Article

Systematic Assessment of Protein C-Termini Mutated in Human Disorders

Zachary T. FitzHugh [1,2] and **Martin R. Schiller** [1,2,3,*]

1   Nevada Institute of Personalized Medicine, University of Nevada Las Vegas, 4505 S. Maryland Pkwy, Las Vegas, NV 89154, USA
2   School of Life Sciences, University of Nevada, 4505 S. Maryland Parkway, Las Vegas, NV 89154, USA
3   Heligenics Inc., 833 Las Vegas Blvd. North, Suite B, Las Vegas, NV 89101, USA
*   Correspondence: martin.schiller@unlv.edu; Tel.: +1-702-895-5546; Fax: +1-702-895-5728

Abstract: All proteins have a carboxyl terminus, and we previously summarized eight mutations in binding and trafficking sequence determinants in the C-terminus that, when disrupted, cause human diseases. These sequence elements for binding and trafficking sites, as well as post-translational modifications (PTMs), are called minimotifs or short linear motifs. We wanted to determine how frequently mutations in minimotifs in the C-terminus cause disease. We searched specifically for PTMs because mutation of a modified amino acid almost always changes the chemistry of the side chain and can be interpreted as loss-of-function. We analyzed data from ClinVar for disease variants, Minimotif Miner and the C-terminome for PTMs, and RefSeq for protein sequences, yielding 20 such potential disease-causing variants. After additional screening, they include six with a previously reported PTM disruption mechanism and nine with new hypotheses for mutated minimotifs in C-termini that may cause disease. These mutations were generally for different genes, with four different PTM types and several different diseases. Our study helps to identify new molecular mechanisms for nine separate variants that cause disease, and this type of analysis could be extended as databases grow and to binding and trafficking motifs. We conclude that mutated motifs in C-termini are an infrequent cause of disease.

Keywords: minimotifs; C-terminus; post-translational modification; acetylation; methylation; phosphorylation; bioinformatics

1. Introduction

For rare diseases of large effect size, most of the mutations are within the coding regions [1]. Proteins are often composed of domains, with the average human protein having 2.06 domains [2]. These domains are collected in several databases, including PROSITE and the conserved domain database (CDD) [3–5]. The domains are often identified by motif signatures. They are also often identified by an algorithmic approach where sequence conservation is not consistent enough to define domains. Furthermore, some domain identities can only be determined after recognition of a common fold through 3D structure analysis [6].

Minimotifs are short functional sequence motifs that encode a function in a protein that is not a domain, such as post-translational modification (PTM) sites, binding sites, and trafficking elements [7,8]. The current datasets of Minimotif Miner (MnM) and Eukaryotic Linear Motif (ELM) have many consensus sequences and instances of these motifs [7,9]. An offshoot of the MnM database is the C-terminome database, which records over 3600 human proteins with a known minimotif in the last ten amino-acid residues [10]. This suggests that most or all human proteins may have functionalized C-termini.

What is not yet known is how often minimotifs are mutated in disease, though there are several reports that indicate that this is a disease mechanism [11–13]. In our first paper on the C-terminome, there were eight C-termini mutated in different diseases (Table 1) [14–22].

Table 1. Eight known C-terminal motifs in disease.

Protein	Minimotif	General Function	Disease	Reference
K_v11.1	RGRX$_{152}$> [1]	Trafficking	Long QT syndrome type 2	Kupershmidt et al. (2002), [15]
Rhodopsin	VAPA>	Trafficking	Vision loss	Deretic et al. (2005), [16]
MNK	LLX$_{12}$> [1]	Trafficking	Menkes disease	Petris et al. (1998), [17]
NKCC2	LLX$_{13}$> [1]	Trafficking	Bartter syndrome	Zaarour et al. (2012), [18]
BSEP	YYKLVX$_7$> [1]	Trafficking	Primary familial intrahepatic cholestasis type 2	Lam et al. (2012), [19]
SANS	TEL>	Binding	Usher syndrome type 1	Reiners et al. (2006), [20]
Claudin-16	TRV>	Binding	Nephrocalcinosis	Müller et al. (2003), [21]
NaV1.5	SIV>	Binding	Cardiac disease	Shy et al. (2013), [22]

[1] The symbol "X$_y$>" indicates that y additional amino acids occur before the C-terminus.

Collectively, these observations led us to question how many other C-termini may be mutated in diseases. In this paper, we used available databases to identify C-termini that are mutated in disease and the subset of those in which the mutation was in a known minimotif. We focused on PTMs because mutation of a key residue that is covalently modified changes the chemistry and can be reliably assigned as a loss-of-function (LOF). We designed a structured approach, taking advantage of current knowledge collected in several databases, to identify several known C-termini disease mechanisms and several new hypotheses where a PTM site is mutated in a disease but is not yet known to cause the disease.

2. Methods

Our bioinformatics analysis examined the presence of mutations in minimotifs from the C-terminome and MnM [7,8,10]. The C-terminome is a comprehensive database of over 3600 minimotifs in human and animal protein C-termini [10]. The MnM database likewise contains over one million motif sequences [7]. For mutation data, we used the NCBI ClinVar, an annotated repository of genetic variants that catalogues their pathogenicity and expressed phenotypes [23,24]. We implemented all data cleaning, analysis, and statistics with custom Python scripts. Most scripts relied on the pandas and numpy Python data science libraries.

Our first task was to procure the mutations from ClinVar and then join other datasets to the resulting table. First, we extracted all missense and nonsense SNPs from ClinVar (release date: May 2021) with the Python lxml library. We then obtained allele frequencies and GERP scores for the ClinVar mutations from gnomAD v2 Exomes and dbNSFP4.2 [25,26]. gnomAD frequencies were joined to ClinVar entries by chromosome, hg38 coordinates, reference nucleotide, and alternative nucleotide. We joined the dbNSFP information to the mutation data by these fields as well as the reference and alternative amino acid. We then constructed reference and alternative C-termini for each mutation using protein sequences and IDs from the NCBI RefSeq database [27]. Each ClinVar entry contains a set of RefSeq IDs designating the protein isoforms in which the mutation occurs. Using these RefSeq IDs

as a set of keys, we joined the protein sequence data to the ClinVar entries. From there, we determined the alternative sequences and alternative C-termini caused by the mutations.

We subsequently prepared the minimotif data. First, we queried the MnM database for motifs on the C-termini. Second, we determined RefSeq IDs for the motif source proteins to facilitate a later join to the ClinVar mutations. Many table entries already had RefSeq IDs, but in some cases, motifs in the C-terminome and MnM data included only a UniProt ID for the source protein. Therefore, we acquired the corresponding NCBI RefSeq IDs for each entry in UniProtKB/SwissProt [28]. Then, we joined the RefSeq IDs to the motif entries by their listed UniProt IDs. When a RefSeq ID was linked to a non-human animal species, we matched the RefSeq ID to the human ortholog from the NCBI HomoloGene database where possible [29]. After determining the corresponding RefSeq IDs, we validated that each motif sequence was present in the C-terminus of the respective RefSeq protein. For that purpose, we used the RefSeq IDs to join the RefSeq sequence data to each motif. We then matched the motifs to the protein sequences with regular expressions. For motifs with PTM sites, we additionally verified that the amino acid of the PTM was present at the listed residue position. Lastly, we removed any duplicate motif entries between the C-terminome and MnM datasets.

In the final stage of our analysis, we compared the ClinVar mutation data and the motif data. The motif data were joined to the ClinVar data by RefSeq IDs. If a mutation occurred in the same protein as a motif, we determined if the mutation disrupted the motif via regular expression matching on the alternative C-terminus. We distinguished cases in which a mutation disrupted a minimotif without affecting the PTM site and cases where it directly substituted the site. We considered a nonsense mutation to disrupt a PTM site if the nonsense mutation occurred before the site residue. Our primary interest was in missense mutations that modified PTM sites, and we included an entry only if the ClinVar mutation was noted to be pathogenic or likely pathogenic. The mutations explored in our results are from this subset of the data.

We developed all figures using PowerPoint and the BioRender software tools. Figure 1 was developed in PowerPoint, whereas all remaining figures were created with BioRender. The graphical abstract was developed with a mixture of PowerPoint and BioRender.
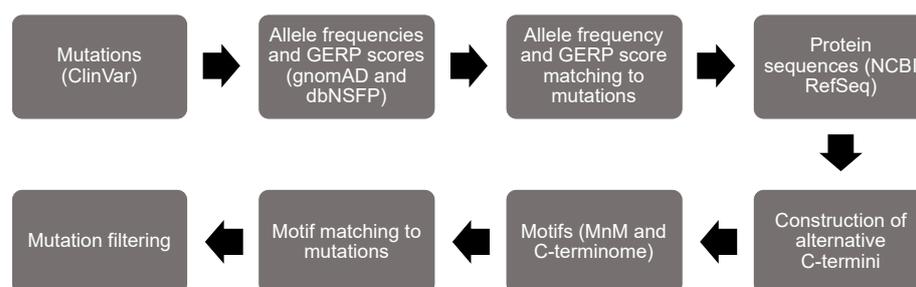

Figure 1. Flow chart of the bioinformatic analysis to identify minimotifs in the C-termini that are mutated in disease.

3. Results

We designed a structured approach to screen for minimotifs in the C-termini that are mutated in human diseases (Figure 1). The first step in the flowchart was to parse missense and nonsense mutations from the ClinVar database [23,24]. We then extracted variant data from gnomAD and dbNSFP and matched it to the ClinVar data [25,26]. From there, we acquired proteome sequence data from the National Center of Biotechnology Information (NCBI) RefSeq database [27]. Next, we matched mutations to C-termini of each protein and created the alternative C-termini. We then parsed motif data from MnM and the C-terminome databases [7,8,10], using UniProt protein data in the process [28]. The motif data were matched to the disease mutation data to identify motif disruption mutations. We focused on PTMs because the amino acid that is covalently modified can confidently be assigned to be a LOF mutation when the chemistry is not possible. For example, substitution

of a reference threonine with an alternative alanine necessarily prohibits phosphorylation. We also sorted out the subset of mutations classified as pathogenic or likely pathogenic. General statistics of our analysis are summarized in Table 2.

Table 2. ClinVar mutation statistics.

Category	Count
Total ClinVar variants	1,260,173
ClinVar nonsense and C-terminal missense variants	37,046
ClinVar C-terminal missense variants	5550
Pathogenic variants with PTM disrupted	20
Uncertain variants with PTM disrupted	119
Total with PTM disrupted	213
Pathogenic with motif but not PTM disrupted	166
Uncertain with motif but not PTM disrupted	782
Total with motif but not PTM disrupted	1457
Total C-terminal motifs	9165
Motifs mutated by a C-terminal variant	1258

Our computational search rediscovered several known mutations in the C-termini that cause one or more human diseases. The filtered set had 20 variant instances across 16 PTM sites. In a secondary screen, we examined each instance in detail to determine if we identified a new hypothesis for a disease mechanism. The nine new instances are included in Table 3. In five other instances, a PTM site was substituted, but we discerned that site disruption is an unlikely disease mechanism. We also identified six instances in C-termini that had mutated PTMs already known in disease but not yet summarized in a mutated motif paper. These were in CSF1R, FUS, GluN2A, and rhodopsin, and the disease mechanisms are summarized below.

Table 3. Variants discovered in our bioinformatics analysis.

Protein	Disease	Pathogenicity	Mutation	Motif	PTM	Frequency	GERP Score	Type
AR	Androgen resistance syndrome	Likely pathogenic	Y915S	KVKPIYFHTQ>	Phosphotyrosine	$<7.95 \times 10^6$	5.18	New hypothesis
APOC-III	NULL [1]	NULL [1]	T74A	PEVRPTSAVAA>	O-glycosylation	2.78×10^5	−0.501	New hypothesis
APOE	NULL [2]	NULL [2]	S296R	TSAAPVPSDNH>	O-glycosylation; phosphoserine	9.48×10^5	−0.0552	Unlikely
CRYM	Congenital sensorineural deafness	Pathogenic	K314T	YDSWSSGK>	Acetylation	$<7.95 \times 10^6$	5.45	New hypothesis
CSF1R	Hematologic neoplasm	Likely pathogenic	Y969F	LQPNNYQFC>	Phosphotyrosine	$<7.95 \times 10^6$	5.25	Rediscovery
CSF1R	Hematologic neoplasm	Likely pathogenic	Y969H	LQPNNYQFC>	Phosphotyrosine	$<7.95 \times 10^6$	5.25	Rediscovery
CSF1R	Hematologic neoplasm	Likely pathogenic	Y969C	LQPNNYQFC>	Phosphotyrosine	$<7.95 \times 10^6$	5.25	Rediscovery
FUS	Juvenile amyotrophic lateral sclerosis	Pathogenic	Y526C	QDRRERPY>	Phosphotyrosine	3.98×10^6	0.158	Rediscovery
GluN2A	Landau-Kleffner syndrome	Likely pathogenic	S1459G	RVYKKMPSIESDV>	Phosphoserine	$<7.95 \times 10^6$	5.79	Rediscovery

Table 3. *Cont.*

Protein	Disease	Pathogenicity	Mutation	Motif	PTM	Frequency	GERP Score	Type
GMPPB	Muscular dystrophy-dystroglycanopathy	Likely pathogenic	R357H	GESVPEPRIIM>	Methylation	3.72×10^5	5.24	New hypothesis
HBA	Familial erythrocytosis	Pathogenic	K139E	VSTVLTSKYR>	Acetylation	$<7.95 \times 10^6$	4.43	Unlikely
HBA	NULL [3]	NULL [3]	Y140H	VSTVLTSKYR>	Phosphotyrosine	$<7.95 \times 10^6$	2.8	Unlikely
HBB	Familial erythrocytosis	Pathogenic	K144N	VANALAHKYH>	Acetylation	$<7.95 \times 10^6$	−2.8	Unlikely
HBB	Familial erythrocytosis	Pathogenic	Y145H	VANALAHKYH>	Phosphotyrosine	$<7.95 \times 10^6$	4.68	New hypothesis
HBB	Familial erythrocytosis	Pathogenic	Y145C	VANALAHKYH>	Phosphotyrosine	$<7.95 \times 10^6$	4.68	New hypothesis
HBB	Familial erythrocytosis	Pathogenic	Y145N	VANALAHKYH>	Phosphotyrosine	$<7.95 \times 10^6$	4.68	New hypothesis
INS	Permanent neonatal diabetes mellitus	Pathogenic	Y108C	SLYQLENYCN>	Phosphotyrosine	$<7.95 \times 10^6$	3.58	Unlikely
RHO	Retinitis pigmentosa	Likely pathogenic	S343N	SKTETSQVAPA>	Phosphoserine; dephosphorylation	$<7.95 \times 10^6$	5.42	Rediscovery
SDHA	Mitochondrial complex II deficiency; paragangliomas 5	Likely pathogenic; uncertain significance	R662C	ATVPPAIRSY>	Methylation	$<7.95 \times 10^6$	4.12	New hypothesis
SMAD3	Familial thoracic aneurysm and aortic dissection	Likely pathogenic	S423N	SPSIRCSSVS>	Phosphoserine	$<7.95 \times 10^6$	4.97	New hypothesis

[1] Labeled in ClinVar as pathogenic for "Apolipoprotein C-III, non-glycosylated". [2] Labeled in ClinVar as pathogenic for "APOE4(+)". [3] Labeled in ClinVar as pathogenic for erythrocytosis but contradicted by the report identifying the mutation [30].

3.1. Rediscoveries of C-Terminal Minimotifs in Disease

3.1.1. CSF1R

Our bioinformatics search identified several likely pathogenic mutations for hematologic neoplasms (HNs) on the C-terminus of colony stimulation factor 1 receptor (CSF1R), all at the same phosphorylated tyrosine residue [23,24,31]. HNs are tumors of white blood cells [32], and CSF1R is causative of these tumors [33–38]. The protein has a largely identical primary structure to the v-fms oncogene of feline sarcoma virus. In v-fms, the C-terminal 40 residues of CSF1R are substituted for an 11-residue sequence, and 9 other positions in the protein are replaced by missense variants [39]. The resemblance implies a negative regulatory function of the CSF1R C-terminus on tumorigenesis [40]. The mutations Y969F, Y969H, and Y969C at the final residue in the protein are likely pathogenic for HNs [23,24,41–43]. CSF1R autophosphorylates Y969 upon binding of CSF1. Phosphorylation subsequently allows c-CBL to bind to CSF1R and terminate CSF1R activity. The Y969F mutation abolishes c-CBL binding, inducing excessive, tumorigenic CSF1R activity [44]. The mutation also instigates cell transformation in mice cells [45].

3.1.2. FUS

Another known C-terminal motif our analysis identified is a FUS mutation of a phosphorylation site that is causative for juvenile amyotrophic lateral sclerosis (JALS), a subtype of ALS. ALS is a neurodegenerative disease marked by fast deterioration of motor neurons that is typically lethal. Pathogenic FUS variants are especially prominent in JALS, and most occur in or near the nuclear localization signal in the C-terminus [46]. The pathogenic mutation Y526C is at the final residue within the nuclear localization signal [23,24,47]. Y526 is also a phosphorylation site that inhibits the FUS interaction with TNPO1, preventing nuclear translocation [31,48]. Therefore, the mechanism of JALS in Y526C is disruption of Y526 phosphorylation and its key role in FUS trafficking, leading to neurodegeneration.

3.1.3. GluN2A

Our bioinformatics investigation identified a mutation in the C-terminus of GluN2A, a neurotransmitter receptor that is pathogenic for Landau–Kleffner syndrome (LKS), a notably rare and intense form of inherited epilepsy [49]. More than 30 mutations of GluN2A are pathogenic for epilepsy [50]. GluN2A is a key subunit in many NMDA receptors, which are involved in excitatory neurotransmission, synaptic plasticity, and memory formation [51]. The mutation S1459G in GluN2A is causative of LKS [23,24,52]. S1459 phosphorylation is an essential step in NMDA-receptor trafficking to the postsynaptic density, and serine binds to PSD-95 after translocation and dephosphorylation [31,53]. As such, disruption of the S1459 phosphorylation site in GluN2 inhibits NMDA trafficking and its interaction with PSD-95, causing LKS.

3.1.4. Rhodopsin

The mutation S343N in a rhodopsin phosphorylation site in the C-terminus is likely pathogenic for retinitis pigmentosa [23,24], which induces progressive loss of vision over time. Deterioration of the rod cells of the retina causes the vision loss [54]. Expressed in rod cells, rhodopsin is a light receptor and transducer of the phototransduction cascade responsible for sight. After rhodopsin stimulation by light, retinal isomerizes, inducing conformational changes and phosphorylation of several threonine and serine residues in the C-terminus [55]. Phosphorylation of these residues, including S343, recruits arrestin and deactivates rhodopsin, terminating the signal [31,55–61]. In the absence of arrestin recruitment, excessive signaling causes neurodegeneration [58,62–64]. Therefore, S343N induces retinitis pigmentosa by disrupting a critical phosphorylation site and arrestin recruitment.

3.2. New Hypotheses for C-Terminal Minimotifs in Disease

In addition to minimotifs in the C-termini that were already mutated in disease, we also identified nine new instances where a PTM site was in the C-termini and mutated in a disease, but the connections were not previously recognized. Below, we summarize these new disease-causing hypotheses that will require further investigation to test their validity.

3.2.1. Androgen Receptor

Androgen insensitivity syndrome (AIS) is a disorder caused by inhibited androgen signaling (Figure 2A). Patients are genetically male and carry the disorder as an X-linked trait. Symptoms can include infertility, undescended testes, excess testosterone, and a female appearance. Mutations in the androgen receptor are the major cause of the disorder [65]. Androgens bind to residues in the C-terminal ligand-binding domain of the androgen receptor. The receptor then translocates to the nucleus, binds to the hormone response element (HRE) element in the promoter, and initiates transcription of genes integral to male sexual development. Androgen receptor mutations can disrupt this process, leading to AIS [66].
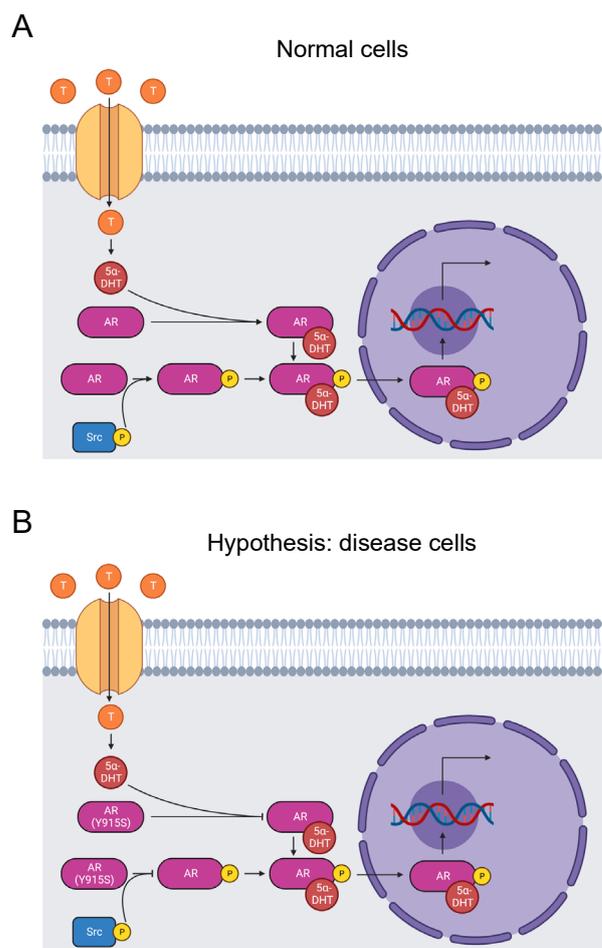
Figure 2. Diagram of the AR signaling pathway. (**A**) Testosterone enters prostate cells, where it converts to 5α-dihydrotestosterone. It binds to and activates the AR, which is phosphorylated on its ligand-binding domain by a Src kinase. The AR translocates to the nucleus and stimulates RNA transcription that promotes male sexual development [66]. Tyrosine phosphorylation by Src kinases can also activate the receptor independent of androgen binding [67]. (**B**) The mutation Y915S may cause AIS by abolishing tyrosine phosphorylation of the residue and disrupting androgen-independent AR activation. Alternatively, Y915S causes AIS by inhibiting androgen binding to the AR [23,24].

Cleaving the last 12 C-terminal residues of the androgen receptor eliminates binding of agonists and antagonists, indicating that the C-terminus is critical for function [68]. Likewise, missense mutations in the region cause AIS [69]. Site-directed mutagenesis of 10 of the last 13 residues identified that all but three of the 10 mutated residues contribute to ligand binding [70].

Y915 is one of the androgen receptor residues that affect ligand binding [70], and Y915S is causative of AIS [23,24]. Our bioinformatics analysis identified the Y915S mutation and phosphorylation of Y915 by Src kinases [31,67]. In fact, when the androgen receptor is tyrosine-phosphorylated, it is activated in an androgen-independent manner. It is not certain that Y915 is the key site because there are 10 other phosphorylated tyrosine sites throughout the protein [67]. Additionally, it is not yet known if androgen-independent activation is abolished by mutation of Y915. Given the pathogenic nature of Y915S, we propose the hypothesis that mutation of Y915 completely blocks Src kinase phosphorylation of this site and induces AIS by inhibiting androgen-independent activation (Figure 2B). Alternatively, Y915S may just block androgen binding. Further experimentation is needed to test these hypotheses.

3.2.2. APOC-III

Apolipoprotein C-III (APOC-III) is a key constituent of HDL, LDL, and VLDL, transporting lipids through the bloodstream. It additionally inhibits the activity of lipoprotein lipase and the uptake of triglycerides by hepatocytes [71]. Elevated bloodstream APOC-III is associated with cardiovascular diseases, hypertriglyceridemia, and diabetes [72,73].

The T74A mutation in the C-terminus of APOC-III abolishes its principal glycosylation site (Figure 3) [74]. The mutation was first observed in a normolipidemic, heterozygous Japanese family and is reported in ClinVar [23,24,75]. Most APOC-III molecules are glycosylated, where an O-linkage consisting of a single galactose molecule and a single N-acetylgalactosamine molecule is established (Figure 3) [76]. Glycosylation and sialylation of the O-linkage distinguish four glycoforms: APOC-III$_{0A}$ is non-glycosylated, APOC-III$_{0B}$ is non-sialylated, APOC-III$_1$ is monosialylated, and APOC-III$_2$ is disialylated [77].
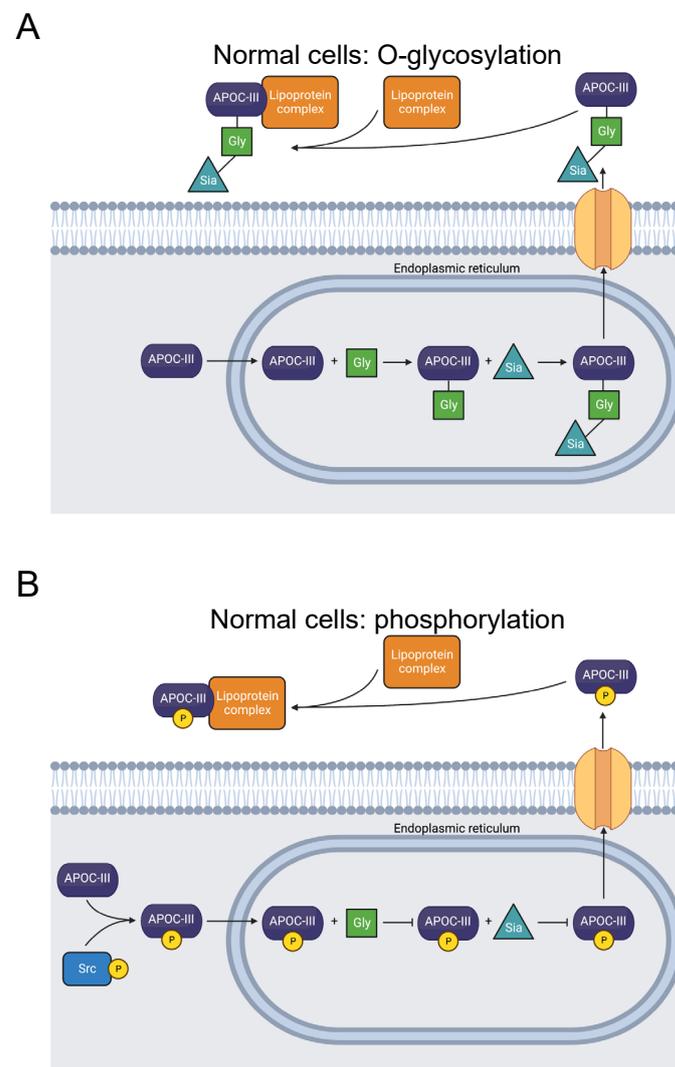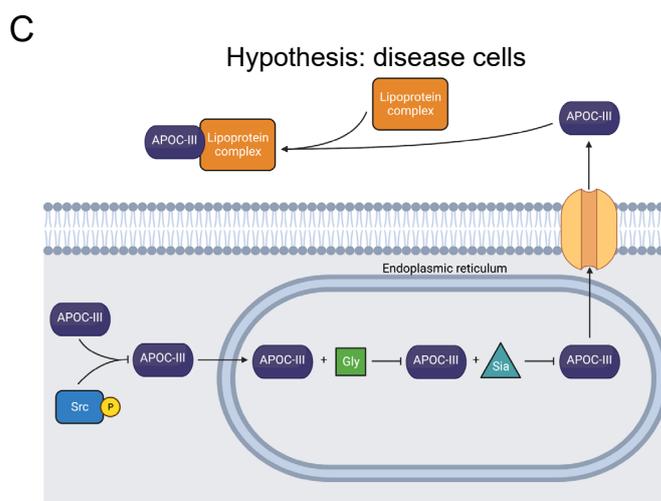

Figure 3. *Cont.*

Figure 3. Diagram showing O-glycosylation, sialylation, phosphorylation, and extracellular transport of APOC-III. (**A**) APOC-III is O-glycosylated and sialylated in the endoplasmic reticulum [76,77]. After extracellular transport, it becomes a constituent of lipoproteins, regulating lipid transport and metabolism [71]. (**B**) Alternatively, T74 is phosphorylated, precluding O-glycosylation [31]. (**C**) The mutation T74A abolishes the O-glycosylation and phosphorylation site [23,24,74,75]. Inhibited O-glycosylation and phosphorylation of APOC-III may be related to dysfunctional lipid metabolism, though secretion is not prevented [78].

Several studies have observed a correlation between a greater degree of APOC-III sialylation and reduced risk of atherosclerosis, as well as lower plasma triglyceride levels [77,79–85]. Others have reported the opposite effect: a link between increased APOC-III sialylation, hypertriglyceridemia, and diabetic hemodialysis [86–89]. Nevertheless, the function of APOC-III glycosylation is not yet fully understood, and there is conflicting evidence regarding its effect on disease. For example, an in vitro experiment expressing APOC-III with T74A indicated that the mutation does not inhibit APOC-III secretion and association with VLDL [78].

Our computational study revealed phosphorylation of T74 in APOC-III [31]. We expect that this PTM competes with T74 O-glycosylation (Figure 3). Therefore, we hypothesize that T74 phosphorylation blocks APOC-III glycosylation and sialylation and may affect lipid metabolism and lipidemic diseases. Studies of the site's phosphorylation and how it affects glycosylation and lipid particle physiology may clarify the pathology of APOC-III-related diseases.

3.2.3. CRYM

Nonsyndromic sensorineural deafness (NSD) is deafness resulting from dysfunction of the inner ear that is not symptomatic of an underlying syndrome. There are several genes associated with NSD, with one being μ-Crystallin. A screen of 192 Japanese families with NSD discovered causative mutations in μ-Crystallin [90]. Moreover, a Chinese family with NSD inherited an N-terminal P51L mutation, further supporting a role for μ-Crystallin in NSD pathology [91]. Thyroid hormone binds μ-Crystallin, stimulating nuclear translocation and activation of a thyroid-hormone signaling pathway. When μ-Crystallin binds $NADP^+$ and then to T_3, it is translocated to the nucleus (Figure 4A). Alternatively, when μ-Crystallin binds the reduced form of $NADP^+$, NADPH, translocation of the complex with T_3 is inhibited (Figure 4B) [92].
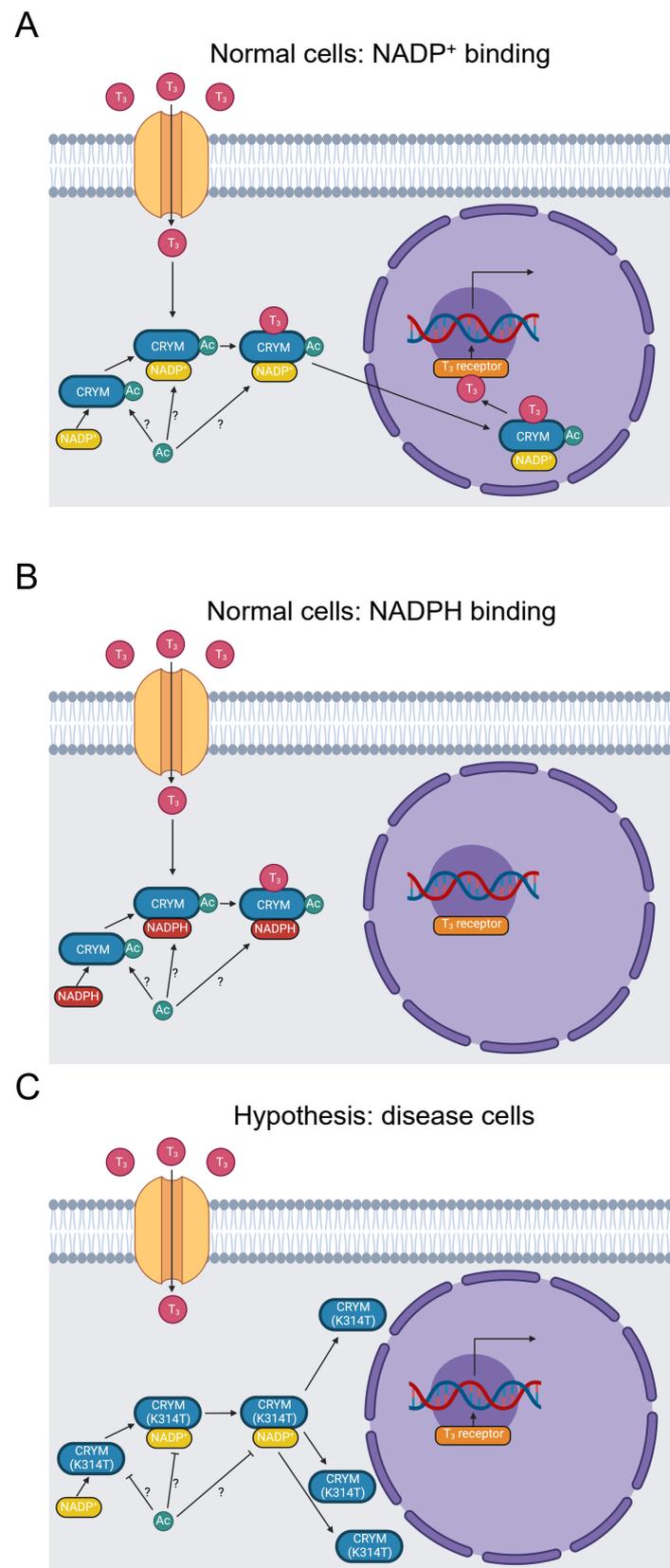
Figure 4. Diagram of thyroid-hormone trafficking by μ-Crystallin and subsequent transcription promoted by a thyroid-hormone receptor. (**A**) Thyroid hormone enters the cytoplasm and binds

to μ-Crystallin in the presence of NADP⁺. The complex translocates to the nucleus, where thyroid hormone binds to a receptor and stimulates RNA transcription [92]. K314 is acetylated [31]. Whether this occurs when μ-Crystallin is free, when it binds to NADP⁺, or when it binds to thyroid hormone is indeterminate, as indicated by the "?" symbol. (**B**) When NADPH binds to μ-Crystallin instead of NADP⁺, the complex fails to translocate to the nucleus [92]. (**C**) The K314T mutation abolishes thyroid-hormone binding to μ-Crystallin and results in a perinuclear distribution of μ-Crystallin [90,93]. It, therefore, prevents thyroid-hormone signaling downstream of μ-Crystallin binding. It further inhibits acetylation of K314.

Like the trafficking role of many C-termini [14], mutations in the C-terminus of μ-Crystallin interfere with its subcellular localization, which is critical for its thyroid-hormone signaling function. Abe et al. reported two variants in NSD patients in the C-terminus of μ-Crystallin. A stop loss mutation X315Y extends the C-terminus by five residues and is mislocalized to vacuoles in transfected COS-7 cells. Similarly, a K314T missense mutation is mislocalized perinuclearly. These mutations may also affect potassium ion recycling stimulated by thyroid-hormone activation of the Na,K-ATPase ion transporter [90].

The causative K314T mutation in the last residue of μ-Crystallin (YDSWSSGK>) is pathogenic for NSD [23,24,90]. The homologous residue in mouse μ-Crystallin is acetylated [31], suggesting our hypothesis that loss of K314 acetylation is pathogenic (Figure 4C). K314 acetylation is not yet linked to disease pathology. Disruption of K314 acetylation and the K314T mutation may contribute to NSD etiology by three possible mechanisms. First, the subcellular mislocalization established by Abe et al. may impede thyroid-hormone trafficking. Second, potassium-ion recycling may be disrupted, negatively impacting hearing function. Third, μ-Crystallin with the K314T mutation cannot bind to T₃ at all [93], which eliminates nuclear translocation of thyroid hormone by μ-Crystallin. These three mechanisms could potentially work in concert with each other. Additional work will be needed to clarify the role of K314 acetylation in these processes and the etiology of NSD.

3.2.4. GMPPB

Muscular dystrophy dystroglycanopathy (MDDG) is a major class of muscular disease. In MDDG, insufficient activity of α-dystroglycan weakens the stability of muscle and brain cell membranes [94,95]. The GDP-mannose pyrophosphorylase B (GMPPB) gene is causative of the disorder [94,96]. In zebrafish models, sufficient GMPPB enzymatic activity is necessary for normal neuronal and muscular development [97]. GMPPB is part of a complex that catalyzes the mannose-1-phosphate and GTP substrate conversion to the GDP-mannose product (Figure 5A) [98]. GDP-mannose reacts with Ser or Thr hydroxyl groups, O-mannosylating α-dystroglycan. α-dystroglycan interaction with extracellular proteins is dependent on its mannosylation. Thus, dysfunctional GMPPB activity induces MDDG through this pathway [94,99,100].

GMPPB mutations also cause a related disease called congenital myasthenic syndrome (CMS). CMS constitutes a distinct spectrum of neuromuscular disease, characterized by dysfunction of the neuromuscular joint [101]. GDP-mannose is a key precursor of N-glycans [94]. Deficient GDP-mannose production, therefore, likely impairs the glycosylation necessary for neuromuscular communication, instigating CMS. These two diseases have similar mechanisms involving GMPPB mutations, thus the proposed mechanism may also be related to CMS.

In total, there are 21 disease-associated variants of GMPPB in its C-terminal domain that are essential for interaction with the C-termini of other subunits in the catalytic GDP-mannose pyrophosphorylase complex. The complex consists of several GMPPA and GMPPB subunit homodimers and heterodimers. Therefore, C-terminal GMPPB mutations likely impair complex formation, producing MDDG pathology [98].

The mutation R357H in GMPPB's C-terminus was identified in five Chinese patients afflicted with MDDG and CMS [23,24,95]. These results suggest that the R357H mutation is

likely pathogenic for MDDG and CMS. R357H, like other MDDG missense mutations, is in the C-terminus of GMPPB that is critical for assembly of the catalytic complex.
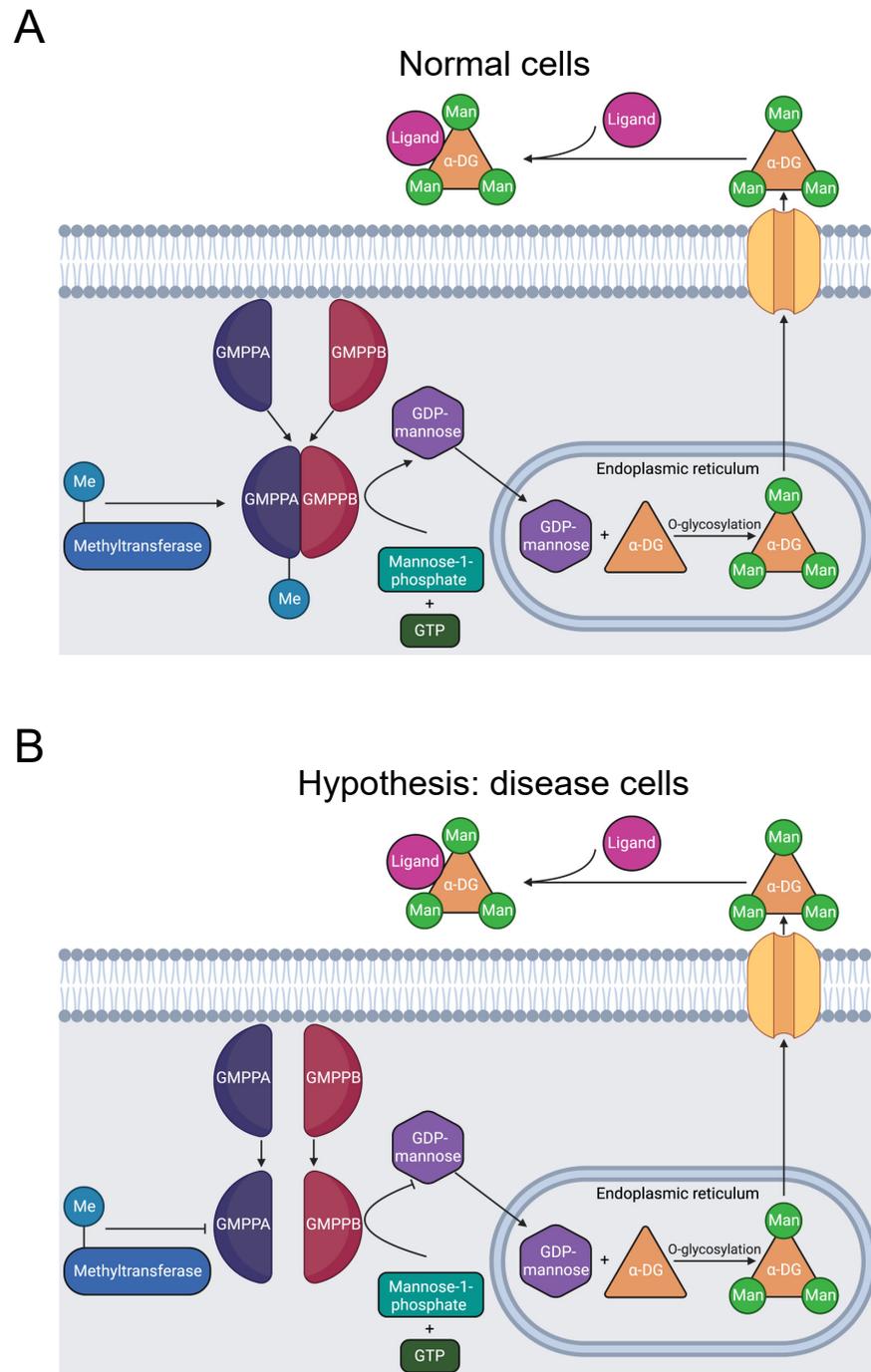

Figure 5. Diagram of GMPPB catalysis of α-DG O-mannosylation. (**A**) GMPPA and GMPPB subunits form a stable complex, including methylation of R357 [31]. The complex catalyzes the reaction of GTP and mannose-1-phosphate that produces GDP-mannose. In the endoplasmic reticulum, GDP-mannose O-mannosylates α-DG. Once transported to the extracellular matrix, α-DG binds to ligands and stimulates downstream signaling pathways [94]. (**B**) R357H inhibits R357 methylation and causes MDDG [23,24,95], likely by preventing stable complex formation [98]. The resulting instability reduces the quantity of GDP-mannose produced and decreases O-mannosylation of α-DG, causing MDDG [94,99,100].

Our bioinformatics investigation identified a separate study where R357 is methylated [31]. There are three possibilities: that methylation has no effect, promotes, or inhibits formation of the catalytic complex. Given that the mutation is likely pathogenic and that disrupted GMPPB activity is pathogenic, we hypothesize that R357 methylation is required for complex formation and that the R357H mutation inhibits methylation and contributes to muscular disorder etiology (Figure 5B). Further experimentation should be conducted to solidify the connection between R357 methylation and muscular dysfunction.

3.2.5. Hemoglobin

Secondary erythrocytosis is a disease characterized by abnormally high red blood cell (RBC) count [102]. Hemoglobin A, the principal form of hemoglobin in adult RBCs, is a heterotetramer composed of two hemoglobin alpha (HBA) and two hemoglobin beta (HBB) subunits [103]. Mutations in HBA and HBB are causative of erythrocytosis [104]. Over 100 HBB and HBA mutations increase the oxygen affinity of hemoglobin, often causing erythrocytosis. Many such mutations are concentrated in the C-termini [105,106].

In its unliganded T (tense) state, salt bridges on the C-termini stabilize hemoglobin A's heterotetrameric structure, which is made up of two HBB-HBA heterodimers (Figure 6A) [107–109]. As oxygen molecules bind to hemoglobin, critical interdimeric $\alpha_1\beta_2$ contacts rupture when a penultimate tyrosine residue on each chain is displaced from pockets between the F and H helices. The new conformer is called the R (relaxed) state. Additionally, the final residue in each HBA subunit forms salt bridges with the HBA subunit of the opposing dimer [109]. The penultimate tyrosine residues in HBB and HBA also participate in hydrogen bonds in the T state that are lost in the R state [109,110]. Therefore, hemoglobin mutations in the C-terminus destabilize the T state and favor the R state, raising its affinity for oxygen [104,111–117]. The consequent increase in oxygen affinity inhibits the release of oxygen to the tissues, provoking tissue hypoxia. The kidneys compensate by elevating erythropoietin secretion, stimulating increased RBC count [102,104].

The mutations Y145H (Hemoglobin Bethesda), Y145C (Hemoglobin Rainier), and Y145N (Hemoglobin Osler) in HBB induce erythrocytosis [23,24,118–124]. Residue Y145 corresponds to the penultimate residue of HBB and is also a phosphorylation site [31]. The coincidence of these mutations with a PTM site suggests our hypothesis that elimination of Y145 phosphorylation disrupts regulation of the R-to-T-state transition and causes erythrocytosis (Figure 6B). In addition to the role played by the tyrosine residues in interdimeric bonds, we propose that phosphorylation of Y145 contributes to the stability of the T-state heterotetramer. Therefore, inhibiting phosphorylation favors the R state and elevates oxygen affinity. Further experimental work should be pursued to test this hypothesis.

The mutations K144N (Hb Andrew-Minneapolis) in HBB and K139E (Hb Hanamaki) in HBA1 also produce erythrocytosis [23,24,125–128]. The residues K144 in HBB and K139 in HBA are acetylated [31]. However, studies of hemoglobin acetylation by aspirin, including K144, demonstrate either no effect on oxygen affinity or increased oxygen affinity [129–131]. Therefore, we do not believe that disrupted acetylation is a major factor in erythrocytosis etiology. Instead, these mutations likely cause disease by some other mechanism, such as destabilizing T-state salt bridges. We additionally did not consider the mutation Y140H (Hemoglobin Ethiopia/Rouen) in HBA, which substitutes a phosphorylated residue, since its probands displayed higher oxygen affinity but not compensatory erythrocytosis [23,24,30,31,132]. Given the lack of evidence of PTM disruption as a disease mechanism, we have included these three mutations in the "unlikely" category in Table 3.

3.2.6. SDHA

Defective mitochondrial complex II activity produces dysfunctional metabolism, causing mitochondrial complex II deficiency (MCIID). MCIID is a rare disease, accounting for 2% of patients with mitochondrial disease, and can cause neuropathies and other diseases [133,134]. Succinate dehydrogenase A (SDHA) is a subunit of mitochondrial complex II (SDH), a key enzyme in the tricarboxylic acid (TCA) cycle and oxidative phosphorylation

(Figure 7A) [133]. In addition to its role in cell metabolism, the SDH complex operates as a tumor suppressor [135].
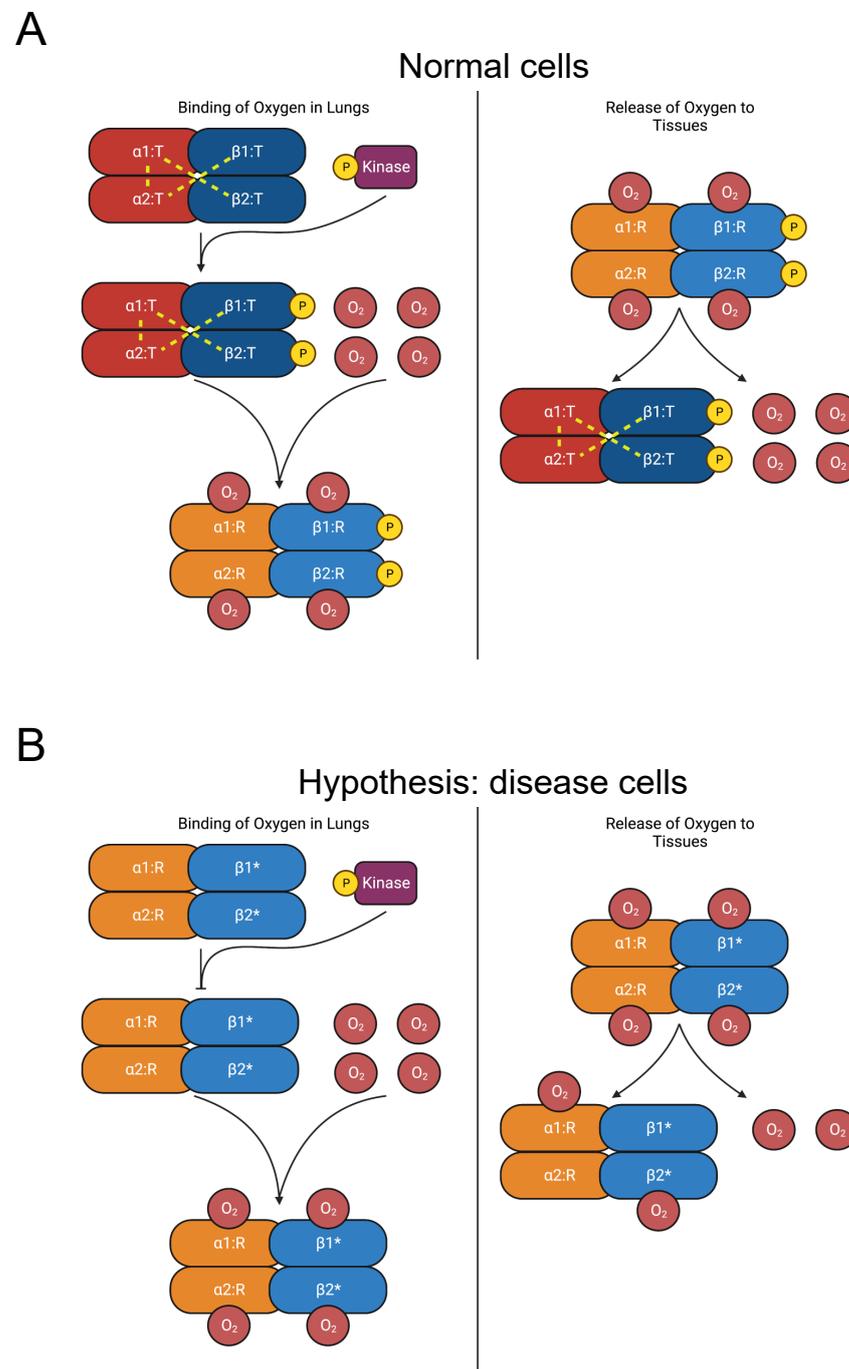

Figure 6. Diagram of oxygen binding to hemoglobin and subsequent release to tissues. (**A**) In the lungs, oxygen binds to hemoglobin, which is composed of two αβ heterodimers [107–109]. Y145 on the β chains is phosphorylated [31]. Interdimeric bonds that stabilize the T state are broken as oxygen binds, and the molecule transitions to the R state, increasing oxygen affinity [109,110]. In the tissues, oxygen is released, and hemoglobin reverts to the T state. (**B**) Mutations of Y145 (such as Y145H, Y145C, and Y145N, indicated by "β1*" and "β2*") disrupt interdimeric bonds and phosphorylation, favoring the R state over the T state and increasing oxygen affinity [104,111–117]. The consequent increase in oxygen affinity diminishes the release of oxygen to the tissues and provokes compensatory erythrocytosis [102,104].
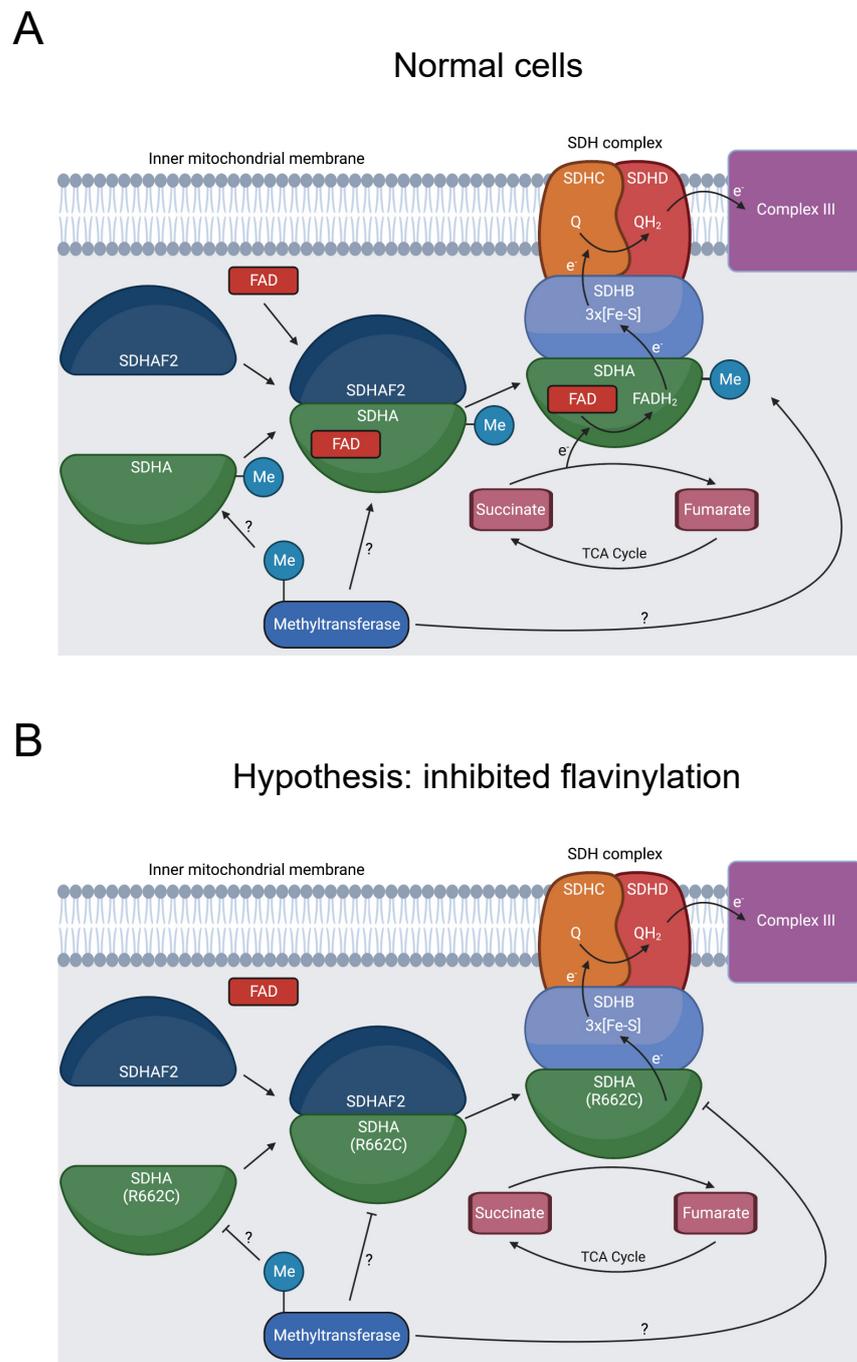
A

Normal cells

B

Hypothesis: inhibited flavinylation

Figure 7. *Cont.*

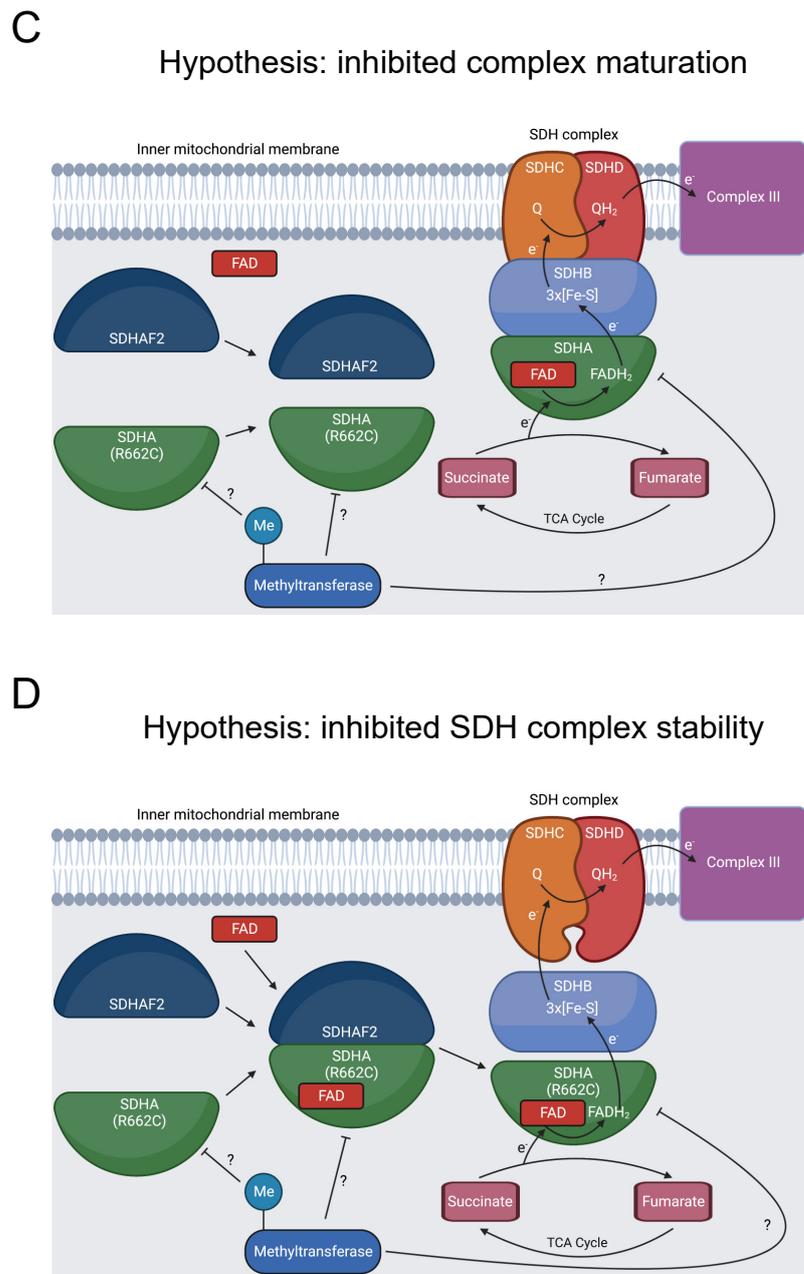**Figure 7.** Diagram showing formation of the SDHA-SDHAF2 complex, SDHA methylation, and SDH's role in oxidative phosphorylation and the TCA cycle. (**A**) SDHA and SDHAF2 form an intermediate complex, which promotes binding of the FAD cofactor to SDHA [136,137]. The full SDH complex then assembles. The heterotetramer operates as complex II in the electron transport chain and oxidizes succinate to fumarate in the TCA cycle [138]. At some point, R662 of SDHA is methylated [31]. This could occur prior to the SDHA-SDHAF2 complex formation, during SDHA-SDHAF2 complex formation, during SDH complex assembly, or after SDH complex assembly, as indicated by the "?" symbol. (**B**) R662C disrupts R662 methylation and precipitates MCIID [23,24,139]. A possible mechanism is disruption of SDHA flavinylation. (**C**) Another possible mechanism is inhibited complex maturation and formation of the SDHA-SDHAF2 complex. (**D**) A final proposed mechanism is increased instability of the full SDH complex.

Seventeen pathogenic variants of SDHA cause MCIID, and SDHA mutations can also induce familial paragangliomas and pheochromocytomas [135]. R662C is likely pathogenic for MCIID and was additionally present in a patient with a paraganglioma [23,24,139].

This position has multiple roles, affecting binding to the flavin adenine dinucleotide (FAD) cofactor, forming a salt bridge in a complex intermediate with the dedicated assembly factor subunit SDHAF2, and potentially contributing to the stability of the SDH complex [136,139,140].

Our bioinformatics analysis discovered that R662 is methylated [31]. We hypothesize that inhibited R662 methylation in the R662C mutant disrupts SDHA flavinylation (Figure 7B), complex maturation (Figure 7C), and/or SDH complex stability (Figure 7D), contributing to MCIID pathogenesis and paragangliomas. Further experimentation should be pursued to assess the impact of R662 methylation on these processes.

3.2.7. SMAD3

Familial thoracic aortic aneurysmal disease (FTAAD) is an inherited propensity for the development of aortic aneurysms. A thoracic aortic aneurysm occurs when the diameter of the thoracic aorta enlarges to 150% of its normal size. In extreme cases, an aneurysm precipitates dissection of the aorta [141]. Related conditions include Marfan syndrome and Loeys–Dietz syndrome [142].

Mutations of SMAD3 cause FTAAD in 2% of cases and are also causative of Marfan syndrome and Loeys–Dietz syndrome [143]. This is supported by studies of SMAD3 knockout mice, which experience greatly accelerated development of aortic aneurysms and death [144]. A 2018 study compiled 67 SMAD3 pathogenic and likely pathogenic variants associated with Loeys–Dietz syndrome [145], and a 2019 study of patients with SMAD3 mutations found that more than half experienced dissections [146].

SMAD3 is a key transducer of the TGF-β signaling pathway, and its C-terminus is critical for signal transduction (Figure 8A). Disruption of the pathway instigates aortic aneurysms. Upon activation, TGF-β binds to TβRII and TβRI receptors. TβRII then phosphorylates TβRI. Subsequently, TβRI phosphorylates two of the C-terminal serine residues in the SSXS> motif of SMAD2 or SMAD3. The activated SMAD2 or SMAD3 forms a complex with co-SMADs. The resulting complex translocates to the nucleus, where it regulates gene transcription [147].

Our analysis identified the likely pathogenic mutation S423N in the SSXS> motif of the C-terminus, along with S423 phosphorylation [31]. It was reported in ClinVar in two individuals with FTAAD, one of which also suffered Marfan syndrome [23,24]. Moreover, the aforementioned 2018 study also documented this mutation in Loeys–Dietz syndrome [145]. However, neither study recognized that this mutation disrupted this key signaling motif. Therefore, we hypothesize that this variant blocks SMAD3 phosphorylation and disrupts the TGF-β signaling pathway, resulting in disease of the aortic wall (Figure 8B).

3.3. Unlikely Hypotheses for C-Terminal Minimotifs in Disease

3.3.1. APOE

APOE is a constituent molecule of lipoproteins such as chylomicrons, HDL, LDL, and VLDL. It is widely present in the central nervous system, where it is secreted by astrocytes [148]. Possession of the APOE4 isoform is the greatest risk factor for Alzheimer's disease, and APOE dysfunction is related to coronary heart disease [149]. Our study identified an O-glycosylation and phosphorylation site at S296 that is abolished by an S296R mutation [23,24,31,148,150–152]. The function of O-glycosylation in APOE is disputed, with some studies linking greater glycosylation to superior health outcomes and others implying the opposite [79,153–156]. The protein kinase CK2 preferentially phosphorylates S296 in vitro. However, phosphorylation likely does not greatly impact APOE function in vivo since most APOE is lipidated in neurons and since there are no known kinases in the secretory pathway [152]. Similarly, disruption of APOE O-glycosylation does not impede secretion [157,158], and the mutation was found in a normolipidemic proband [150].
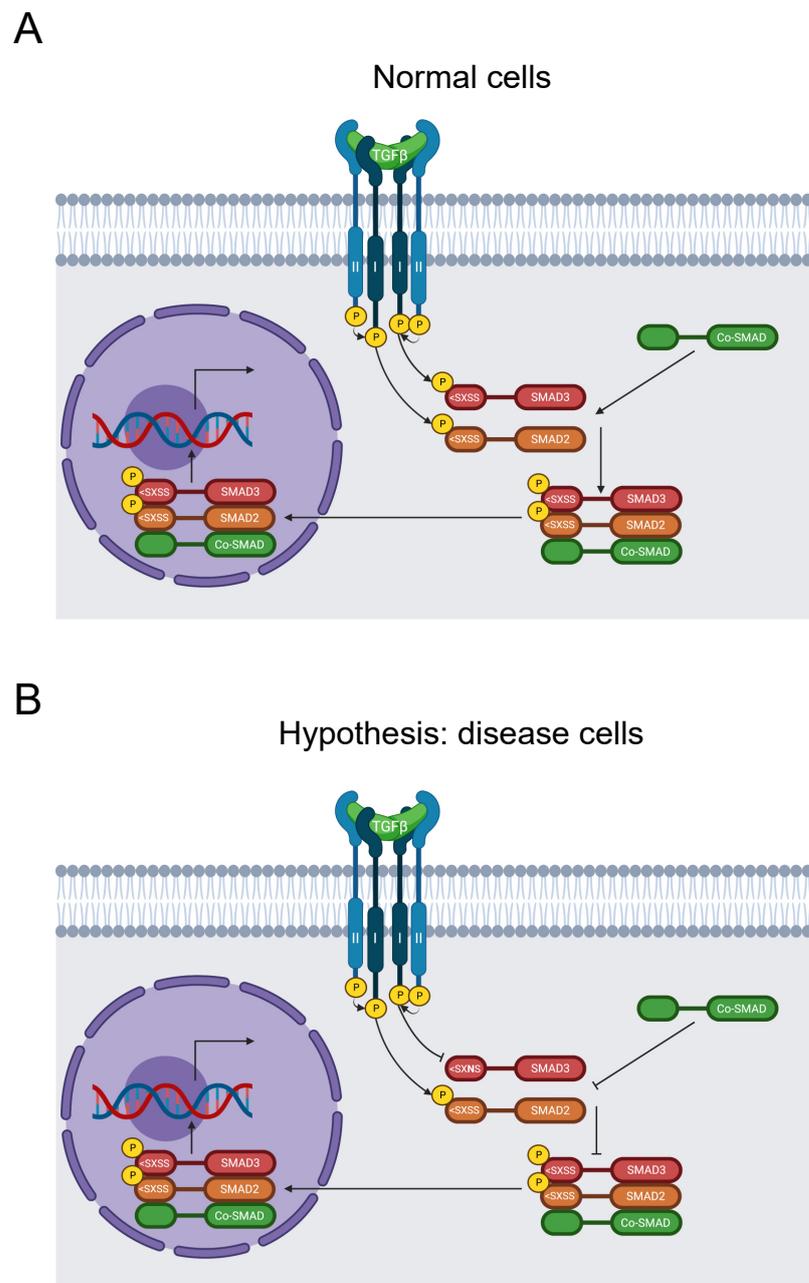
A

Normal cells

B

Hypothesis: disease cells

Figure 8. Diagram of the SMAD signaling pathway. (**A**) TGF-β binds to TβRII and subsequently recruits TβRI. TβRII subsequently activates TβRI by phosphorylation, which transduces the signal by phosphorylating SMAD2 and SMAD3 on their SXSS > motifs. Activated SMAD2 and SMAD3 then form a complex with a co-SMAD before translocating to the nucleus and stimulating gene expression [147]. (**B**) The S423N mutation disrupts the key SSXS > motif for SMAD3 phosphorylation and activation, thereby inhibiting further signal transduction and regulation of RNA transcription by the SMAD complex [147]. This causes FTAAD [23,24].

3.3.2. INS

Monogenic mutations are causative in about 1–2% of diabetes cases, and mutations in the *INS* gene are common [159]. The insulin mutation Y108C is pathogenic for congenital neonatal diabetes and juxtaposed to a disulfide bridge between the A and B chains [23,24]. Our bioinformatic analysis identified phosphorylation of Y108, which is also on the C-terminus (YCN>) [31]. While disruption of Y108 phosphorylation could contribute to insulin activity and cause congenital neonatal diabetes, insulin is very well-studied, there

are no known tyrosine kinases in the lumen of the secretory pathway, and insulin is not known to be tyrosine phosphorylated in vivo. Therefore, Y108 phosphorylation is not a likely mechanism in congenital neonatal diabetes etiology.

4. Discussion

Our analysis of PTMs in protein C-termini that are mutated in disease led to several conclusions:

1. Twenty missense mutations in PTM sites are reported to cause human disease.
2. We developed novel hypotheses for nine mutations in seven genes and PTM sites where disruption of a PTM causes a human disease.
3. The analysis of available databases for disease shows that the disruption of C-terminal PTMs, while present, is far less frequent than missense mutations throughout coding regions.

The mutation rate of our dataset estimates the prevalence of pathogenic mutations in PTM sites in protein C-termini. There are roughly 19,800 protein-coding genes in the human genome [160]. Considering the 20 mutations identified from our study of the C-termini in the human proteome (10 amino acids), the mutation rate is $\frac{20}{10*19800} = 0.0001$, or roughly 0.01%. Eliminating the 5 mutations in the "unlikely" category, the rate is instead about 0.008%. We can contrast this metric with the overall pathogenic missense mutation rate across the entire human proteome. There are currently 50,001 pathogenic and likely pathogenic missense variants in ClinVar [23,24], and there are at least 10.8 million amino acid residue positions in the human proteome [161]. This gives an estimated mutation rate of roughly 0.5%, though the rate is based on current knowledge and is possibly much higher. The overall pathogenic missense mutation rate is thus approximately 63 times higher than the pathogenic mutation rate of C-terminal PTM sites. These results suggest that PTM sites in the C-terminal region are highly conserved relative to the rest of the proteome and that these mutations are rare. However, the total rate of pathogenic mutations in minimotifs may be underestimated since we did not assess binding or trafficking motifs because we cannot conclusively assign an LOF mutation for most minimotifs.

All 20 of the reported mutations were classified as rare, either with a low frequency as estimated in gnomAD or not appearing at all. If a mutation was absent in gnomAD, we assigned it a frequency of <7.95 × 10^6 in Table 3. Genome evolutionary rate profiling (GERP) scores were used to assess the conservation of each mutation given that a highly conserved position is more likely to be disease-causing [162]. All but 4 mutations had a GERP above score above 2, with these 4 split among the unlikely category (n = 2), new hypothesis category (n = 1), and the rediscovery category (n = 1). The results of low frequencies and high GERP scores for most mutations that we identified are consistent with their pathogenic character.

We also assessed potential pathogenicity with scores from commonly used algorithms (SIFT, SIFT4G, PolyPhen-2 HumDiv, PolyPhen-2 HumVar, MPC, and CADD); see Supplementary Materials, Table S1 [163–169]. Scores were acquired from dbNSFP [25]. Most scores were consistent with the mutations being damaging. Only three mutations were predicted to be "tolerated" by SIFT. Similarly, SIFT4G predicted only two to be "tolerated." Thus, 85% and 90% of mutations were predicted to be damaging by SIFT and SIFT4G, respectively. Three mutations had no PolyPhen-2 predictions in dbNSFP. Of the mutations with PolyPhen-2 scores, approximately 76.5% and 58.8% were predicted to be damaging by PolyPhen-2 HumDiv and PolyPhen-2 HumVar, respectively. CADD PHRED-scaled scores were above the threshold of 20 for deleteriousness for 85% of mutations. On the other hand, most MPC scores were somewhat low on the 0 to 5 scale, with all but 2 mutations having a score below 2. The results from SIFT, SIFT4G, PolyPhen-2 HumDiv, and CADD support the pathogenic nature of the mutations in our study, though the MPC scores and PolyPhen-2 HumVar scores are less favorable.

Our analysis of C-terminal minimotifs in disease has several limitations. We only analyzed PTMs, and not binding or trafficking motifs, because of the simpler interpretation

with reliable assignment of LOF alleles. Moreover, we only analyzed the PTM site itself rather than other motif positions such as the S or T in the Nx(S/T) N-glycosylation site because the requirement for these positions is less certain. Furthermore, we had previously identified 8 C-terminal binding and trafficking sites that cause human disease; thus, a future study could investigate these other classes. The challenge is that the motif datasets would require annotation for which residues are required, and this is not yet available in the MnM and ELM databases. An additional limitation is that not all disease-causing mutations are in ClinVar.

We would like to conclude that mutation of C-terminal motifs is not a significant disease mechanism. Our data suggest that this is the case, with only dozens of causative mutations. However, the variant and motif databases we used are not yet comprehensive. Nevertheless, our investigation suggests that C-terminal motifs that are mutated and cause disease do occur but have low prevalence.

References

1. Choi, M.; Scholl, U.I.; Ji, W.; Liu, T.; Tikhonova, I.R.; Zumbo, P.; Nayir, A.; Bakkaloğlu, A.; Özen, S.; Sanjad, S.; et al. Genetic Diagnosis by Whole Exome Capture and Massively Parallel DNA Sequencing. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 19096–19101. [CrossRef] [PubMed]
2. Barrera, A.; Alastruey-Izquierdo, A.; Martín, M.J.; Cuesta, I.; Vizcaíno, J.A. Analysis of the Protein Domain and Domain Architecture Content in Fungi and Its Application in the Search of New Antifungal Targets. *PLoS Comput. Biol.* **2014**, *10*, e1003733. [CrossRef] [PubMed]
3. Lu, S.; Wang, J.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; Gwadz, M.; Hurwitz, D.I.; Marchler, G.H.; Song, J.S.; et al. CDD/SPARCLE: The Conserved Domain Database in 2020. *Nucleic Acids Res.* **2020**, *48*, D265–D268. [CrossRef] [PubMed]
4. Sigrist, C.J.A.; de Castro, E.; Cerutti, L.; Cuche, B.A.; Hulo, N.; Bridge, A.; Bougueleret, L.; Xenarios, I. New and Continuing Developments at PROSITE. *Nucleic Acids Res.* **2012**, *41*, D344–D347. [CrossRef] [PubMed]
5. Hulo, N. The PROSITE Database. *Nucleic Acids Res.* **2006**, *34*, D227–D230. [CrossRef] [PubMed]
6. Wang, Y.; Zhang, H.; Zhong, H.; Xue, Z. Protein Domain Identification Methods and Online Resources. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 1145–1153. [CrossRef]

7.   Lyon, K.F.; Cai, X.; Young, R.J.; Mamun, A.-A.; Rajasekaran, S.; Schiller, M.R. Minimotif Miner 4: A Million Peptide Minimotifs and Counting. *Nucleic Acids Res.* **2018**, *46*, D465–D470. [CrossRef]

8.   Balla, S.; Thapar, V.; Verma, S.; Luong, T.; Faghri, T.; Huang, C.-H.; Rajasekaran, S.; del Campo, J.J.; Shinn, J.H.; Mohler, W.A.; et al. Minimotif Miner: A Tool for Investigating Protein Function. *Nat Methods* **2006**, *3*, 175–177. [CrossRef] [PubMed]

9.   Kumar, M.; Gouw, M.; Michael, S.; Sámano-Sánchez, H.; Pancsa, R.; Glavina, J.; Diakogianni, A.; Valverde, J.A.; Bukirova, D.; Čalyševa, J.; et al. ELM—The Eukaryotic Linear Motif Resource in 2020. *Nucleic Acids Res.* **2020**, *48*, D296–D306. [CrossRef] [PubMed]

10. Sharma, S.; Toledo, O.; Hedden, M.; Lyon, K.F.; Brooks, S.B.; David, R.P.; Limtong, J.; Newsome, J.M.; Novakovic, N.; Rajasekaran, S.; et al. The Functional Human C-Terminome. *PLoS ONE* **2016**, *11*, e0152731. [CrossRef] [PubMed]

11. Kadaveru, K. Viral Infection and Human Disease—Insights from Minimotifs. *Front. Biosci.* **2008**, *13*, 6455. [CrossRef] [PubMed]

12. Sharma, S.; Young, R.J.; Chen, J.; Chen, X.; Oh, E.C.; Schiller, M.R. Minimotifs Dysfunction Is Pervasive in Neurodegenerative Disorders. *Alzheimer's Dement. Transl. Res. Clin. Interv.* **2018**, *4*, 414–432. [CrossRef] [PubMed]

13. Dampier, W.; Evans, P.; Ungar, L.; Tozeren, A. Host Sequence Motifs Shared by HIV Predict Response to Antiretroviral Therapy. *BMC Med. Genom.* **2009**, *2*, 47. [CrossRef] [PubMed]

14. Sharma, S.; Schiller, M.R. The Carboxy-Terminus, a Key Regulator of Protein Function. *Crit. Rev. Biochem. Mol. Biol.* **2019**, *54*, 85–102. [CrossRef]

15. Kupershmidt, S.; Yang, T.; Chanthaphaychith, S.; Wang, Z.; Towbin, J.A.; Roden, D.M. Defective Human Ether-à-Go-Go-Related Gene Trafficking Linked to an Endoplasmic Reticulum Retention Signal in the C Terminus. *J. Biol. Chem.* **2002**, *277*, 27442–27448. [CrossRef]

16. Deretic, D.; Williams, A.H.; Ransom, N.; Morel, V.; Hargrave, P.A.; Arendt, A. Rhodopsin C Terminus, the Site of Mutations Causing Retinal Disease, Regulates Trafficking by Binding to ADP-Ribosylation Factor 4 (ARF4). *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 3301–3306. [CrossRef]

17. Petris, M. A C-Terminal Di-Leucine Is Required for Localization of the Menkes Protein in the Trans-Golgi Network. *Hum. Mol. Genet.* **1998**, *7*, 2063–2071. [CrossRef]

18. Zaarour, N.; Demaretz, S.; Defontaine, N.; Zhu, Y.; Laghmani, K. Multiple Evolutionarily Conserved Di-Leucine Like Motifs in the Carboxyl Terminus Control the Anterograde Trafficking of NKCC2. *J. Biol. Chem.* **2012**, *287*, 42642–42653. [CrossRef]

19. Lam, P.; Xu, S.; Soroka, C.J.; Boyer, J.L. A C-Terminal Tyrosine-Based Motif in the Bile Salt Export Pump Directs Clathrin-Dependent Endocytosis. *Hepatology* **2012**, *55*, 1901–1911. [CrossRef]

20. Reiners, J.; Nagel-Wolfrum, K.; Jürgens, K.; Märker, T.; Wolfrum, U. Molecular Basis of Human Usher Syndrome: Deciphering the Meshes of the Usher Protein Network Provides Insights into the Pathomechanisms of the Usher Disease. *Exp. Eye Res.* **2006**, *83*, 97–119. [CrossRef]

21. Müller, D.; Kausalya, P.J.; Claverie-Martin, F.; Meij, I.C.; Eggert, P.; Garcia-Nieto, V.; Hunziker, W. A Novel Claudin 16 Mutation Associated with Childhood Hypercalciuria Abolishes Binding to ZO-1 and Results in Lysosomal Mistargeting. *Am. J. Hum. Genet.* **2003**, *73*, 1293–1301. [CrossRef] [PubMed]

22. Shy, D.; Gillet, L.; Abriel, H. Cardiac Sodium Channel NaV1.5 Distribution in Myocytes via Interacting Proteins: The Multiple Pool Model. *Biochim. Et Biophys. Acta (BBA) Mol. Cell Res.* **2013**, *1833*, 886–894. [CrossRef]

23. Landrum, M.J.; Lee, J.M.; Benson, M.; Brown, G.; Chao, C.; Chitipiralla, S.; Gu, B.; Hart, J.; Hoffman, D.; Hoover, J.; et al. ClinVar: Public Archive of Interpretations of Clinically Relevant Variants. *Nucleic Acids Res.* **2016**, *44*, D862–D868. [CrossRef] [PubMed]

24. Landrum, M.J.; Chitipiralla, S.; Brown, G.R.; Chen, C.; Gu, B.; Hart, J.; Hoffman, D.; Jang, W.; Kaur, K.; Liu, C.; et al. ClinVar: Improvements to Accessing Data. *Nucleic Acids Res.* **2020**, *48*, D835–D844. [CrossRef] [PubMed]

25. Liu, X.; Li, C.; Mou, C.; Dong, Y.; Tu, Y. DbNSFP v4: A Comprehensive Database of Transcript-Specific Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Genome Med.* **2020**, *12*, 103. [CrossRef]

26. Karczewski, K.J.; Francioli, L.C.; Tiao, G.; Cummings, B.B.; Alföldi, J.; Wang, Q.; Collins, R.L.; Laricchia, K.M.; Ganna, A.; Birnbaum, D.P.; et al. The Mutational Constraint Spectrum Quantified from Variation in 141,456 Humans. *Nature* **2020**, *581*, 434–443. [CrossRef]

27. O'Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; et al. Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation. *Nucleic Acids Res.* **2016**, *44*, D733–D745. [CrossRef]

28. Bateman, A.; Martin, M.-J.; Orchard, S.; Magrane, M.; Agivetova, R.; Ahmad, S.; Alpi, E.; Bowler-Barnett, E.H.; Britto, R.; The UniProt Consortium; et al. UniProt: The Universal Protein Knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480–D489. [CrossRef]

29. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2016**, *44*, D7–D19. [CrossRef]

30. Wajcman, H.; Kister, J.; Marden, M.; Lahary, A.; Monconduit, M.; Galacteros, F. Homoglobin Rouen (α-140 (HC2) Tyr→His): Alteration of the α-Chain C-Terminal Region and Moderate Increase in Oxygen Affinity. *Biochim. Et Biophys. Acta (BBA) Mol. Basis Dis.* **1992**, *1180*, 53–57. [CrossRef]

31. Hornbeck, P.V.; Kornhauser, J.M.; Tkachev, S.; Zhang, B.; Skrzypek, E.; Murray, B.; Latham, V.; Sullivan, M. PhosphoSitePlus: A Comprehensive Resource for Investigating the Structure and Function of Experimentally Determined Post-Translational Modifications in Man and Mouse. *Nucleic Acids Res.* **2012**, *40*, D261–D270. [CrossRef] [PubMed]

32. Zierhut, M.; Haen, S.P.; Moehle, R.; Chan, C.-C. Hematological Neoplasms. In *Intraocular Inflammation*; Zierhut, M., Pavesio, C., Ohno, S., Orefice, F., Rao, N.A., Eds.; Springer: Berlin/Heidelberg, Germany, 2016; pp. 1493–1510, ISBN 978-3-540-75385-8.

33. Murga-Zamalloa, C.; Rolland, D.C.M.; Polk, A.; Wolfe, A.; Dewar, H.; Chowdhury, P.; Onder, O.; Dewar, R.; Brown, N.A.; Bailey, N.G.; et al. Colony-Stimulating Factor 1 Receptor (CSF1R) Activates AKT/MTOR Signaling and Promotes T-Cell Lymphoma Viability. *Clin. Cancer Res.* **2020**, *26*, 690–703. [CrossRef] [PubMed]

34. Martín-Moreno, A.M.; Roncador, G.; Maestre, L.; Mata, E.; Jiménez, S.; Martínez-Torrecuadrada, J.L.; Reyes-García, A.I.; Rubio, C.; Tomás, J.F.; Estévez, M.; et al. CSF1R Protein Expression in Reactive Lymphoid Tissues and Lymphoma: Its Relevance in Classical Hodgkin Lymphoma. *PLoS ONE* **2015**, *10*, e0125203. [CrossRef]

35. Ullrich, K.; Wurster, K.D.; Lamprecht, B.; Köchert, K.; Engert, A.; Dörken, B.; Janz, M.; Mathas, S. BAY 43-9006/Sorafenib Blocks CSF1R Activity and Induces Apoptosis in Various Classical Hodgkin Lymphoma Cell Lines. *Br. J. Haematol.* **2011**, *155*, 398–402. [CrossRef]

36. Lamprecht, B.; Walter, K.; Kreher, S.; Kumar, R.; Hummel, M.; Lenze, D.; Köchert, K.; Bouhlel, M.A.; Richter, J.; Soler, E.; et al. Derepression of an Endogenous Long Terminal Repeat Activates the CSF1R Proto-Oncogene in Human Lymphoma. *Nat. Med.* **2010**, *16*, 571–579. [CrossRef] [PubMed]

37. Skinnider, B.F.; Mak, T.W. The Role of Cytokines in Classical Hodgkin Lymphoma. *Blood* **2002**, *99*, 4283–4297. [CrossRef]

38. Komohara, Y.; Noyori, O.; Saito, Y.; Takeya, H.; Baghdadi, M.; Kitagawa, F.; Hama, N.; Ishikawa, K.; Okuno, Y.; Nosaka, K.; et al. Potential Anti-Lymphoma Effect of M-CSFR Inhibitor in Adult T-Cell Leukemia/Lymphoma. *JCEH* **2018**, *58*, 152–160. [CrossRef]

39. Woolford, J.; McAuliffe, A.; Rohrschneider, L.R. Activation of the Feline C-Fms Proto-Oncogene: Multiple Alterations Are Required to Generate a Fully Transformed Phenotype. *Cell* **1988**, *55*, 965–977. [CrossRef]

40. Browning, P.J.; Bunn, H.F.; Cline, A.; Shuman, M.; Nienhuis, A.W. "Replacement" of COOH-Terminal Truncation of v-Fms with c-Fms Sequences Markedly Reduces Transformation Potential. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 7800–7804. [CrossRef]

41. Ridge, S.A.; Worwood, M.; Oscier, D.; Jacobs, A.; Padua, R.A. FMS Mutations in Myelodysplastic, Leukemic, and Normal Subjects. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 1377–1380. [CrossRef]

42. Tobal, K.; Pagliuca, A.; Bhatt, B.; Bailey, N.; Layton, D.M.; Mufti, G.J. Mutation of the Human FMS Gene (M-CSF Receptor) in Myelodysplastic Syndromes and Acute Myeloid Leukemia. *Leukemia* **1990**, *4*, 486–489.

43. Yu, W.; Chen, J.; Xiong, Y.; Pixley, F.J.; Dai, X.-M.; Yeung, Y.-G.; Stanley, E.R. CSF-1 Receptor Structure/Function in MacCsf1r-/- Macrophages: Regulation of Proliferation, Differentiation, and Morphology. *J. Leukoc. Biol.* **2008**, *84*, 852–863. [CrossRef]

44. Wilhelmsen, K.; Burkhalter, S.; van der Geer, P. C-Cbl Binds the CSF-1 Receptor at Tyrosine 973, a Novel Phosphorylation Site in the Receptor's Carboxy-Terminus. *Oncogene* **2002**, *21*, 1079–1089. [CrossRef] [PubMed]

45. Roussel, M.F.; Dull, T.J.; Rettenmier, C.W.; Ralph, P.; Ullrich, A.; Sherr, C.J. Transforming Potential of the C-Fms Proto-Oncogene (CSF-1 Receptor). *Nature* **1987**, *325*, 549–552. [CrossRef] [PubMed]

46. Sharma, A.; Lyashchenko, A.K.; Lu, L.; Nasrabady, S.E.; Elmaleh, M.; Mendelsohn, M.; Nemes, A.; Tapia, J.C.; Mentis, G.Z.; Shneider, N.A. ALS-Associated Mutant FUS Induces Selective Motor Neuron Degeneration through Toxic Gain of Function. *Nat. Commun.* **2016**, *7*, 10465. [CrossRef]

47. Corcia, P.; Danel, V.; Lacour, A.; Beltran, S.; Andres, C.; Couratier, P.; Blasco, H.; Vourc'h, P. A Novel Mutation of the C-Terminal Amino Acid of *FUS* (Y526C) Strengthens *FUS* Gene as the Most Frequent Genetic Factor in Aggressive Juvenile ALS. *Amyotroph. Lateral Scler. Front. Degener.* **2017**, *18*, 298–301. [CrossRef] [PubMed]

48. Darovic, S.; Mihevc, S.P.; Župunski, V.; Gunčar, G.; Štalekar, M.; Lee, Y.; Shaw, C.E.; Rogelj, B. Phosphorylation of C-Terminal Tyrosine 526 in FUS Impairs Its Nuclear Import. *J. Cell Sci.* **2015**, *128*, 4151–4159. [CrossRef]

49. Pearl, P.L.; Carrazana, E.J.; Holmes, G.L. The Landau-Kleffner Syndrome. *Epilepsy Curr.* **2001**, *1*, 39–45. [CrossRef]

50. Addis, L.; Virdee, J.K.; Vidler, L.R.; Collier, D.A.; Pal, D.K.; Ursu, D. Epilepsy-Associated GRIN2A Mutations Reduce NMDA Receptor Trafficking and Agonist Potency—Molecular Profiling and Functional Rescue. *Sci. Rep.* **2017**, *7*, 66. [CrossRef]

51. Sapkota, K.; Dore, K.; Tang, K.; Irvine, M.; Fang, G.; Burnell, E.S.; Malinow, R.; Jane, D.E.; Monaghan, D.T. The NMDA Receptor Intracellular C-Terminal Domains Reciprocally Interact with Allosteric Modulators. *Biochem. Pharmacol.* **2019**, *159*, 140–153. [CrossRef]

52. Bowling, K.M.; Thompson, M.L.; Amaral, M.D.; Finnila, C.R.; Hiatt, S.M.; Engel, K.L.; Cochran, J.N.; Brothers, K.B.; East, K.M.; Gray, D.E.; et al. Genomic Diagnosis for Children with Intellectual Disability and/or Developmental Delay. *Genome Med.* **2017**, *9*, 43. [CrossRef]

53. Mota Vieira, M.; Nguyen, T.A.; Wu, K.; Badger, J.D.; Collins, B.M.; Anggono, V.; Lu, W.; Roche, K.W. An Epilepsy-Associated GRIN2A Rare Variant Disrupts CaMKIIα Phosphorylation of GluN2A and NMDA Receptor Trafficking. *Cell Rep.* **2020**, *32*, 108104. [CrossRef]

54. Nash, B.M.; Wright, D.C.; Grigg, J.R.; Bennetts, B.; Jamieson, R.V. Retinal Dystrophies, Genomic Applications in Diagnosis and Prospects for Therapy. *Transl. Pediatr.* **2015**, *4*, 139–163. [CrossRef] [PubMed]

55. Zhou, X.E.; Melcher, K.; Xu, H.E. Structure and Activation of Rhodopsin. *Acta Pharm. Sin* **2012**, *33*, 291–299. [CrossRef] [PubMed]

56. Kennedy, M.J.; Lee, K.A.; Niemi, G.A.; Craven, K.B.; Garwin, G.G.; Saari, J.C.; Hurley, J.B. Multiple Phosphorylation of Rhodopsin and the In Vivo Chemistry Underlying Rod Photoreceptor Dark Adaptation. *Neuron* **2001**, *31*, 87–101. [CrossRef] [PubMed]

57. Ohguro, H.; Rudnicka-Nawrot, M.; Buczyko, J.; Zhao, X.; Taylor, J.A.; Walsh, K.A.; Palczewski, K. Structural and Enzymatic Aspects of Rhodopsin Phosphorylation. *J. Biol. Chem.* **1996**, *271*, 5215–5224. [CrossRef] [PubMed]

58.   Azevedo, A.W.; Doan, T.; Moaven, H.; Sokal, I.; Baameur, F.; Vishnivetskiy, S.A.; Homan, K.T.; Tesmer, J.J.; Gurevich, V.V.; Chen, J.; et al. C-Terminal Threonines and Serines Play Distinct Roles in the Desensitization of Rhodopsin, a G Protein-Coupled Receptor. *eLife* **2015**, *4*, e05981. [CrossRef] [PubMed]

59.   Vishnivetskiy, S.A.; Raman, D.; Wei, J.; Kennedy, M.J.; Hurley, J.B.; Gurevich, V.V. Regulation of Arrestin Binding by Rhodopsin Phosphorylation Level. *J. Biol. Chem.* **2007**, *282*, 32075–32083. [CrossRef] [PubMed]

60.   Bownds, D.; Dawes, J.; Miller, J.; Stahlman, M. Phosphorylation of Frog Photoreceptor Membranes Induced by Light. *Nat. New Biol.* **1972**, *237*, 125–127. [CrossRef] [PubMed]

61.   Kühn, H.; Dreyer, W.J. Light Dependent Phosphorylation of Rhodopsin by ATP. *FEBS Lett.* **1972**, *20*, 1–6. [CrossRef] [PubMed]

62.   Dolph, P.J.; Ranganathan, R.; Colley, N.J.; Hardy, R.W.; Socolich, M.; Zuker, C.S. Arrestin Function in Inactivation of G Protein-Coupled Receptor Rhodopsin in Vivo. *Science* **1993**, *260*, 1910–1916. [CrossRef] [PubMed]

63.   Lee, E.S.; Flannery, J.G. Transport of Truncated Rhodopsin and Its Effects on Rod Function and Degeneration. *Investig. Ophthalmol. Vis. Sci.* **2007**, *48*, 2868. [CrossRef] [PubMed]

64.   Baylor, D.A.; Burns, M.E. Control of Rhodopsin Activity in Vision. *Eye* **1998**, *12*, 521–525. [CrossRef] [PubMed]

65.   Hughes, I.A.; Davies, J.D.; Bunch, T.I.; Pasterski, V.; Mastroyannopoulou, K.; MacDougall, J. Androgen Insensitivity Syndrome. *Lancet* **2012**, *380*, 1419–1428. [CrossRef]

66.   Tan, M.E.; Li, J.; Xu, H.E.; Melcher, K.; Yong, E. Androgen Receptor: Structure, Role in Prostate Cancer and Drug Discovery. *Acta Pharm. Sin* **2015**, *36*, 3–23. [CrossRef] [PubMed]

67.   Guo, Z.; Dai, B.; Jiang, T.; Xu, K.; Xie, Y.; Kim, O.; Nesheiwat, I.; Kong, X.; Melamed, J.; Handratta, V.D.; et al. Regulation of Androgen Receptor Activity by Tyrosine Phosphorylation. *Cancer Cell* **2006**, *10*, 309–319. [CrossRef]

68.   Jenster, G.; van der Korput, H.A.G.M.; van Vroonhoven, C.; van der Kwast, T.H.; Trapman, J.; Brinkmann, A.O. Domains of the Human Androgen Receptor Involved in Steroid Binding, Transcriptional Activation, and Subcellular Localization. *Mol. Endocrinol.* **1991**, *5*, 1396–1404. [CrossRef]

69.   Gottlieb, B.; Beitel, L.K.; Nadarajah, A.; Paliouras, M.; Trifiro, M. The Androgen Receptor Gene Mutations Database: 2012 Update. *Hum. Mutat.* **2012**, *33*, 887–894. [CrossRef]

70.   Tahiri, B.; Auzou, G.; Nicolas, J.-C.; Sultan, C.; Lupo, B. Participation of Critical Residues from the Extreme C-Terminal End of the Human Androgen Receptor in the Ligand Binding Function. *Biochemistry* **2001**, *40*, 8431–8437. [CrossRef]

71.   Fredenrich, A.; Giroux, L.M.; Tremblay, M.; Krimbou, L.; Davignon, J.; Cohn, J.S. Plasma Lipoprotein Distribution of ApoC-III in Normolipidemic and Hypertriglyceridemic Subjects: Comparison of the ApoC-III to ApoE Ratio in Different Lipoprotein Fractions. *J. Lipid Res.* **1997**, *38*, 1421–1432. [CrossRef]

72.   Borén, J.; Packard, C.J.; Taskinen, M.-R. The Roles of ApoC-III on the Metabolism of Triglyceride-Rich Lipoproteins in Humans. *Front. Endocrinol.* **2020**, *11*, 474. [CrossRef] [PubMed]

73.   Rodríguez, M.; Rehues, P.; Iranzo, V.; Mora, J.; Balsells, C.; Guardiola, M.; Ribalta, J. Distribution of Seven ApoC-III Glycoforms in Plasma, VLDL, IDL, LDL and HDL of Healthy Subjects. *J. Proteom.* **2022**, *251*, 104398. [CrossRef] [PubMed]

74.   Maeda, H.; Hashimoto, R.K.; Ogura, T.; Hiraga, S.; Uzawa, H. Molecular Cloning of a Human ApoC-III Variant: Thr 74—-Ala 74 Mutation Prevents O-Glycosylation. *J Lipid Res* **1987**, *28*, 1405–1409. [CrossRef] [PubMed]

75.   Maeda, H.; Uzawa, H.; Kamei, R. Unusual Familial Lipoprotein C-III Associated with Apolipoprotein C-III-0 Preponderance. *Biochim. Et Biophys. Acta (BBA) Lipids Lipid Metab.* **1981**, *665*, 578–585. [CrossRef]

76.   Jian, W.; Edom, R.W.; Wang, D.; Weng, N.; Zhang, S. (Weihua) Relative Quantitation of Glycoisoforms of Intact Apolipoprotein C3 in Human Plasma by Liquid Chromatography–High-Resolution Mass Spectrometry. *Anal. Chem.* **2013**, *85*, 2867–2874. [CrossRef]

77.   Yassine, H.N.; Trenchevska, O.; Ramrakhiani, A.; Parekh, A.; Koska, J.; Walker, R.W.; Billheimer, D.; Reaven, P.D.; Yen, F.T.; Nelson, R.W.; et al. The Association of Human Apolipoprotein C-III Sialylation Proteoforms with Plasma Triglycerides. *PLoS ONE* **2015**, *10*, e0144138. [CrossRef]

78.   Roghani, A.; Zannis, V.I. Mutagenesis of the Glycosylation Site of Human ApoCIII. O-Linked Glycosylation Is Not Required for ApoCIII Secretion and Lipid Binding. *J. Biol. Chem.* **1988**, *263*, 17925–17932. [CrossRef]

79.   Savinova, O.V.; Fillaus, K.; Jing, L.; Harris, W.S.; Shearer, G.C. Reduced Apolipoprotein Glycosylation in Patients with the Metabolic Syndrome. *PLoS ONE* **2014**, *9*, e104833. [CrossRef] [PubMed]

80.   Tertov, V.V.; Sobenin, I.A.; Tonevitsky, A.G.; Orekhov, A.N.; Smirnov, V.N. Isolation of Atherogenic Modified (Desialylated) Low Density Lipoprotein from Blood of Atherosclerotic Patients: Separation from Native Lipoprotein by Affinity Chromatography. *Biochem. Biophys. Res. Commun.* **1990**, *167*, 1122–1127. [CrossRef]

81.   Orekhov, A.N.; Tertov, V.V.; Mukhin, D.N.; Mikhailenko, I.A. Modification of Low Density Lipoprotein by Desialylation Causes Lipid Accumulation in Cultured Cells: Discovery of Desialylated Lipoprotein with Altered Cellular Metabolism in the Blood of Atherosclerotic Patients. *Biochem. Biophys. Res. Commun.* **1989**, *162*, 206–211. [CrossRef]

82.   Tertov, V.V.; Orekhov, A.N.; Sobenin, I.A.; Morrisett, J.D.; Gotto, A.M.; Guevara, J.G. Carbohydrate Composition of Protein and Lipid Components in Sialic Acid-Rich and -Poor Low Density Lipoproteins from Subjects with and without Coronary Artery Disease. *J. Lipid Res.* **1993**, *34*, 365–375. [CrossRef]

83.   Millar, J.S.; Anber, V.; Shepherd, J.; Packard, C.J. Sialic Acid-Containing Components of Lipoproteins Influence Lipoprotein-Proteoglycan Interactions. *Atherosclerosis* **1999**, *145*, 253–260. [CrossRef]

84. Mendoza, S.; Trenchevska, O.; King, S.M.; Nelson, R.W.; Nedelkov, D.; Krauss, R.M.; Yassine, H.N. Changes in Low-Density Lipoprotein Size Phenotypes Associate with Changes in Apolipoprotein C-III Glycoforms after Dietary Interventions. *J. Clin. Lipidol.* **2017**, *11*, 224–233.e2. [CrossRef]

85. Koska, J.; Yassine, H.; Trenchevska, O.; Sinari, S.; Schwenke, D.C.; Yen, F.T.; Billheimer, D.; Nelson, R.W.; Nedelkov, D.; Reaven, P.D. Disialylated Apolipoprotein C-III Proteoform Is Associated with Improved Lipids in Prediabetes and Type 2 Diabetes. *J. Lipid Res.* **2016**, *57*, 894–905. [CrossRef]

86. Krishnan, S.; Shimoda, M.; Sacchi, R.; Kailemia, M.J.; Luxardi, G.; Kaysen, G.A.; Parikh, A.N.; Ngassam, V.N.; Johansen, K.; Chertow, G.M.; et al. HDL Glycoprotein Composition and Site-Specific Glycosylation Differentiates Between Clinical Groups and Affects IL-6 Secretion in Lipopolysaccharide-Stimulated Monocytes. *Sci. Rep.* **2017**, *7*, 43728. [CrossRef]

87. Kashyap, M.L.; Srivastava, L.S.; Hynd, B.A.; Gartside, P.S.; Perisutti, G. Quantitation of Human Apolipoprotein C-III and Its Subspecie by Radioimmunoassay and Analytical Isoelectric Focusing: Abnormal Plasma Triglyceride-Rich Lipoprotein Apolipoprotein C-III Subspecie Concentrations in Hypertriglyceridemia. *J. Lipid Res.* **1981**, *22*, 800–810. [CrossRef] [PubMed]

88. Holdsworth, G.; Stocks, J.; Dodson, P.; Galton, D.J. An Abnormal Triglyceride-Rich Lipoprotein Containing Excess Sialylated Apolipoprotein C-III. *J. Clin. Investig.* **1982**, *69*, 932–939. [CrossRef]

89. Holleboom, A.G.; Karlsson, H.; Lin, R.-S.; Beres, T.M.; Sierts, J.A.; Herman, D.S.; Stroes, E.S.G.; Aerts, J.M.; Kastelein, J.J.P.; Motazacker, M.M.; et al. Heterozygosity for a Loss-of-Function Mutation in GALNT2 Improves Plasma Triglyceride Clearance in Man. *Cell Metab.* **2011**, *14*, 811–818. [CrossRef] [PubMed]

90. Abe, S.; Katagiri, T.; Saito-Hisaminato, A.; Usami, S.; Inoue, Y.; Tsunoda, T.; Nakamura, Y. Identification of CRYM as a Candidate Responsible for Nonsyndromic Deafness, through CDNA Microarray Analysis of Human Cochlear and Vestibular Tissues**Nucleotide Sequence Data Reported Herein Are Available in the DDBJ/EMBL/GenBank Databases; for Details, See the Electronic-Database Information Section of This Article. *Am. J. Hum. Genet.* **2003**, *72*, 73–82. [CrossRef] [PubMed]

91. Wang, M.; Li, Q.; Deng, A.; Zhu, X.; Yang, J. Identification of a Novel Mutation in CRYM in a Chinese Family with Hearing Loss Using Whole-exome Sequencing. *Exp. Ther. Med.* **2020**, *20*, 1447–1454. [CrossRef]

92. Kinney, C.J.; Bloch, R.J. M-Crystallin: A Thyroid Hormone Binding Protein. *Endocr. Regul.* **2021**, *55*, 89–102. [CrossRef]

93. Oshima, A. CRYM Mutations Cause Deafness through Thyroid Hormone Binding Properties in the Fibrocytes of the Cochlea. *J. Med. Genet.* **2006**, *43*, e25. [CrossRef]

94. Carss, K.J.; Stevens, E.; Foley, A.R.; Cirak, S.; Riemersma, M.; Torelli, S.; Hoischen, A.; Willer, T.; van Scherpenzeel, M.; Moore, S.A.; et al. Mutations in GDP-Mannose Pyrophosphorylase B Cause Congenital and Limb-Girdle Muscular Dystrophies Associated with Hypoglycosylation of α-Dystroglycan. *Am. J. Hum. Genet.* **2013**, *93*, 29–41. [CrossRef] [PubMed]

95. Luo, S.; Cai, S.; Maxwell, S.; Yue, D.; Zhu, W.; Qiao, K.; Zhu, Z.; Zhou, L.; Xi, J.; Lu, J.; et al. Novel Mutations in the C-Terminal Region of GMPPB Causing Limb-Girdle Muscular Dystrophy Overlapping with Congenital Myasthenic Syndrome. *Neuromuscul. Disord.* **2017**, *27*, 557–564. [CrossRef]

96. Sun, L.; Shen, D.; Xiong, T.; Zhou, Z.; Lu, X.; Cui, F. Limb-Girdle Muscular Dystrophy Due to GMPPB Mutations: A Case Report and Comprehensive Literature Review. *Bosn J. Basic. Med. Sci.* **2020**, *2*, 275. [CrossRef]

97. Liu, Z.; Wang, Y.; Yang, F.; Yang, Q.; Mo, X.; Burstein, E.; Jia, D.; Cai, X.; Tu, Y. GMPPB-Congenital Disorders of Glycosylation Associate with Decreased Enzymatic Activity of GMPPB. *Mol. Biomed.* **2021**, *2*, 13. [CrossRef]

98. Zheng, L.; Liu, Z.; Wang, Y.; Yang, F.; Wang, J.; Huang, W.; Qin, J.; Tian, M.; Cai, X.; Liu, X.; et al. Cryo-EM Structures of Human GMPPA–GMPPB Complex Reveal How Cells Maintain GDP-Mannose Homeostasis. *Nat. Struct. Mol. Biol.* **2021**, *28*, 1–12. [CrossRef] [PubMed]

99. Yoshida-Moriguchi, T.; Yu, L.; Stalnaker, S.H.; Davis, S.; Kunz, S.; Madson, M.; Oldstone, M.B.A.; Schachter, H.; Wells, L.; Campbell, K.P. O-Mannosyl Phosphorylation of Alpha-Dystroglycan Is Required for Laminin Binding. *Science* **2010**, *327*, 88–92. [CrossRef]

100. Endo, T.; Toda, T. Glycosylation in Congenital Muscular Dystrophies. *Biol. Pharm. Bull.* **2003**, *26*, 1641–1647. [CrossRef]

101. Belaya, K.; Rodríguez Cruz, P.M.; Liu, W.W.; Maxwell, S.; McGowan, S.; Farrugia, M.E.; Petty, R.; Walls, T.J.; Sedghi, M.; Basiri, K.; et al. Mutations in *GMPPB* Cause Congenital Myasthenic Syndrome and Bridge Myasthenic Disorders with Dystroglycanopathies. *Brain* **2015**, *138*, 2493–2504. [CrossRef]

102. Bento, C. Genetic Basis of Congenital Erythrocytosis. *Int. J. Lab. Hem.* **2018**, *40*, 62–67. [CrossRef]

103. Ahmed, M.H.; Ghatge, M.S.; Safo, M.K. Hemoglobin: Structure, Function and Allostery. In *Vertebrate and Invertebrate Respiratory Proteins, Lipoproteins and other Body Fluid Proteins*; Hoeger, U., Harris, J.R., Eds.; Subcellular Biochemistry; Springer International Publishing: Cham, Switzerland, 2020; Volume 94, pp. 345–382, ISBN 978-3-030-41768-0.

104. Wajcman, H.; Galactéros, F. Hemoglobins with High Oxygen Affinity Leading to Erythrocytosis. New Variants and New Concepts. *LHEM* **2005**, *29*, 91–106. [CrossRef]

105. Hardison, R.C.; Chui, D.H.K.; Giardine, B.; Riemer, C.; Patrinos, G.P.; Anagnou, N.; Miller, W.; Wajcman, H. HbVar: A Relational Database of Human Hemoglobin Variants and Thalassemia Mutations at the Globin Gene Server. *Hum. Mutat.* **2002**, *19*, 225–233. [CrossRef] [PubMed]

106. Giardine, B.; van Baal, S.; Kaimakis, P.; Riemer, C.; Miller, W.; Samara, M.; Kollia, P.; Anagnou, N.P.; Chui, D.H.K.; Wajcman, H.; et al. HbVar Database of Human Hemoglobin Variants and Thalassemia Mutations: 2007 Update. *Hum. Mutat.* **2007**, *28*, 206. [CrossRef] [PubMed]

107. Perutz, M.F.; Rossmann, M.G.; Cullis, A.F.; Muirhead, H.; Will, G.; North, A.C.T. Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. *Nature* **1960**, *185*, 416–422. [CrossRef]

108. Perutz, M.F.; Muirhead, H.; Cox, J.M.; Goaman, L.C.G. Three-Dimensional Fourier Synthesis of Horse Oxyhaemoglobin at 2.8 Å Resolution: The Atomic Model. *Nature* **1968**, *219*, 131–139. [CrossRef] [PubMed]

109. Perutz, M.F. Stereochemistry of Cooperative Effects in Haemoglobin: Haem–Haem Interaction and the Problem of Allostery. *Nature* **1970**, *228*, 726–734. [CrossRef]

110. Kneipp, J.; Balakrishnan, G.; Chen, R.; Shen, T.-J.; Sahu, S.C.; Ho, N.T.; Giovannelli, J.L.; Simplaceanu, V.; Ho, C.; Spiro, T.G. Dynamics of Allostery in Hemoglobin: Roles of the Penultimate Tyrosine H Bonds. *J. Mol. Biol.* **2006**, *356*, 335–353. [CrossRef] [PubMed]

111. Rajko Igić, B.M.; Leray, V.; Deddish, P.A.; Erdös, E.G. Removal of Arg ^{141} From the α Chain of Human Hemoglobin by Carboxypeptidases N and M. *Circ. Res.* **1996**, *78*, 635–642. [CrossRef]

112. Tosqui, P.; Bonini-Domingos, C.R.; Colombo, M.F. Characterization and Oxygen Binding Properties of Des-Arg Human Hemoglobin. *Braz J. Med. Biol. Res.* **2009**, *42*, 494–500. [CrossRef]

113. Bare, G.H.; Bromberg, P.A.; Alben, J.O.; Brimhall, B.; Jones, R.T.; Mintz, S.; Rother, I. Altered C-Terminal Salt Bridges in Haemoglobin York Cause High Oxygen Affinity. *Nature* **1976**, *259*, 155–156. [CrossRef] [PubMed]

114. Wang, D.; Spiro, T.G. Structure Changes in Hemoglobin upon Deletion of C-Terminal Residues, Monitored by Resonance Raman Spectroscopy. *Biochemistry* **1998**, *37*, 9940–9951. [CrossRef]

115. Antonini, E.; Wyman, J.; Zito, R.; Rossi-Fanelli, A.; Caputo, A. Studies on Carboxypeptidase Digests of Human Hemoglobin. *J. Biol. Chem.* **1961**, *236*, PC60–PC63. [CrossRef] [PubMed]

116. Bettati, S.; Mozzarelli, A.; Perutz, M.F. Allosteric Mechanism of Haemoglobin: Rupture of Salt-Bridges Raises the Oxygen Affinity of the T-Structure 1 1Edited by D. Rees. *J. Mol. Biol.* **1998**, *281*, 581–585. [CrossRef] [PubMed]

117. Jones, R.T.; Shih, T.-B. Hemoglobin Variants with Altered Oxygen Affinity. *Hemoglobin* **1980**, *4*, 243–261. [CrossRef] [PubMed]

118. Schmidt, R.M.; Jue, D.L.; Ali, M.A.M.; Lyonnais, J.; Moo-Penn, W.F. Hemoglobiri $_{Bethesda}$, B145 (HC2) Tyr→His, in a Canadian Family. *Am J Clin Pathol* **1976**, *66*, 449–452. [CrossRef]

119. Adamson, J.W.; Hayashi, A.; Stamatoyannopoulos, G.; Burger, W.F. Erythrocyte Function and Marrow Regulation in Hemoglobin Bethesda (B145 Histidine). *J. Clin. Investig.* **1972**, *51*, 2883–2888. [CrossRef]

120. Bunn, H.F.; Bradley, T.B.; Davis, W.E.; Drysdale, J.W.; Burke, J.F.; Beck, W.S.; Laver, M.B. Structural and Functional Studies on Hemoglobin Bethesda (A2β2145 His), a Variant Associated with Compensatory Erythrocytosis. *J. Clin. Investig.* **1972**, *51*, 2299–2309. [CrossRef]

121. Hayashi, A.; Stamatoyannopoulos, G.; Yoshida, A.; Adamson, J. Haemoglobin Rainier: B145 (HC2) Tyrosine→Cysteine and Haemoglobin Bethesda: B145 (HC2) Tyrosine → Histidine. *Nat. New Biol.* **1971**, *230*, 264–267. [CrossRef]

122. Gacon, G.; Wajcman, H.; Labie, D.; Vigneron, C. Structural and Functional Study of Hb Nancy β 145 (HC 2) Tyr → Asp a High Oxygen Affinity Hemoglobin. *FEBS Lett.* **1975**, *56*, 39–42. [CrossRef]

123. Charache, S.; Brimhall, B.; Jones, R.T. Polycythemia Produced by Hemoglobin Osler (Beta-145 (HC2) Tyr Yields Asp). *Johns Hopkins Med. J.* **1975**, *136*, 132–136. [PubMed]

124. Hutt, P.J.; Donaldson, M.H.; Khatri, J.; Fairbanks, V.F.; Hoyer, J.D.; Thibodeau, S.N.; Moxness, M.S.; McMorrow, L.E.; Green, M.M.; Jones, R.T. Hemoglobin S/Hemoglobin Osler: A Case with 3 β Globin Chains. DNA Sequence (AAT) Proves That Hb Osler Is β 145 Tyr→Asn. *Am. J. Hematol.* **1996**, *52*, 305–309. [CrossRef]

125. Hebbel, R.P.; Eaton, J.W.; Kronenberg, R.S.; Zanjani, E.D.; Moore, L.G.; Berger, E.M. Human Llamas: Adaptation to Altitude in Subjects with High Hemoglobin Oxygen Affinity. *J. Clin. Investig.* **1978**, *62*, 593–600. [CrossRef]

126. Zak, S.J.; Brimhall, B.; Jones, R.T.; Kaplan, M.E. Hemoglobin Andrew-Minneapolis Alpha 2 A Beta 2 144 Lys Leads to Asn: A New High-Oxygen-Affinity Mutant Human Hemoglobin. *Blood* **1974**, *44*, 543–549. [CrossRef]

127. Rahbar, S.; Lee, T.D.; Davis, M.; Novotny, W.F.; Ranney, H.M. A Second Case of Hb Hanamaki [α $_2$ 139 (HC1) LYS→GLU β $_2$] in an American Family with Erythrocytosis. *Hemoglobin* **1994**, *18*, 221–226. [CrossRef] [PubMed]

128. Orisaka, M.; Tajima, T.; Harano, T.; Harano, K.; Kushida, Y.; Imai, K. A New α Chain Variant, HB Hanamaki or α $_2$ 139(HC1)LYS→GLUβ $_2$, Found in a Japanese Family. *Hemoglobin* **1992**, *16*, 67–71. [CrossRef]

129. Klotz, I.M.; Tam, J.W.O. Acetylation of Sickle Cell Hemoglobin by Aspirin. *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 1313–1315. [CrossRef] [PubMed]

130. Shamsuddin, M.; Mason, R.G.; Ritchey, J.M.; Honig, G.R.; Klotz, I.M. Sites of Acetylation of Sickle Cell Hemoglobin by Aspirin. *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 4693–4697. [CrossRef]

131. de Furia, F.G.; Cerami, A.; Bunn, H.F.; Lu, Y.S.; Peterson, C.M. The Effect of Aspirin on Sickling and Oxygen Affinity of Erythrocytes. *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 3707–3710. [CrossRef]

132. Webber, B.B.; Wilson, J.B.; Gu, L.-H.; Huisman, T.H.J. HB Ethiopia or α $_2$ 140(HC2)TYR→HISβ $_2$. *Hemoglobin* **1992**, *16*, 441–443. [CrossRef]

133. Fullerton, M.; McFarland, R.; Taylor, R.W.; Alston, C.L. The Genetic Basis of Isolated Mitochondrial Complex II Deficiency. *Mol. Genet. Metab.* **2020**, *131*, 53–65. [CrossRef]

134. Renkema, G.H.; Wortmann, S.B.; Smeets, R.J.; Venselaar, H.; Antoine, M.; Visser, G.; Ben-Omran, T.; van den Heuvel, L.P.; Timmers, H.J.L.M.; Smeitink, J.A.; et al. SDHA Mutations Causing a Multisystem Mitochondrial Disease: Novel Mutations and Genetic Overlap with Hereditary Tumors. *Eur. J. Hum. Genet* **2015**, *23*, 202–209. [CrossRef]

135. Burnichon, N.; Brière, J.-J.; Libé, R.; Vescovo, L.; Rivière, J.; Tissier, F.; Jouanno, E.; Jeunemaitre, X.; Bénit, P.; Tzagoloff, A.; et al. SDHA Is a Tumor Suppressor Gene Causing Paraganglioma. *Hum. Mol. Genet.* **2010**, *19*, 3011–3020. [CrossRef] [PubMed]

136. Sharma, P.; Maklashina, E.; Cecchini, G.; Iverson, T.M. The Roles of SDHAF2 and Dicarboxylate in Covalent Flavinylation of SDHA, the Human Complex II Flavoprotein. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 23548–23556. [CrossRef]

137. Hao, H.-X.; Khalimonchuk, O.; Schraders, M.; Dephoure, N.; Bayley, J.-P.; Kunst, H.; Devilee, P.; Cremers, C.W.R.J.; Schiffman, J.D.; Bentz, B.G.; et al. *SDH5*, a Gene Required for Flavination of Succinate Dehydrogenase, Is Mutated in Paraganglioma. *Science* **2009**, *325*, 1139–1142. [CrossRef]

138. Rutter, J.; Winge, D.R.; Schiffman, J.D. Succinate Dehydrogenase—Assembly, Regulation and Role in Human Disease. *Mitochondrion* **2010**, *10*, 393–401. [CrossRef] [PubMed]

139. Zehavi, Y.; Saada, A.; Jabaly-Habib, H.; Dessau, M.; Shaag, A.; Elpeleg, O.; Spiegel, R. A Novel de Novo Heterozygous Pathogenic Variant in the SDHA Gene Results in Childhood Onset Bilateral Optic Atrophy and Cognitive Impairment. *Metab. Brain Dis.* **2021**, *36*, 581–588. [CrossRef]

140. Kim, H.J.; Jeong, M.-Y.; Na, U.; Winge, D.R. Flavinylation and Assembly of Succinate Dehydrogenase Are Dependent on the C-Terminal Tail of the Flavoprotein Subunit. *J. Biol. Chem.* **2012**, *287*, 40670–40679. [CrossRef]

141. Jones, J.A.; Spinale, F.G.; Ikonomidis, J.S. Transforming Growth Factor-β Signaling in Thoracic Aortic Aneurysm Development: A Paradox in Pathogenesis. *J. Vasc. Res.* **2009**, *46*, 119–137. [CrossRef]

142. Goldfinger, J.Z.; Halperin, J.L.; Marin, M.L.; Stewart, A.S.; Eagle, K.A.; Fuster, V. Thoracic Aortic Aneurysm and Dissection. *J. Am. Coll. Cardiol.* **2014**, *64*, 1725–1739. [CrossRef]

143. Regalado, E.S.; Guo, D.; Villamizar, C.; Avidan, N.; Gilchrist, D.; McGillivray, B.; Clarke, L.; Bernier, F.; Santos-Cortez, R.L.; Leal, S.M.; et al. Exome Sequencing Identifies *SMAD3* Mutations as a Cause of Familial Thoracic Aortic Aneurysm and Dissection With Intracranial and Other Arterial Aneurysms. *Circ. Res.* **2011**, *109*, 680–686. [CrossRef] [PubMed]

144. van der Pluijm, I.; van Vliet, N.; von der Thusen, J.H.; Robertus, J.L.; Ridwan, Y.; van Heijningen, P.M.; van Thiel, B.S.; Vermeij, M.; Hoeks, S.E.; Buijs-Offerman, R.M.G.B.; et al. Defective Connective Tissue Remodeling in Smad3 Mice Leads to Accelerated Aneurysmal Growth Through Disturbed Downstream TGF-β Signaling. *EBioMedicine* **2016**, *12*, 280–294. [CrossRef] [PubMed]

145. Schepers, D.; Tortora, G.; Morisaki, H.; MacCarrick, G.; Lindsay, M.; Liang, D.; Mehta, S.G.; Hague, J.; Verhagen, J.; van de Laar, I.; et al. A Mutation Update on the LDS-Associated Genes *TGFB2/3* and *SMAD2/3*. *Hum. Mutat.* **2018**, *39*, 621–634. [CrossRef] [PubMed]

146. Hostetler, E.M.; Regalado, E.S.; Guo, D.-C.; Hanna, N.; Arnaud, P.; Muiño-Mosquera, L.; Callewaert, B.L.; Lee, K.; Leal, S.M.; Wallace, S.E.; et al. *SMAD3* Pathogenic Variants: Risk for Thoracic Aortic Disease and Associated Complications from the Montalcino Aortic Consortium. *J. Med. Genet* **2019**, *56*, 252–260. [CrossRef]

147. Brown, K.A.; Pietenpol, J.A.; Moses, H.L. A Tale of Two Proteins: Differential Roles and Regulation of Smad2 and Smad3 in TGF-β Signaling. *J. Cell. Biochem.* **2007**, *101*, 9–33. [CrossRef]

148. Flowers, S.A.; Rebeck, G.W. APOE in the Normal Brain. *Neurobiol. Dis.* **2020**, *136*, 104724. [CrossRef]

149. Kim, J.; Basak, J.M.; Holtzman, D.M. The Role of Apolipoprotein E in Alzheimer's Disease. *Neuron* **2009**, *63*, 287–303. [CrossRef] [PubMed]

150. van den Maagdenberg, A.M.; Weng, W.; de Bruijn, I.H.; de Knijff, P.; Funke, H.; Smelt, A.H.; Gevers Leuven, J.A.; van't Hooft, F.M.; Assmann, G.; Hofker, M.H. Characterization of Five New Mutants in the Carboxyl-Terminal Domain of Human Apolipoprotein E: No Cosegregation with Severe Hyperlipidemia. *Am. J. Hum. Genet* **1993**, *52*, 937–946.

151. Halim, A.; Rüetschi, U.; Larson, G.; Nilsson, J. LC–MS/MS Characterization of O-Glycosylation Sites and Glycan Structures of Human Cerebrospinal Fluid Glycoproteins. *J. Proteome Res.* **2013**, *12*, 573–584. [CrossRef]

152. Raftery, M.; Campbell, R.; Glaros, E.N.; Rye, K.-A.; Halliday, G.M.; Jessup, W.; Garner, B. Phosphorylation of Apolipoprotein-E at an Atypical Protein Kinase CK2 PSD/E Site in Vitro. *Biochemistry* **2005**, *44*, 7346–7353. [CrossRef]

153. Hu, Y.; Meuret, C.; Go, S.; Yassine, H.N.; Nedelkov, D. Simple and Fast Assay for Apolipoprotein E Phenotyping and Glycotyping: Discovering Isoform-Specific Glycosylation in Plasma and Cerebrospinal Fluid. *JAD* **2020**, *76*, 883–893. [CrossRef] [PubMed]

154. Chua, C.-C.; Lim, M.-L.; Wong, B.-S. Altered Apolipoprotein E Glycosylation Is Associated with Aβ(42) Accumulation in an Animal Model of Niemann-Pick Type C Disease. *J. Neurochem.* **2010**, *112*, 1619–1626. [CrossRef] [PubMed]

155. Sugano, M.; Yamauchi, K.; Kawasaki, K.; Tozuka, M.; Fujita, K.; Okumura, N.; Ota, H. Sialic Acid Moiety of Apolipoprotein E3 at Thr194 Affects Its Interaction with β-Amyloid1–42 Peptides. *Clin. Chim. Acta* **2008**, *388*, 123–129. [CrossRef] [PubMed]

156. Ke, L.; Chan, H.; Chen, C.; Chang, C.; Lu, P.; Chu, C.; Lai, W.; Shin, S.; Liu, F.; Chen, C. Increased APOE Glycosylation Plays a Key Role in the Atherogenicity of L5 Low-density Lipoprotein. *FASEB J.* **2020**, *34*, 9802–9813. [CrossRef]

157. Wernette-Hammond, M.E.; Lauer, S.J.; Corsini, A.; Walker, D.; Taylor, J.M.; Rall, S.C. Glycosylation of Human Apolipoprotein E. The Carbohydrate Attachment Site Is Threonine 194. *J. Biol. Chem.* **1989**, *264*, 9094–9101. [CrossRef]

158. Zanni, E.E.; Kouvatsi, A.; Hadzopoulou-Cladaras, M.; Krieger, M.; Zannis, V.I. Expression of ApoE Gene in Chinese Hamster Cells with a Reversible Defect in O-Glycosylation. Glycosylation Is Not Required for ApoE Secretion. *J. Biol. Chem.* **1989**, *264*, 9137–9140. [CrossRef]

159. Edghill, E.L.; Flanagan, S.E.; Patch, A.-M.; Boustred, C.; Parrish, A.; Shields, B.; Shepherd, M.H.; Hussain, K.; Kapoor, R.R.; Malecki, M.; et al. Insulin Mutation Screening in 1044 Patients With Diabetes. *Diabetes* **2008**, *57*, 1034–1042. [CrossRef]

160. Adhikari, S.; Nice, E.C.; Deutsch, E.W.; Lane, L.; Omenn, G.S.; Pennington, S.R.; Paik, Y.-K.; Overall, C.M.; Corrales, F.J.; Cristea, I.M.; et al. A High-Stringency Blueprint of the Human Proteome. *Nat. Commun.* **2020**, *11*, 5301. [CrossRef]

161. Morgan, A.A.; Rubenstein, E. Proline: The Distribution, Frequency, Positioning, and Common Functional Roles of Proline and Polyproline Sequences in the Human Proteome. *PLoS ONE* **2013**, *8*, e53785. [CrossRef]

162. Davydov, E.V.; Goode, D.L.; Sirota, M.; Cooper, G.M.; Sidow, A.; Batzoglou, S. Identifying a High Fraction of the Human Genome to Be under Selective Constraint Using GERP++. *PLoS Comput. Biol.* **2010**, *6*, e1001025. [CrossRef]

163. Ng, P.C. SIFT: Predicting Amino Acid Changes That Affect Protein Function. *Nucleic Acids Res.* **2003**, *31*, 3812–3814. [CrossRef] [PubMed]

164. Kumar, P.; Henikoff, S.; Ng, P.C. Predicting the Effects of Coding Non-Synonymous Variants on Protein Function Using the SIFT Algorithm. *Nat Protoc* **2009**, *4*, 1073–1081. [CrossRef] [PubMed]

165. Vaser, R.; Adusumalli, S.; Leng, S.N.; Sikic, M.; Ng, P.C. SIFT Missense Predictions for Genomes. *Nat. Protoc.* **2016**, *11*, 1–9. [CrossRef]

166. Adzhubei, I.; Jordan, D.M.; Sunyaev, S.R. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *CP Hum. Genet.* **2013**, *76*, 7–20. [CrossRef]

167. Samocha, K.E.; Kosmicki, J.A.; Karczewski, K.J.; O'Donnell-Luria, A.H.; Pierce-Hoffman, E.; MacArthur, D.G.; Neale, B.M.; Daly, M.J. Regional Missense Constraint Improves Variant Deleteriousness Prediction. *BioRxiv* **2017**, 148353.

168. Rentzsch, P.; Witten, D.; Cooper, G.M.; Shendure, J.; Kircher, M. CADD: Predicting the Deleteriousness of Variants throughout the Human Genome. *Nucleic Acids Res.* **2019**, *47*, D886–D894. [CrossRef]

169. Kircher, M.; Witten, D.M.; Jain, P.; O'Roak, B.J.; Cooper, G.M.; Shendure, J. A General Framework for Estimating the Relative Pathogenicity of Human Genetic Variants. *Nat. Genet.* **2014**, *46*, 310–315. [CrossRef]