores obtained from the in-silico analyses using distinct tools routinely applied for evaluating functional significance of protein replacements. CADD scores above 10 and 20 represent missense replacements that are in the 10% and 1%, respectively, of the most deleterious substitutions in the human genome. A good indication of putative functional impact is a CADD score above 20, whereas a score below 10 would indicate a neutral variant.

*Table 2 Values from the used databases. For CADD, a score >10, indicates that a variant is in the 10% most deleterious substitutions in the genome and score > 20 indicates that is in the 1% most deleterious. Polyphen scores range from tolerable (0.0) to deleterious (1.0). FATHMM scores are given in p-values, p-value> 0.5 is deleterious, p-value<0.5 is tolerable. SNPs&GO predicts if a variation in the gene is disease-related or neutral. *Clinvar classification retrieved from GnomAD database*

		Polyphen				
Variant ID	**CADD score**	HDiv	HVar	**FATHMM**	**SNPs&GO**	**Clinical classification***
rs183021839	6.635	0	0	0.09535	Neutral	B/L.B
rs139415350	18.56	0.234	0.149	0.14620	Disease	B
rs781932570	23.1	1.00	0.953	0.21557	Disease	B
rs201455322	24.6	0.992		0.81710	Disease	C.I. P
rs150151955	16.46	0.368	0.140	0.48072	Disease	C.I. P
rs368061976	18.22	No result		0.07197	Disease	B
rs79383557	24.4	1	0.998	0.98743	Disease	No result
rs151201945	23.3	0.939	0.547	0.98292	Disease	B

Variant rs183021839 (Asn13Thr), is the only scored lower than 10 by CADD. According to the literature, Asn13Thr was the first missense replacement reported in the *ABCD1* gene (Dvoráková, Štorkánová et al. 2001). Other authors have concluded that it retains wild type ALDP function. In fact, transfection of X-ALD fibroblasts led to an increase in β-oxidation similar to that observed with normal ABCD1 cDNA transfection and was not different from the level shown in healthy control cells (Dvoráková, Štorkánová et al. 2001). In agreement, a score of 0 in Polyphen indicates that this amino acid replacement is tolerable. Similarly, FATHMM scored this variation below 0.5, indicating that it is a neutral or benign substitution. SNPs&GO and Clinvar also classify it as neutral and benign/likely benign, respectively. Altogether, these data are in agreement with the Clinvar's classification of this variant-benign (Clinvar code: 368044).

The variant p.Leu253Val (rs150151955), revealed a CADD scored of 16.46 which indicates it belongs to the 10% of the most deleterious substitutions in the genome. In agreement with this, SNPs&GO classified it as deleterious. However, Polyphen and FATHMM assume it as tolerable. This type of result goes in line with the classification of Clinvar since is classified as C.I.P. (conflicting interpretations of pathogenicity, Clinvar

code: 377034). Variant p.Arg452Trp, has similar results. CADD scored 18.22, SNPs&GO also classified it as deleterious. Another similar result was obtained for rs1394415350 (p.Val201Met), scoring 18.56 in CADD and predicted as potentially deleterious by SNPs&GO, but received three scores of being tolerable and ClinVar (ClinVar code: 585357) also classified this variation as benign. For these cases, the putative impact is difficult to establish using the in-silico predictions only and these data were analysed further in terms of conservation degree in the next sections.

The p.Arg231Trp received a CADD score above 20 (23.1), meaning that belongs to the 1% of the most deleterious substitutions on the genome. In agreement with this, SNPs&GO and Polyphen classified as deleterious. However, FATHMM classified it as tolerable. The same classification was obtained for variants, p.Arg236His, p.Val583Ala and p.Val604Ile, all receiving scores above 20 in CADD (24.6, 24.4 and 23.3 respectively). The same results were obtained from the other programs. Altogether these findings indicate a putative functional impact for these replacements and justified the analyses presented in the following sections.

3.3. Allelic frequency among human populations

The allele frequency of the variants in the study for eight populations was obtained from GnomAD and is displayed in table 3.

Table 3 Distribution of allelic frequencies of eight ABCD1 across eight populations as retrieved from GnomAD

	p.Asn13Thr	p.Val201Met	p.Arg231Trp	p.Arg236His	p.Leu253Val	p.Arg452Trp	p.Val583Ala	p.Val604Ile
European (Finnish)	0.01184	0.000	0.002997	0.000	0.001571	0.000	0.001383	0.000
European (non-Finnish)	0.01184	0.00003262	0.00005478	0.00001541	0.001335	0.000	0.00003938	0.0001483
East Asian	0.000	0.000	0.000	0.0001368	0.000	0.000	0.000	0.000
South Asian	0.000	0.000	0.000	0.000	0.000115	0.0001050	0.000	0.00005908
Latin/Admixed American	0.0008994	0.000	0.000	0.001467	0.00007323	0.002998	0.000	0.0002279
African/ African American	0.0002841	0.001517	0.000	0.00005322	0.0002166	0.00005261	0.000	0.0002327
Ashkenazi Jewish	0.000	0.000	0.000	0.02523	0.000	0.000	0.000	0.01283
Other	0.004366	0.0003455	0.001137	0.002283	0.001571	0.001505	0.0002119	0.0004085

Although no single variant is polymorphic in all populations, three variants – p.Arg236His, p.Leu253Val, and p.Val604Ile – can be found in four of the populations (European, Latin/Admixed American, African/African American, and Others). Variants with polymorphic (MAF≥1%) are the p.Asn13Thr in Europeans, and both p.Arg236His and p.Val604Ile in Ashkenazi Jewish. All the other variants are present at low or very low frequencies in humans populations. It's important to note that while these variants are rare, they were initially filtered because they were found in at least 10 hemizygous individuals. This suggests that if they have a functional effect, it could indicate a potential modifier role in *cis* or *trans* with other variants (deleterious and/or polymorphic variants). This phenomenon, known as epistasis ((Cheverud and Routman 1995); (de Visser, Cooper and Elena 2011); (Lopes-Marques, Pacheco et al. 2021); (Chan, Longley and Copeland 2006)), is common among protein variants and can result in distinct outcomes for the same variant (Gonzalez and Ostermeier 2019).

3.4. Conservation scores

Proteins carry out a wide range of functions, all of which are mediated by amino acid residues that have a functional role (Studer, Dessailly and Orengo 2013). Proteins are dynamic entities and undergo alternative periods of conservative evolution with those periods of rapid changes (Studer, Dessailly and Orengo 2013). Therefore, each amino acid has a definite and important function within proteins and is involved in interactions with other proteins that render the protein space a dynamic and articulate space (DePristo, Weinreich and Hartl 2005).

With the aim to analyse the conservation degree of the polymorphic variants, we used ConSurf to score each of the amino acid positions. In table 4 the results from ConSurf are shown.

Table 4 Results of the ConSurf analysis

Polymorphism	ConSurf	Amino acids in orthologues positions
Asn13	3 (variable)	N, T, S, G, A, V, K
Val201	6 (average)	V, M, D, E, S, T, A, H, Q, N, P, E
Arg231	4 (variable)	R, Q, S, K, T, E, G, A, Y, V, N, W, H, F, C, L, D
Arg236	5 (average)	R, K, E, Q, M, L, V, G, S, T, F, A
Leu253	2 (variable)	L, V, I, A, F, G, T, S, M
Arg452	1 (variable)	R, S, Q, K, N, T, V, P, E
Val583	8 (conserved)	V, A, L, T, S, C
Val604	8 (conserved)	V, I, R, A, T, S, E, Q

Position Asn13 was classified as variable, indicating that a change at this position would be less harmful. Results from previous analyses (Table 2) show that a change at this position is not deleterious, which is consistent with the conservation score obtained in Consurf. Positions, Val201 and Arg236 both received a classification of average in conservation, whereas Leu253 and Arg452 received the classification of variable in terms of conservation. Positions Val583 and Val604, were classified was being conserved, and previous results (Table 2) showed that these two replacements might have a functional impact in the protein. Overall, the ConSurf results are in agreement with previous analyses.

3.5. In silico prediction of structural changes

To predict the changes introduced by all missense variants (except Asn13Thr and Arg452Trp for which no information is available at the PDB 7RRA), we used Missense3D. The results (Table 5) clearly demonstrated that none of the six changes analysed result in structural damage to the protein. However some of them may result in subtle changes, such was observed in the relative surface area (RSA), important for the special rearrangement and protein folding ((Marsh and Teichmann 2011); (Tien, Meyer et al. 2013)) . This could indicate that these variants may have an impact when they co-occur with other polymorphic or deleterious variants.

Table 5 Prediction of structural damage introduced by replacements of wild-type aminoacids

Variants	Buried charge replaced	Buried H-bond breakage	Alteration
Asn13Thr	-	-	-
Val201Met	This substitution does not replace a buried charged residue with an uncharged residue. The wild-type residue VAL is buried uncharged with RSA 2.1% and the mutant residue MET is buried uncharged with RSA 1.0%.	The buried wild-type residue is not involved in any side-chain / side-chain H-bond(s) and/or side-chain / main-chain bond(s) H-bonds.	No structural damage detected
Arg231Trp	This substitution does not replace a buried charged residue with an uncharged residue. The wild-type residue ARG is exposed charged with RSA 37.9% and the mutant residue TRP is exposed uncharged with RSA 79.7%.	This substitution does not result in a complete disruption of all side-chain / side-chain H-bond(s) and/or side-chain / main-chain bond(s) bonds, and the wild-type residue is not buried (RSA 37.9%).	No structural damage detected
Arg236His	This substitution does not replace a buried charged residue with an uncharged residue. The wild-type residue ARG is exposed charged with RSA 69.7% and the mutant residue HIS is exposed charged with RSA 71.1%.	This substitution does not result in a complete disruption of all side-chain / side-chain H-bond(s) and/or side-chain / main-chain bond(s) bonds, and the wild-type residue is not buried (RSA 69.7%).	No structural damage detected
Leu253Val	This substitution does not replace a buried charged residue with an uncharged residue. The wild-type residue LEU is exposed uncharged with RSA 69.5% and the mutant residue VAL is exposed uncharged with RSA 54.9%.	The wild-type residue is not involved in any side-chain / side-chain H-bond(s) and/or side-chain / main-chain bond(s) H-bonds.	No structural damage detected
Arg452Trp	-	-	-
Val583Ala	This substitution does not replace a buried charged residue with an uncharged residue. The wild-type residue VAL is buried uncharged with RSA 0.7% and the mutant residue ALA is buried uncharged with RSA 0.0%.	The buried wild-type residue is not involved in any side-chain / side-chain H-bond(s) and/or side-chain / main-chain bond(s) H-bonds.	No structural damage detected
Val604Ile	This substitution does not replace a buried charged residue with an uncharged residue. The wild-type residue VAL is exposed uncharged with RSA 37.3% and the mutant residue ILE is exposed uncharged with RSA 30.7%.	The wild-type residue is not involved in any side-chain / side-chain H-bond(s) and/or side-chain / main-chain bond(s) H-bonds.	No structural damage detected

3.6. Structural visualization and analysis

To locate the six missense variants at the available 3D structure of the protein, the ABCD1 structure corresponding to the PDB code 7RR9 was used (Le, Thompson et al. 2022). Five (Val201, Arg231, Arg236, Leu253 and Val583) positions analysed are in the α-helix (Figure 3), a structure in the protein where DNA binding motifs are typically located. The α-helix is commonly the element that crosses biological membranes and is, more robust than β-sheets in relation to the impact of amino acid replacements. Val604 is in a different structural element - a loop. These structures are important in the folding of the protein and allow interactions between other secondary structures.

Since our previous analyses indicate a putative deleterious impact for the replacements at positions p. Val583 and Val604 (Figure 4) we analysed in more detail the structural environment. These two residues are the core of the structure, and it is possible that interactions with neighbouring amino acids render more likely that the replacement would result in the disturbance of this network of interactions. It is also possible that, although these alterations might not be responsible for great alterations in the protein, nor result in any clinical symptom, they might create an unstable 3D environment that, in combination with other variants, can have a more evident impact.

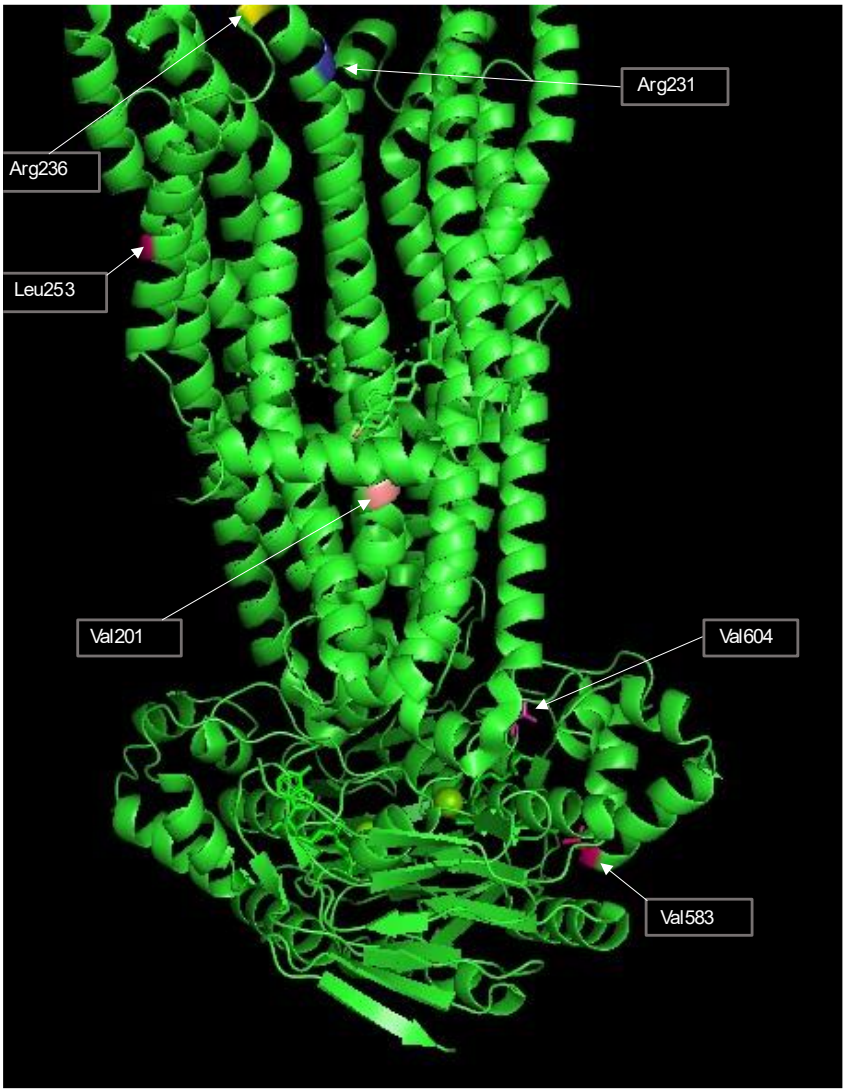

Figure 3- 3D structure of the ABCD1 protein showing the location of the six missense variants

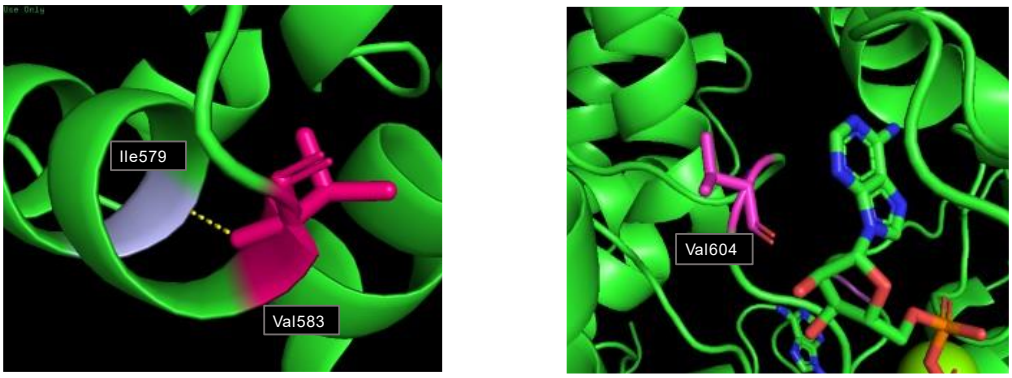

Figure 4- Detailed structural information for positions Val583 and the polar contact with Ile579 and Val604 at the ABCD1 3D structure.

4. Conclusion

This study is integrated in the Forensics Genestics master course because the transmission characteristics of the X-chromosome make it a valuable tool for population studies, kinship analysis, forensic investigations and clinical genetics. The low rate of recombination and mutation results in faster genetic drift, leading to stronger linkage disequilibrium and population structure on the X-chromosome (Garcia, Bessa et al. 2022). Therefore, the study of X-linked diseases from a population perspective might result in the discovery of novel variants or instances of interactions between distinct variants. This is possible because while females have two copies of the X-chromosome that recombine during meiosis and are passed on to both male and female offspring, males transmit the X-chromosome almost entirely as a block to their female offspring. The aim of this study was to investigate the putative functional role of eight missense variants reported for ABCD1, an X-linked gene. Using a batery of *in silico* predictors of pathogeneicity, allelic frequency data and structural evaluations, we concluded that none of these variants have *per se* an important impact. However, two of them (Val583Ala and Val604Ile) are well conserved and predicted as potentially relevant. Although they are rare in the general population, they may be important when analyzing a group of X-ALD patients. Currently, there is no data available on the frequency of these polymorphisms in X-ALD patients.

Future pespectives

This study suggests that some polymorphic variants in the *ABCD1* gene are conserved and warrant further investigation into their role. We anticipate that the next steps in this project will involve evaluating the allelic frequency of these variants in a cohort of X-ALD affected individuals. Analyzing male samples will be particularly valuable as it will allow us to immediately establish linkage disequilibrium between variants and provide a direct genotype-phenotype correlation. Additionally, the impact of these variants should be evaluated through *in vitro* experimental assays.

5. References

Adzhubei, I., D. M. Jordan and S. R. Sunyaev (2013). "Predicting functional effect of human missense mutations using PolyPhen-2." Curr Protoc Hum Genet **Chapter 7**: Unit7.20.

Berntorp, E., et al. (2021). "Haemophilia." Nature Reviews Disease Primers **7**(1): 45.

Chan, S. S., M. J. Longley and W. C. Copeland (2006). "Modulation of the W748S mutation in DNA polymerase gamma by the E1143G polymorphismin mitochondrial disorders." Hum Mol Genet **15**(23): 3473-3483.

Chen, Z.-P., et al. (2022). "Structural basis of substrate recognition and translocation by human very long-chain fatty acid transporter ABCD1." Nature Communications **13**(1): 3299.

Cheverud, J. M. and E. J. Routman (1995). "Epistasis and its contribution to genetic variance components." Genetics **139**(3): 1455-1461.

de Visser, J. A. G. M., T. F. Cooper and S. F. Elena (2011). "The causes of epistasis." Proceedings of the Royal Society B: Biological Sciences **278**(1725): 3617-3624.

DePristo, M. A., D. M. Weinreich and D. L. Hartl (2005). "Missense meanderings in sequence space: a biophysical view of protein evolution." Nat Rev Genet **6**(9): 678-687.

Duan, D., et al. (2021). "Duchenne muscular dystrophy." Nature Reviews Disease Primers **7**(1): 13.

Dvoráková, L., et al. (2001). "Eight novel ABCD1 gene mutations and three polymorphisms in patients with X-linked adrenoleukodystrophy: The first polymorphism causing an amino acid exchange." Human Mutation **18**(1): 52-60.

Fujitani, N., et al. (2022). "Detection of Characteristic Phosphatidylcholine Containing Very Long Chain Fatty Acids in Cerebrospinal Fluid from Patients with X-Linked Adrenoleukodystrophy." Biol Pharm Bull **45**(11): 1725-1727.

Gomes, I., et al. (2020). "Twenty Years Later: A Comprehensive Review of the X Chromosome Use in Forensic Genetics." Frontiers in Genetics **11**.

Gonzalez, C. E. and M. Ostermeier (2019). "Pervasive Pairwise Intragenic Epistasis among Sequential Mutations in TEM-1 β-Lactamase." J Mol Biol **431**(10): 1981-1992.

Gorlov, I. P. and C. I. Amos (2023). "Why does the X chromosome lag behind autosomes in GWAS findings?" PLOS Genetics **19**(2): e1010472.

Honey, M. I. J., et al. (2021). "Molecular Biomarkers for Adrenoleukodystrophy: An Unmet Need." Cells **10**(12).

Hudson, L. D., et al. (1989). "Mutation of the proteolipid protein gene PLP in a human X chromosome-linked myelin disorder." Proceedings of the National Academy of Sciences **86**(20): 8128-8131.

Huffnagel, I. C., et al. (2019). "Disease progression in women with X-linked adrenoleukodystrophy is slow." Orphanet Journal of Rare Diseases **14**(1): 30.

Ittisoponpisan, S., et al. (2019). "Can Predicted Protein 3D Structures Provide Reliable Insights into whether Missense Variants Are Disease Associated?" Journal of Molecular Biology **431**(11): 2197-2212.

Jia, Y., et al. (2022). "Structural and functional insights of the human peroxisomal ABC transporter ALDP." eLife **11**: e75039.

Karczewski, K. J., et al. (2020). "The mutational constraint spectrum quantified from variation in 141,456 humans." Nature **581**(7809): 434-443.

Kaul, R., et al. (1993). "Cloning of the human aspartoacylase cDNA and a common missense mutation in Canavan disease." Nature Genetics **5**(2): 118-123.

Kemp, S., et al. (2016). "Adrenoleukodystrophy - neuroendocrine pathogenesis and redefinition of natural history." Nat Rev Endocrinol **12**(10): 606-615.

Kemp, S. and R. Wanders (2010). "Biochemical Aspects of X-Linked Adrenoleukodystrophy." Brain Pathology **20**(4): 831-837.

Keur, N., et al. (2022). "A systematic review of analytical methods used in genetic association analysis of the X-chromosome." Briefings in Bioinformatics **23**(5).

Kevelam, S. H., et al. (2016). "Update on leukodystrophies: a historical perspective and adapted definition." Neuropediatrics **47**(06): 349-354.

Kim, H. W., et al. (2021). "Adrenomyeloneuropathy with cerebral involvement due to a novel frameshift variant in ABCD1 gene." Annals of Clinical Neurophysiology **23**: 61-64.

König, I. R., et al. (2014). "How to include chromosome X in your genome-wide association study." Genet Epidemiol **38**(2): 97-103.

Korenke, G. C., et al. (1996). "Cerebral adrenoleukodystrophy (ALD) in only one of monozygotic twins with an identical ALD genotype." Ann Neurol **40**(2): 254-257.

Landau, M., et al. (2005). "ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures." Nucleic Acids Research **33**(suppl_2): W299-W302.

Le, L. T. M., et al. (2022). "Structures of the human peroxisomal fatty acid transporter ABCD1 in a lipid environment." Communications Biology **5**(1): 7.

Lopes-Marques, M., et al. (2021). "Common polymorphic OTC variants can act as genetic modifiers of enzymatic activity." Hum Mutat **42**(8): 978-989.

Mallack, E. J., et al. (2022). "Structure and Function of the ABCD1 Variant Database: 20 Years, 940 Pathogenic Variants, and 3400 Cases of Adrenoleukodystrophy." Cells **11**(2): 283.

Marsh, J. A. and S. A. Teichmann (2011). "Relative solvent accessible surface area predicts protein conformational changes upon binding." Structure **19**(6): 859-867.

Migeon, B. R. (2020). "X-linked diseases: susceptible females." Genetics in Medicine **22**(7): 1156-1174.

Mosser, J., et al. (1993). "Putative X-linked adrenoleukodystrophy gene shares unexpected homology with ABC transporters." Nature **361**(6414): 726-730.

Nogales-Gadea, G., et al. (2015). "McArdle Disease: Update of Reported Mutations and Polymorphisms in the PYGM Gene." Hum Mutat **36**(7): 669-678.

Polten, A., et al. (1991). "Molecular Basis of Different Forms of Metachromatic Leukodystrophy." New England Journal of Medicine **324**(1): 18-22.

Rentzsch, P., et al. (2021). "CADD-Splice-improving genome-wide variant effect prediction using deep learning-derived splice scores." Genome Med **13**(1): 31.

Sakai, N., et al. (1994). "Krabbe Disease: Isolation and Characterization of a Full-Length cDNA for Human Galactocerebrosidase." Biochemical and Biophysical Research Communications **198**(2): 485-491.

Schaffner, S. F. (2004). "The X chromosome in population genetics." Nature Reviews Genetics **5**(1): 43-51.

Shihab, H. A., et al. (2013). "Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models." Hum Mutat **34**(1): 57-65.

Slatkin, M. (2008). "Linkage disequilibrium — understanding the evolutionary past and mapping the medical future." Nature Reviews Genetics **9**(6): 477-485.

Studer, Romain A., Benoit H. Dessailly and Christine A. Orengo (2013). "Residue mutations and their impact on protein structure and function: detecting beneficial and pathogenic changes." Biochemical Journal **449**(3): 581-594.

Sun, Z., J. Fan and Y. Wang (2022). "X-Chromosome Inactivation and Related Diseases." Genet Res (Camb) **2022**: 1391807.

Tien, M. Z., et al. (2013). "Maximum Allowed Solvent Accessibilites of Residues in Proteins." PLOS ONE **8**(11): e80635.

Uffelmann, E., et al. (2021). "Genome-wide association studies." Nature Reviews Methods Primers **1**(1): 59.

Volmrich, A. M., et al. (2022). "ABCD1 Gene Mutations: Mechanisms and Management of Adrenomyeloneuropathy." Appl Clin Genet **15**: 111-123.

Zemanova, M., et al. (2021). "X-linked adrenoleukodystrophy: phenotype-genotype correlation in hemizygous males and heterozygous females with ABCD1 mutations." Neuro Endocrinol Lett **42**(5): 359-367.