



Article

# Rare Human Missense Variants can affect the Function of Disease-Relevant Proteins by Loss and Gain of Peroxisomal Targeting Motifs

Cheng-Shoong Chong <sup>1,2,†</sup>, Markus Kunze <sup>3,†</sup>, Bernhard Hochreiter <sup>4</sup>, Martin Krenn <sup>5,6</sup>, Johannes Berger <sup>3</sup> and Sebastian Maurer-Stroh <sup>1,2,7,8\*</sup>

<sup>1</sup> Bioinformatics Institute, Agency for Science, Technology and Research (A\*STAR), Singapore 138671, Singapore; e0161111@u.nus.edu

<sup>2</sup> National University of Singapore Graduate School for Integrative Sciences and Engineering (NGS), National University of Singapore, Singapore 119077, Singapore

<sup>3</sup> Medical University of Vienna, Center for Brain Research, Department of Pathobiology of the Nervous System, 1090 Vienna, Austria; markus.kunze@meduniwien.ac.at (M.K.); johannes.berger@meduniwien.ac.at (J.B.)

<sup>4</sup> Medical University of Vienna, Center for Physiology and Pharmacology, Institute for Vascular Biology and Thrombosis Research, 1090 Vienna, Austria; bernhard.hochreiter@meduniwien.ac.at

<sup>5</sup> Department of Neurology, Medical University of Vienna, 1090 Vienna, Austria; martin.krenn@meduniwien.ac.at

<sup>6</sup> Institute of Human Genetics, Technical University Munich, 81675 Munich, Germany

<sup>7</sup> Department of Biological Sciences, National University of Singapore, Singapore 117558, Singapore

<sup>8</sup> Innovations in Food and Chemical Safety Programme (IFCS), Agency for Science, Technology and Research (A\*STAR), Singapore 138671, Singapore

\* Correspondence: sebastianms@bii.a-star.edu.sg

† These authors contributed equally to this work.

Received: 23 July 2019; Accepted: 14 September 2019; Published: 17 September 2019

**Abstract:** Single nucleotide variants (SNVs) resulting in amino acid substitutions (i.e., missense variants) can affect protein localization by changing or creating new targeting signals. Here, we studied the potential of naturally occurring SNVs from the Genome Aggregation Database (gnomAD) to result in the loss of an existing peroxisomal targeting signal 1 (PTS1) or gain of a novel PTS1 leading to mistargeting of cytosolic proteins to peroxisomes. Filtering down from 32,985 SNVs resulting in missense mutations within the C-terminal tripeptide of 23,064 human proteins, based on gene annotation data and computational prediction, we selected six SNVs for experimental testing of loss of function (LoF) of the PTS1 motif and five SNVs in cytosolic proteins for gain in PTS1-mediated peroxisome import (GoF). Experimental verification by immunofluorescence microscopy for subcellular localization and FRET affinity measurements for interaction with the receptor PEX5 demonstrated that five of the six predicted LoF SNVs resulted in loss of the PTS1 motif while three of five predicted GoF SNVs resulted in *de novo* PTS1 generation. Overall, we showed that a complementary approach incorporating bioinformatics methods and experimental testing was successful in identifying SNVs capable of altering peroxisome protein import, which may have implications in human disease.

**Keywords:** peroxisomal targeting signal 1; PEX5; peroxisome; protein transport; missense variant; mistargeting; gnomAD; disease

## 1. Introduction

Inter-individual DNA sequence variations in the human genome provoke diverse consequences ranging from physical trait differences to altered disease susceptibility and even drug therapy failure

[1–3]. Non-synonymous single nucleotide variants (SNVs) resulting in amino acid substitutions or missense variants have been implicated in many human genetic diseases [4]. To date, more than 100 million validated SNVs are listed in the Single Nucleotide Polymorphism Database (dbSNP build 151) [5], and with the advent of inexpensive next-generation sequencing technology, many more novel and rare SNVs are being discovered [6].

While the ability of missense variants to cause diseases is clear, their effects on the protein can be complex and manifest in several ways. At the protein level, they have been found to cause protein instability, change protein flexibility, abrogate protein–macromolecular and protein–chemical interactions, reduce enzyme activity, and even modify protein function [7–9]. Importantly, such SNVs can also lead to aberrant protein localization in the cell by inactivating targeting signals [10,11], although such effects have hardly been investigated since most targeting signals initiating these translocation processes are rather robust against individual mutations.

In this study, we focused on SNVs leading to the generation or destruction of peroxisomal targeting signal type-1 (PTS1) and their potential impact on sorting of proteins into the peroxisomes. Peroxisomes are single membrane bound organelles housing a variety of metabolic reactions that degrade various fatty acids through  $\beta$ -oxidation, break down hydrogen peroxide and other reactive oxygen species (ROS), and catabolize D-amino acids [12]. In humans, peroxisomes also contribute to the biosynthesis of ether-phospholipids, including plasmalogens or bile acids [13]. The importance of peroxisomes for human physiology is highlighted in many inherited human diseases caused by a complete dysfunction of peroxisomes (Zellweger syndrome spectrum) [14,15] or by the lack of an individual enzyme or transporter protein [16].

Soluble proteins destined for the peroxisomal lumen are equipped with a peroxisomal targeting signal which resides either at the extreme C-terminus (PTS1) or close to the N-terminus (PTS2) [13]. PTS1 motifs have been characterized as a C-terminal tripeptide Ser-Lys-Leu (SKL) or conserved variants thereof ([S/A/C]–[K/R/H]–L<sup>COO</sup>–) [17,18], although later, a much broader number of tripeptides were found to act as targeting signals. Moreover, the sequence preceding the C-terminal tripeptide, the upstream sequence, modulates the binding strength to the PTS1 receptor PEX5 [19,20]. Mistargeting of proteins has been linked to diseases in the context of arbitrary generation of alternative targeting signals for peroxisomal proteins [21–23]. Furthermore, cytosolic proteins should be sensitive to mislocalization because the *de novo* generation of a targeting signal should induce transport, and thus, it should deplete the protein from the cytosol. In this respect, the PTS1 signal is a good candidate for study as it is (i) exposed to the extreme C-terminal end, (ii) reliably predictable [24], and (iii) can be modulated by individual point mutations, whereas other targeting signals are rather tolerant against mutations, but also hard to induce.

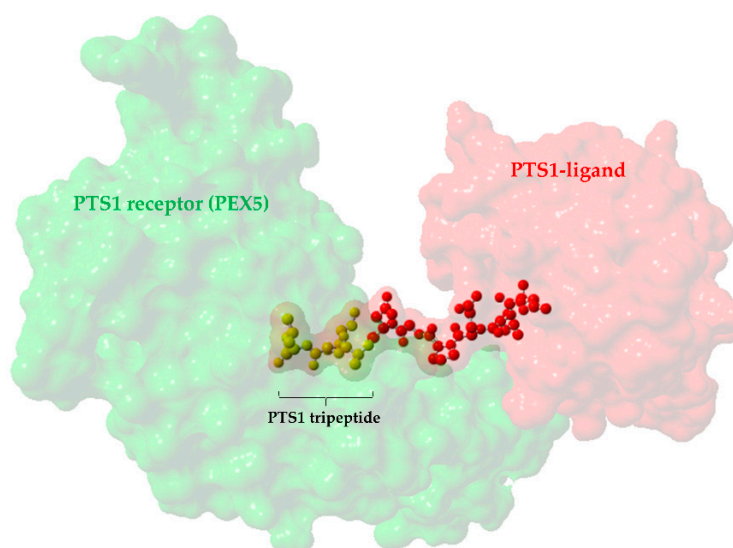
However, much is still unknown about the potential of SNVs in affecting protein import into peroxisomes by gain or loss of PTS1, although computational algorithms and tools [25,26] including integrative methods [27] have been developed to aid in delineating between SNVs with and without functional consequences. Using a combination of such tools and along with experimental validation of PTS1 signal quality, we performed a systematic analysis of SNVs affecting protein transport into peroxisomes after mining the entire Genome Aggregation Database (gnomAD) [28] for relevant missense variants occurring in the extreme C-termini of all human proteins. At the first level, we identified loss of function variants (LoF) inactivating PTS1 in known, disease-relevant peroxisomal proteins. These LoF mutants are expected to ablate the proper localization of these enzymes, and thereby interrupting peroxisomal metabolism similar to a loss of the enzymatic activity. In a second step, we searched for SNVs introducing a PTS1 motif in a cytosolic protein, which was expected to induce peroxisomal import of the protein, thereby functionally depleting the cytosol of this protein. To the best of our knowledge, these are the first descriptions of the ability of missense variants occurring naturally in the human population to abolish or generate a PTS1 signal and consequently alter the localization of the affected protein into or out of the peroxisome. Our findings would be important in the context of analyzing the effects of variant-induced aberrant localization on protein function, and reinforces the need to evaluate targeting signal changes when determining the disease-relevance of a protein mutation amongst other factors.

## 2. Results

### 2.1. Mining of gnomAD for SNVs Causing Missense Mutations in C-terminal Tripeptides

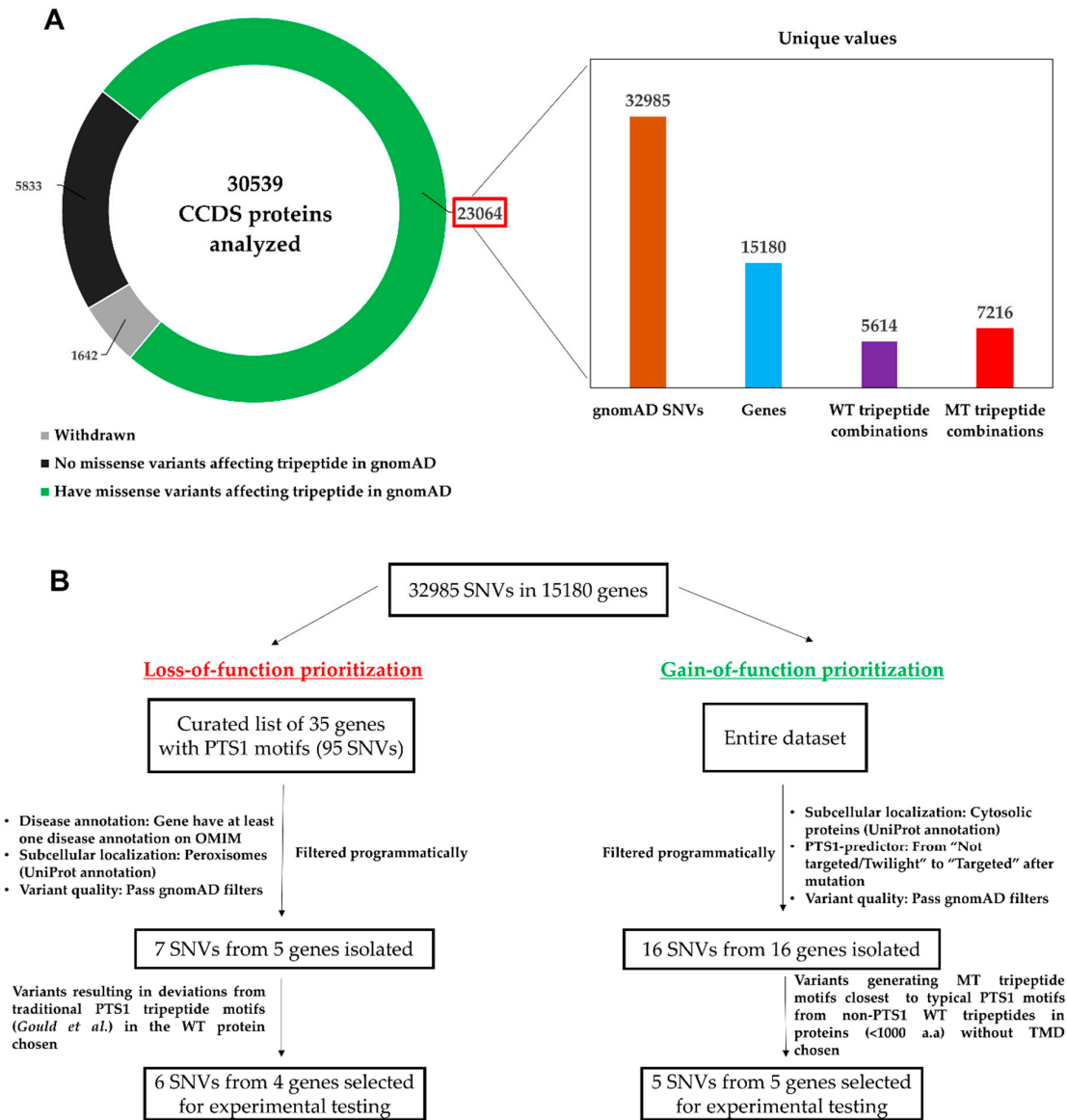
From a core set of manually curated protein coding regions from the Consensus Coding Sequence project (CCDS), we obtained protein and DNA sequence information along with protein coding genomic coordinates for 30,539 proteins and their isoforms. After exclusion of 1642 CCDS-withdrawn proteins, we applied a chromosome and position-based query of 125,748 whole exome and 15,708 whole genome sequencing data from gnomAD (see Method 4.1) to filter out SNVs located at the last three codons preceding the stop codon of the gene transcripts encoding the remaining proteins. gnomAD is a sequence variation database containing harmonized variant data from more than 140,000 human samples collected from a broad range of studies (full list available at <https://gnomad.broadinstitute.org/about>). We chose this database for our analysis as it is currently the largest publicly available human sequence variation database and the data is also easily accessible.

Additionally, we chose to focus on SNVs lying in the last three codons of each transcript as they can potentially generate or abolish a PTS1 mediating the interaction with the PTS1 receptor (PEX5) by mutating the C-terminal tripeptide of the analyzed proteins. The importance of this tripeptide to PTS1-mediated peroxisome targeting has been extensively studied compared to its upstream sequence, and hence, it is well suited to predict [25] and proof functional changes in the PTS1. Figure 1 depicts the location of the tripeptide motif in PTS1 in a previously crystallized complex (pdb ID: 2c0l) that consists of the TPR region of PEX5 and its PTS1-containing ligand (human SCP2).



**Figure 1.** Three-dimensional crystal structure of a PTS1 receptor-ligand complex from Reference [29] (pdb ID = 2c0l), illustrated using Yasara. The PTS1 tripeptide motif (yellow) is located at the extreme C-terminus of the protein (pale red) and is extended and bound to the tetratricopeptide repeat (TPR) region of the receptor (pale green) (C-terminus extension from ligand is highlighted in opaque colors of red and yellow).

In this region, we found that 23,064 protein encoding transcripts covering 15,180 genes contained SNVs (Figure 2A). Given that not all SNVs cause non-synonymous amino acid changes, we only retained missense variants, and in total, we attained 32,985 bi- or multi-allelic SNVs capable of mutating one of the last three amino acids of these proteins. Considering the entire gene ( $n = 18,592$ ) and protein set analyzed ( $n = 28,897$ ), approximately 80% of the proteins and 81.6% of genes had at least one missense variant from at least one human sample in gnomAD capable of mutating the C-terminal tripeptides under study.



**Figure 2.** Overall bioinformatics workflow for the mining of sequence variation data from gnomAD and the selection of SNVs for experimental testing. **(A)** 32,985 bi- or multi-allelic gnomAD SNVs capable of altering at least one amino acid in the tripeptide motifs of 23,064 proteins and generating over 7000 unique mutant (MT) tripeptide combinations from 5614 unique wild-type (WT) tripeptides attained **(B)** Prioritization of variants for testing from the SNVs retrieved from gnomAD. TMD, Transmembrane domain; a.a, amino acid.

## 2.2. Investigation of SNVs Resulting in PTS1 Ablation (Loss-of-Function)

### 2.2.1. Identification and Prioritization of SNVs Resulting in Loss-of-Function PTS1 for Experimental Testing

From curation of the peroxisomal databases [30] and our literature survey, we identified 35 genes encoding PTS1-containing human peroxisomal proteins (Figure 2B, Table S1). To focus on studying potential mistargeting in disease-relevant genes, we narrowed the list to six genes that had been broadly linked to at least one human disease in the Online Mendelian Inheritance in Man (OMIM) database. These gene products must also be localized solely in the peroxisome based on UniProt subcellular localization annotations (see Sections 4.3, 4.5). After filtering for missense

variants that passed gnomAD associated quality filters and that affected proteins with a traditional PTS1 consensus motif [18], six SNVs in four genes remained to be tested for motif inactivation (Table 1, Figure 2B).

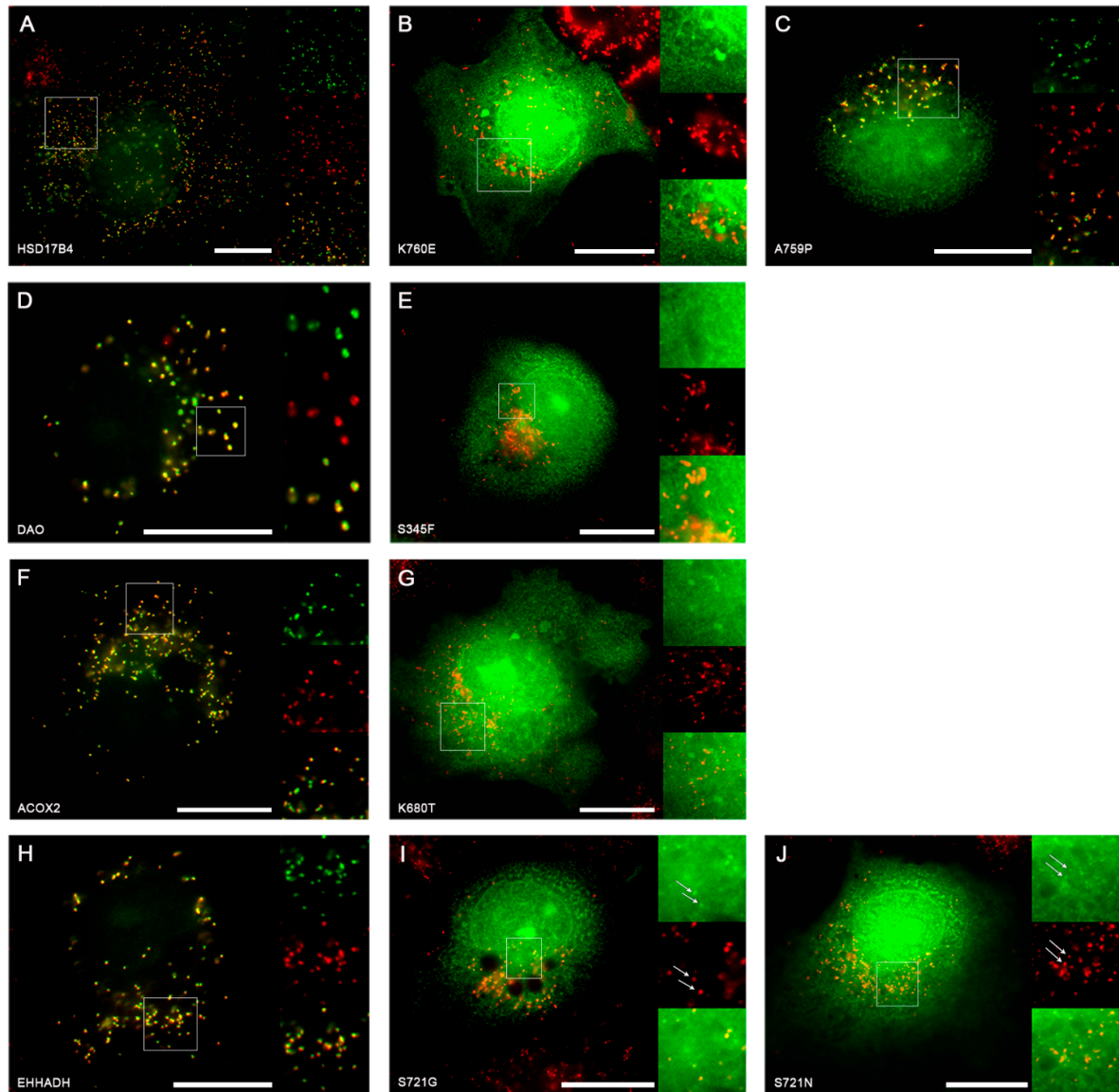
**Table 1.** SNVs selected for experimental testing for the ablation of PTS1 motif.

rsID. (Nucleotide Change)	Gene	CCDS ID	PTS1 Motif		UniProt Subcellular Localization *
			WT	MT	
rs1049954328 (A>G)	<i>HSD17B4</i> **	CCDS56379.1	AKL	AEL	Peroxisome
rs751064948 (G>C)	<i>HSD17B4</i> **	CCDS56379.1	AKL	PKL	Peroxisome
rs143732132 (C>T)	<i>DAO</i>	CCDS9122.1	SHL	FHL	Peroxisome
rs1182979757 (T>G)	<i>ACOX2</i>	CCDS33775.1	SKL	STL	Peroxisome
rs138945273 (T>C)	<i>EHHADH</i>	CCDS33901.1	SKL	GKL	Peroxisome
rs968361154 (C>T)	<i>EHHADH</i>	CCDS33901.1	SKL	NKL	Peroxisome

\* Native protein, \*\* Longest CCDS transcript chosen for analysis. rsID, Reference single nucleotide polymorphism cluster identification; CCDS ID, Consensus Coding Sequence Project accession identification; WT, Wild-type tripeptide; MT, Mutated tripeptide.

## 2.2.2. Experimental Investigation of Functional Consequences of SNVs in PTS1 Motifs

Next, we experimentally verified the functional consequences of the selected SNVs. These SNVs changed the C-terminal tripeptide from typical PTS1 motifs to derivatives thereof, which were not covered by the original suggested PTS1 consensus sequence of the C-terminal tripeptide [18]. To verify the effect of these point mutations on the quality of the PTS1, we first generated EGFP fusion proteins extended by the last 12 amino acids of the native enzymes or with variants thereof harboring the mutated C-terminal tripeptide. When these proteins were expressed in COS7 cells, we found that all EGFP-proteins extended by the PTS1 motifs of the native peroxisomal proteins (*HSD17B4*, *DAO*, *ACOX2*, and *EHHADH*) were found in punctate structures colocalizing with the peroxisomal marker PMP70, which confirmed the functionality of these PTS1 motifs (Figures 3A,D,F,H). In contrast, fusion proteins encoding the same PTS1, but harboring the point mutations *HSD17B4*-K760E (Figure 3B), *DAO*-S345F (Figure 3E), and *ACOX2*-K680T (Figure 3G) appeared exclusively cytosolic indicating an inactivation of the PTS1, whereas *HSD17B4*-A759P (Figure 3C) was clearly punctate suggesting that the functionality of the PTS1 was hardly affected. In contrast, the variant *EHHADH*-S721G was predominantly cytosolic (Figure 3I) and the variant *EHHADH*-S721N was nearly exclusively cytosolic with a faint peroxisomal background in a fraction of cells (Figure 3J).



**Figure 3.** Study of PTS1 loss of function mutations by SNV: (A–J) COS7 cells were transfected with expression plasmids for different EGFP-PTS1 variants, encoding the C-terminus of multifunctional protein (HSD17B4) (A) or the variants K760E (B) and A759P (C), of D-amino acid oxidase (DAO) (D), or the variant S345F (E), of acyl-CoA oxidase 2 (ACOX2) (F), or the variant K680T (G), and of EHHADH (H) or the variants S721G (I) or S721N (J). The subcellular localization was determined by fluorescence microscopy (EGFP, green) in combination with immunofluorescence microscopy of the peroxisomal marker PMP70 (red). I,J: white arrows indicate co-localization between PMP70 and a small fraction of EGFP. Scale bars indicate 20  $\mu$ m and white squares define enlarged areas.

### 2.3. Investigation of SNVs Resulting *de novo* PTS1 Generation in Cytosolic Proteins (Gain-of-Function)

#### 2.3.1. Identification and Prioritization of Gain-of-Function SNVs Generating *de novo* PTS1 for Experimental Testing

From the entire collection of 32,985 SNVs described above, we concurrently applied three filters that were based on the annotation of subcellular localization data, on variant quality, and on the evaluation by the PTS1 predictor [25], to search for SNVs with the potential to generate a *de novo* PTS1 at the C-terminus of a cytosolic protein (see Methods 4.5, Figure 2B). The strict filters identified 16

SNVs in 22 CCDS transcripts of 16 genes (Table S2). Among these, we chose five SNVs which generated PTS1-like tripeptides (MT) from non-PTS1 WT tripeptides at the C-termini of reasonably long proteins (<1000 a.a) without transmembrane domains. (Figure 2B, Table 2).

**Table 2.** SNVs selected for experimental testing of *de novo* generation of PTS1 motif.

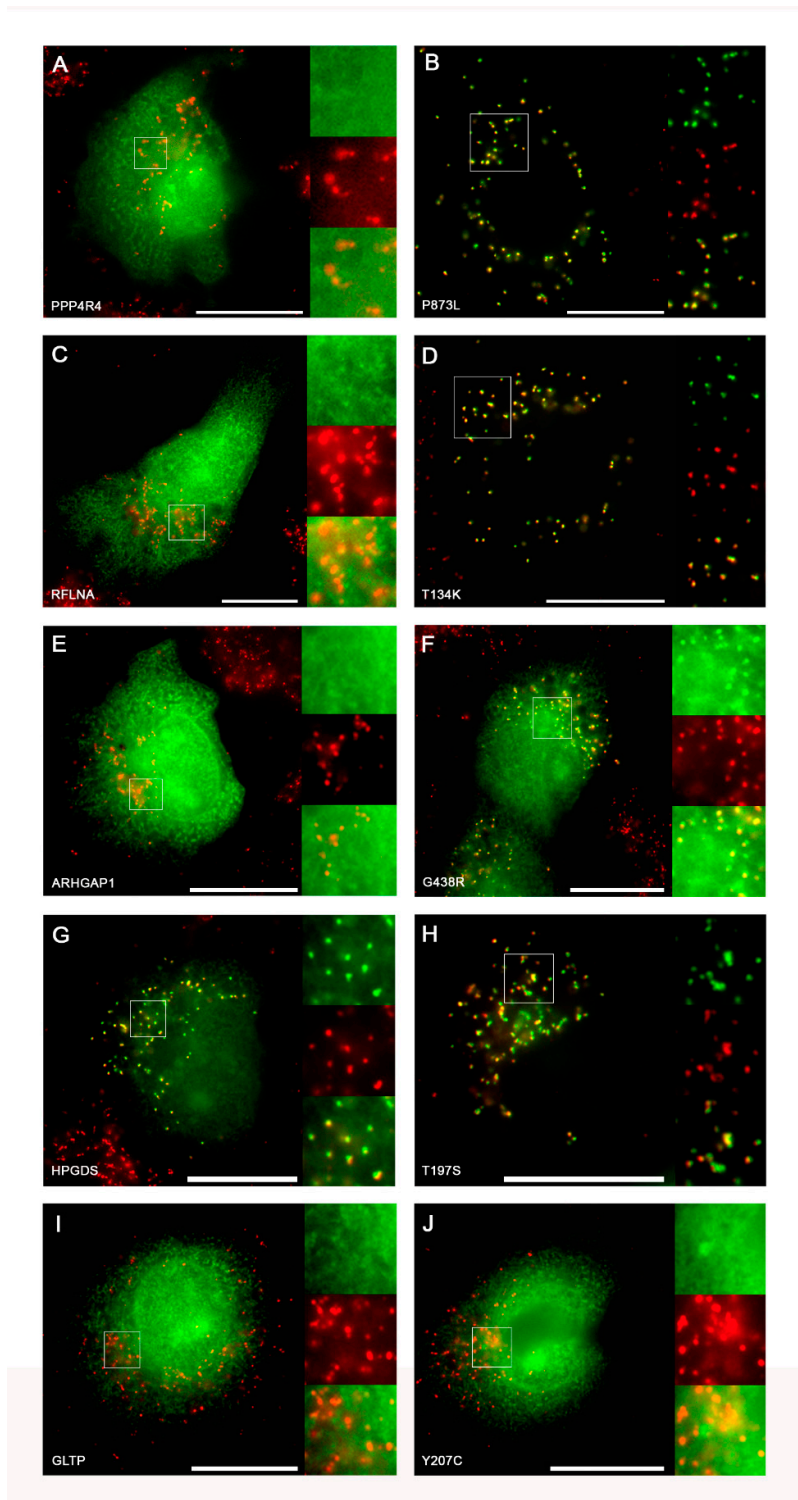
rsID (Nucleotide Change)	Gene	CCDS ID	Motif		PTS1 Prediction (WT)	PTS1 Prediction (MT)	UniProt Subcellular Localization *
			WT	MT			
rs576288488 (C>T)	PPP4R4	CCDS9921.1	SNP	SNL	Not targeted	Targeted	Cytoplasm
rs760206157 (C>A)	RFLNA	CCDS9258.1	ATL	AKL	Not targeted	Targeted	Cytoplasm, cytoskeleton
rs747965330 (C>T)	ARHGAP1	CCDS7922.1	SGL	SRL	Not targeted	Targeted	Cytoplasm
rs765481546 (G>C)	HPGDS	CCDS3640.1	TKL	SKL	Not targeted	Targeted	Cytoplasm
rs773190605 (T>C)	GLTP	CCDS9136.1	YKV	CKV	Twilight zone	Targeted	Cytoplasm

\* native protein. rsID, Reference single nucleotide polymorphism cluster identification; CCDS ID, Consensus Coding Sequence Project accession identification; WT, Wild-type tripeptide; MT, Mutated tripeptide.

### 2.3.2. Experimental Validation of Gain-of-function SNVs Leading to *de novo* Generation of PTS1

Similarly, we generated expression plasmids for EGFP extended either by the C-terminus of the native proteins or by the mutated form. When the EGFP fusion proteins encoding the native C-terminus of these proteins (PPP4R4, RFLNA, ARHGAP1, HPGDS, and GLTP) were expressed in COS7 cells and the subcellular localization of the reporter protein was determined by fluorescence microscopy, we found that EGFP was evenly distributed across the cytosol for all (Figures 4A,C,E,I), except for HPGDS. This result showed punctate staining in front of a strong background, and these punctate structures colocalized with PMP70 (Figure 4G) suggesting that four out of five of these C-terminal ends were non-functional, whereas the C-terminus of HPGDS appeared to act as weak PTS1. However, when the variants of the C-terminal sequences encoding the point mutation were expressed, we found that the fusion proteins encoding the C-terminus of the mutants PPP4R4-P873L (Figure 4B), RFLNA-T134K (Figure 4D), and HPGDS-T197S (Figure 4H) showed a clear punctate staining indicating the functionality of the modified C-terminus as PTS1. Moreover, the EGFP-fusion protein with the C-terminus of ARHGAP1-G438R presented with a weak peroxisomal staining (Figure 4F), and that of GLTP-Y207C appeared completely cytosolic (Figure 4J). These results confirmed the ability of SNVs to generate novel functional PTS1 variants *de novo*. Moreover, we found that in one case the native C-terminus had a weak ability to mediate peroxisomal import (HPGDS), whereas, in another, the mutation did not generate a functional PTS1 (GLTP). However, in this case, the mutation increased the numerical PTS1-score only slightly, although this was sufficient to cross a threshold and change the qualitative prediction from “Twilight zone” to “Targeted.” The inability of this GLTP mutant to act as a PTS1 might be due to the valine being the last residue of the tripeptide as this amino acid has been shown to weaken known PTS1 motifs [31].





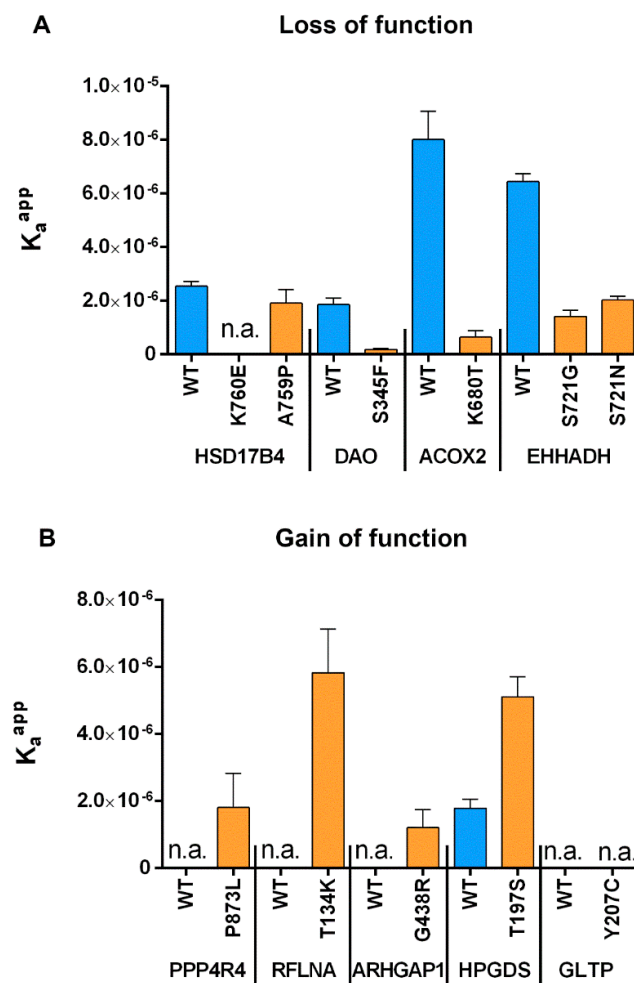
**Figure 4.** Study of PTS1 gain of function mutations by SNV: (A–J) COS7 cells were transfected with expression plasmids for different EGFP-PTS1 variants, encoding the C-terminus of PPP4R4 (A) or the variant P873L (B), of RFLNA (C) or the variant T134K (D), ARHGAP1 (E) or the variant G438R (F), of HPGDS (G) or the variant T197S (H), and of GLTP (I) or the variant Y207C (J). The subcellular localization was determined by fluorescence microscopy (EGFP, green) in combination with immunofluorescence microscopy of the peroxisomal marker PMP70 (red). Scale bars indicate 20  $\mu$ m and white squares define enlarged areas.



#### 2.4. Measurement of Relative Affinity Before and After Point Mutations for both GoF and LoF Mutants

To retrace the effect of SNVs in the last three amino acids more closely to changes in the binding strength to the PTS1-receptor PEX5, we took advantage of a recently developed novel approach [32] to determine the relative binding affinity of PEX5 to diverse PTS1 motifs. Subjecting the abovementioned peptides to this FRET-based measurement of protein–protein interactions in living cells, we were able to determine the change in the affinity to PEX5 caused by individual point mutations for each PTS1 (Figure 5).

We found that for all point mutations reflecting SNVs that inactivated the PTS1, the affinity to PEX5 was also drastically reduced (HSD17B4-K760E, DAO-S345F, ACOX2-K680T, EHHADH-S721N) just as in the case of slight residual import (EHHADH-S721G) (Figure 5A). In contrast, hardly any change was observed for the SNV that did not ablate peroxisomal import (HSD17B4-A759P). This outcome confirmed the results of our previous experiments using an orthogonal method and showed that the functionality of the PTS1 correlates with its affinity to PEX5. Similarly, all SNVs that converted a C-terminal end into a functional PTS1 also increased the affinity of the peptide to PEX5 (PPP4R4-P873L, RFLNA-T134K, ARHGAP1-G438R) although to different extents, whereas the mutation that did not act as PTS1 did not interact with PEX5 (GLTP-Y207C) (Figure 5B). Finally, for HPGDS, the C-terminal sequence in the native form not only enabled peroxisomal import but it mediated interaction with PEX5, although the SNV (HPGDS-T197S) presented with an even higher affinity.



**Figure 5.** FlowFRET analysis based measurement of affinity between the PTS1 receptor (PEX5) and diverse EGFP-PTS1 variants harboring either the native C-termini of peroxisomal proteins (blue) or

LoF mutants thereof (orange) (A), or the C-termini of cytosolic proteins (blue) and variants (orange) thereof harboring SNVs (B);  $K_a^{app}$ : apparent interaction strength as a correlative measure of affinity obtained by fitting; blue: native C-terminus; orange: C-terminus harboring the SNV; n.a: not analyzed (the low affinity of the interaction partners does not allow fitting).

## 2.5. Assessment of PTS1 Predictor and FoldX Predictions of PTS1 Signal Quality

Given that we used computational predictions such as the PTS1-predictor to identify SNVs that could affect peroxisomal targeting and experimentally determined the effect of the SNVs, we could use the experimental data generated in this study as an independent validation of the performance of such tools for future applications. Here, we assessed the overall performance of the PTS1-predictor and another predictor that we had used previously (structural binding prediction using FoldX) [33] by comparing the predictions with the experimental results (immunofluorescence microscopic analysis (IFA) and FRET affinity measurements) (Table 3).

**Table 3.** Experimental measurements of apparent affinity strength ( $K_a^{app}$ ) of the WT and MT PTS1 motif along with predicted values from FoldX ( $ddG_{bind}$ , kcal/mol) and the PTS1 predictor (score difference, MT-WT) upon motif change.

Gene	Motif		WT $K_a^{app}$ ( $10^{-6}$ )	MT $K_a^{app}$ ( $10^{-6}$ )	log $K_a^{app}$ Ratio	FoldX $ddG_{bind}$ (kcal/mol)	PTS1 Predictor Score Difference (MT-WT)
	WT	MT					
<i>HSD17B4</i>	AKL	AEL	2.544	0.065	1.592	5.138	−37.8
<i>HSD17B4</i>	AKL	PKL	2.544	1.915	0.123	0.154	−4.4
<i>DAO</i>	SHL	FHL	1.859	0.186	1.000	7.910	−18.9
<i>ACOX2</i>	SKL	STL	8.015	0.646	1.094	3.800	−25.0
<i>EHHADH</i>	SKL	GKL	6.444	1.424	0.655	0.920	−8.5
<i>EHHADH</i>	SKL	NKL	6.444	2.027	0.502	1.687	−5.8
<i>PPP4R4</i>	SNP	SNL	0.065	1.808	−1.444	0.359	33.9
<i>RFLNA</i>	ATL	AKL	0.059	5.826	−1.995	−0.723	31.1
<i>ARHGAP1</i>	SGL	SRL	0.047	1.209	−1.410	−2.432	28.5
<i>HPGDS</i>	TKL	SKL	1.792	5.110	−0.455	1.397	22.2
<i>GLTP</i>	YKV	CKV	0.108	0.020	0.738	−0.877	10.1

WT, Wild-type tripeptide; MT, Mutated tripeptide.

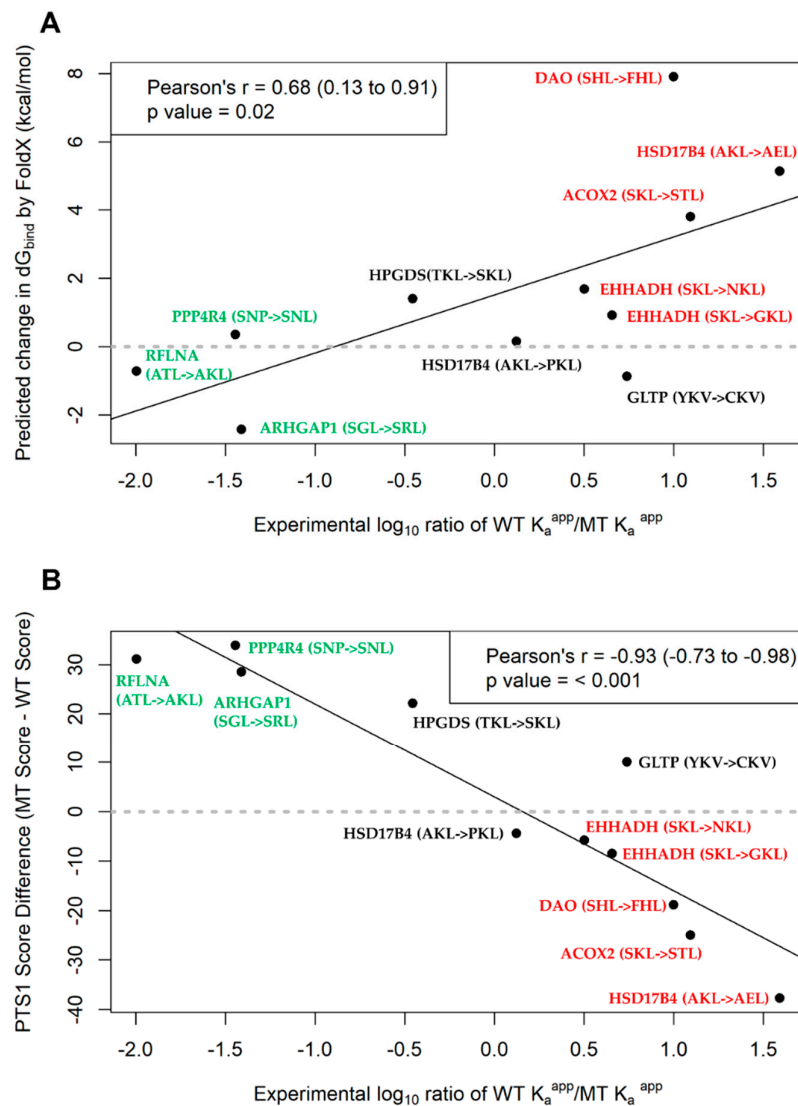
FoldX is an empirical force field that can be used to predict a change in interaction energy between the ligand and receptor upon one or more amino acid mutations at their interaction interface. Positive  $ddG_{bind}$  values denote a reduction in binding strength, and conversely, a negative value suggests an improved interaction. Based on the results in Table 3, seven of the 11 point mutations were predicted to destabilize the PTS1–PEX5 interaction while the rest of the point mutants were predicted to stabilize further the interaction between the two binding partners. Specifically, for the LoF mutants that were confirmed by IFA, FoldX correctly predicted ablation of PTS1 for five of the six variants as most of PTS1 generated were nearly or completely cytosolic after introduction of the point mutant. For *HSD17B4* A759P which retained peroxisome localization even after mutation, FoldX predicted a slight ablated interaction ( $ddG_{bind} = +0.154$  kcal/mol) between the *HSD17B4* mutant and PTS1 receptor (PEX5) which was not a wrong prediction given the level of significance for FoldX energy changes is  $\pm 0.5$  kcal/mol.

At the same time, of the five GoF related SNVs, FoldX projected that three would lead to higher interaction affinity and thus, the peptide can act as PTS1. In comparison to the IFA results where all

point mutants generated PTS1 signals except for GLTP Y207C, we observed that two (RLFNA and ARHGAP1) missense variants were correctly predicted to stabilize the interaction. In total, FoldX correctly predicted eight of the 11 SNVs to change affinity to PEX5 resulting in a change in localization. Further comparison of the absolute  $\Delta G_{\text{bind}}$  values with the FRET affinity measurements using correlation analysis (Figure 6A) revealed that the  $\Delta G_{\text{bind}}$  values had a statistically significant correlation with the  $\log_{10} K_a^{\text{app}}$  ratio of the affinity measurements (Pearson's  $r = 0.68$ ,  $p = 0.02$ ).

The PTS1-predictor is a prediction tool that assesses PTS1 quality using a position-specific scoring matrix developed from a database of known PEX5 binding PTS1-containing proteins and peptides from two hybrid assays. For an input consisting of the primary sequence of a protein, the algorithm categorizes the protein into three categories: non-peroxisomal ("Not Targeted"), peroxisomal ("Targeted"), and intermediate between the two categories ("Twilight zone").

For the LoF mutants, the tool predicted that half of the six SNVs would cause a loss in peroxisome targeting from "Twilight zone" to "Not targeted" or "Targeted" to "Twilight zone"/"Not targeted." Conversely, the other three variants were predicted to have no change in localization even after point mutation. When compared to the IFA results, four of these predictions were in concordance while the other two EHHADH gene variants were incorrectly predicted to have no change in localization. Concurrently, the GoF-associated variants were all predicted to change from being "Not targeted" or "Twilight zone" to being "Targeted" to the peroxisome after point mutation (Table 2). Three of these five predictions were true based on the IFA results (ARHGAP1, RFLNA, PPP4R4 gene variants), while for the other two variants no change in localization was observed. Altogether, the prediction tool correctly predicted localization change in seven of the 11 SNVs tested when considering the crude output labels. A more detailed analysis of the absolute score values from the algorithm was also performed, where we computed the score difference between the WT variant and the point mutant (Table 3) and correlated this with the FRET affinity measurements similar to the FoldX analysis. A statistically highly significant negative correlation between the score difference and the experimental measurements ( $\log_{10} K_a^{\text{app}}$  ratio, see Method 4.7) was obtained (Pearson's  $r = -0.93$ ,  $p < 0.001$ ) suggesting that as relative affinity of the WT over the MT variant increases (log ratio will increase), the predicted score difference decreases (Figure 6B). Since the method is sequence-based and the calculation is fast over a larger number of sequences (<1 sec per sequence), it would be well suited for high-throughput screening of effects of SNVs on PTS1 motifs. Additionally, in large datasets the calculated score differences could serve as a finer differentiation point for mutations resulting in the same pairs of categorical outcomes (e.g., "Targeted" -> "Not targeted").



**Figure 6.** Correlation analysis of the experimental measure of affinity ( $\log_{10} K_a^{\text{app}}$  ratio) with the prediction values: (A) FoldX-computed median  $ddG_{\text{bind}}$ , and (B) PTS1 Score difference (MT-WT) upon introducing a single point mutation into the native protein's tripeptide motif. Proteins and their mutants in green are associated with a GoF change from cytosolic to peroxisomal localization, while mutants in red are associated with a LoF in peroxisome localization. The rest had no change in subcellular localization compartments based on the IFA results. For each plot, we report the Pearson's  $r$  values and its associated  $p$ -value and 95% confidence interval (in brackets).

## 2.6. Disease-Relevance of Investigated SNVs as Consequence of Abolishment or Gain in Peroxisome Targeting

To evaluate disease relevance of the selected SNVs, we used the gnomAD database (comprising >140,000 control individuals) to assess the minor allele frequencies (MAF) of all investigated SNVs in addition to obtaining in silico predictions from variant pathogenicity prediction tools, such as SIFT and Polyphen-2. While no homozygous carriers were listed for any SNVs, MAFs ranged from singletons (one allele in gnomAD) to a maximum of 0.028% (79 alleles in gnomAD). To account for potential existing links to human diseases, we screened publicly available databases that provide evidence for associations between genetic variants and human diseases. In the ClinVar database, we

found the most frequent variant S721G (in EHHADH) to be classified as “likely benign” while all other SNVs were not listed. Collectively in the entire dataset of over 30,000 SNVs, 29 of them were associated with at least one “pathogenic” or “likely pathogenic” clinical significance in this database. Of note, only four of the genes have so far been linked to human diseases according to the Online Mendelian Inheritance in Man (OMIM) database. To date, there is only limited evidence for disease associations of DAO (schizophrenia) [34–36], RFLNA (spondylocarpotarsal synostosis) [37], and EHHADH (Fanconi renotubular syndrome) [21,22]. However, homozygous variants in HSD17B4 (D-bifunctional protein deficiency (MIM# 261515), Perrault syndrome 1 (MIM# 233400)) and ACOX2 (Bile acid synthesis defect, congenital (MIM# 617308)) have reliably been associated with monogenic diseases. In contrast, no human diseases have so far been linked to the genes PPP4R4, ARHGAP1, HPGDS, and GLTP according to OMIM. Independent of known disease associations, effects of the studied SNVs could be detrimental for the respective protein’s molecular and cellular functions as discussed below.

For the D-bifunctional enzyme (HSD17B4) which is associated with severe genetic disease (Table 4), inactivation of the PTS1 creates a purely cytosolic enzyme that cannot exert its peroxisomal function of fatty acid degradation and bile acid side chain shortening [38]. Therefore, this mutation should mimic an inactive enzyme with a recessive Mendelian inheritance.

**Table 4.** Summary of gene and variant-specific information from gnomAD, OMIM, and ClinVar.

Gene	Associated Disease (OMIM)	rsID	ClinVar	gnomAD Allele Frequency	PolyPhen-2	SIFT
<i>HSD17B4</i>	DBP deficiency (type I-III), Perrault Syndrome 1	rs1049954328	NA	Singleton	Possibly damaging	Deleterious
		rs751064948	NA	Singleton	Probably damaging	Deleterious
<i>DAO</i>	Schizophrenia *	rs143732132	NA	0.018%, No homozygotes	Possibly damaging	Deleterious
<i>ACOX2</i>	Bile acid synthesis defect, congenital	rs1182979757	NA	Singleton	Possibly damaging	Deleterious
<i>EHHADH</i>	Fanconi renotubular syndrome 3	rs138945273	Likely benign	0.028%, No homozygotes	Benign	Deleterious
		rs968361154	NA	Singleton	Benign	Tolerated
<i>PPP4R4</i>	None	rs576288488	NA	0.009%, No homozygotes	Benign	Deleterious **
<i>RFLNA</i>	None	rs760206157	NA	0.003%, No homozygotes	Benign	Tolerated
<i>ARHGAP1</i>	None	rs747965330	NA	<0.001%, No homozygotes	Probably damaging	Deleterious **
<i>HPGDS</i>	None	rs765481546	NA	Singleton	Possibly damaging	Deleterious
<i>GLTP</i>	None	rs773190605	NA	Singleton	Possibly damaging	Deleterious

\* Association awaiting confirmation on OMIM, \*\* Low confidence prediction. rsID, Reference single nucleotide polymorphism cluster identification; gnomAD, Genome Aggregation Database.

Acyl-CoA oxidase 2 (ACOX2) is an enzyme that mediates FAD-dependent dehydrogenation of the side chain of bile acids [39]. The lack of peroxisomal ACOX2 due to the SNV inactivating the PTS1 should ablate side-chain shortening of the bile-acid precursor. Thus, patients expressing only this variant of ACOX2 (-STL) should present a comparable severity in pathology as patients lacking the enzyme.

D-amino acid oxidase (DAO) is an enzyme that degrades amino acids, and its gene locus was found linked to schizophrenia [34–36] along with several links to amyotrophic lateral sclerosis (ALS) [40]. However, it is unclear whether the peroxisomal localization of DAO is linked to these pathologies or whether a cytosolic subfraction of the enzyme performs the relevant activity. This finding renders a clear prediction of the phenotype of patients equipped exclusively with this

predominantly cytosolic DAO variant (DAO-S345F) difficult. However, such variant could also serve as an interesting tool to disentangle the relationship between the localization of DAO and its different substrates.

EHHADH functions as enoyl-CoA hydratase/3-hydroxyacyl CoA dehydrogenase. Interestingly, a mutation has been described as being very close to the N-terminus of EHHADH (E3K), which generates a mitochondrial targeting signal in the enzyme [21] and is assumed to interfere with mitochondrial metabolism. Patients with this mutation suffer from renal dysfunction (Inherited Renal Fanconi's Syndrome) [21,22] which is inherited in an autosomal dominant manner. Thus, the latter model differs from the expected outcome of an SNV affecting the PTS1, because in this case the peroxisomal import is prevented, but no mitochondrial import is expected. In contrast, the phenotype is expected to resemble the mouse model lacking EHHADH, which did not present with any obvious pathology [41], although recently, a defect in the degradation of dicarboxylic fatty acids has been reported [42].

Besides the above peroxisomal proteins that are postulated to lose their PTS1 through the effects of the studied SNVs, we also studied normally cytosolic proteins that have gained a PTS1 motif through SNVs. Four such proteins have been discussed below as they were found to have gene variants potentially generating a PTS1 motif or have increased affinity to the PTS1 receptor after point mutation.

A part of the regulatory subunit of protein phosphatase 4, PPP4R4 is a binding partner of protein phosphatase 4 (PP4) with which it forms a phosphatase specific complex within the cytosol [43] and shares structural similarity with other phosphatase binding proteins such as PR65/A, including several HEAT-motifs [43]. Thus, PPP4R4 might be involved in the regulation of phosphatase activity or target protein selection. PP4 has been linked to diverse physiological functions such as DNA-repair [44], and mice lacking PP4 are embryonically lethal. However, complex formation of PPP4R4 and the core-enzyme (PP4) might even induce a partial peroxisomal co-import of PP4 with PPP4R4 variants harboring a PTS1, which should have even more drastic effects.

Refilin A (RFLNA) is a member of the refilin family (RFLNA and RFLNB), which are filamin-binding short-lived actin regulators coordinating nuclear movements in complex processes such as cell migration and differentiation [45], but also the dynamics of lamellipodium protrusions [46]. Since both refilin proteins are unstable proteins, the consequence of a cytosolic reduction of RFLNA by its mistargeting to peroxisomes is hard to predict. Recently, a homozygous mutation in RFLNA was found in a patient suffering from spondylarcatarsal synostosis syndrome, a rare syndromic skeletal disorder characterized by disrupted vertebral segmentation with vertebral fusion, carpal and tarsal synostosis, but also with scoliosis and short stature [37].

Rho GTPase activating protein 1 (ARHGAP1) activates the GTPase activity of the Ras-homologue Rho and can be found in the literature under different names (RhoGAP, RhoGAP1, CDC42GAP, and p50rhoGAP). Interestingly, ARHGAP1 has been found as a target of the micro-RNA hs-miR-940, which is spread via exosomes from diverse cancer cells and promotes osteogenic differentiation of human mesenchymal stem cells [47]. Similarly, miR-130b downregulates ARHGAP1 to drive the development of Ewing sarcoma [48], whereas overexpression of ARHGAP1 reduced the proliferation and migration of human cervical carcinoma cells [49]. The physiological relevance of the modulation of ARHGAP1-levels by micro-RNAs suggests that the level of this protein is rather critical, and experimental downregulation of ARHGAP1 causes an increase in tumor marker alkaline phosphatase activity, whereas overexpression of ARHGAP1 caused its reduction [47]. Thus, the level of ARHGAP1 appears critical for the aggressiveness of cancer cells, rendering a dysregulation of its cytosolic level by mislocalization to peroxisomes a prime candidate for further studies.

The hematopoietic prostaglandin D synthase (HPGDS) converts the prostaglandin PGH<sub>2</sub> into PGD<sub>2</sub> in a glutathione dependent manner [50], but this reaction in vertebrates is performed either by HPGDS (hematopoietic) or PTGDS (lipocalin type). Importantly, the enzyme serves the production of prostanoids in the immune system and mast cells, and its product, PGD<sub>2</sub>, mediates allergic asthma [51–53]. Moreover, the protein is also protective against cerebral ischemia of mice [54] and was

specifically found in microglia of mice brains [55]. Finally, an SNV in HPGDS has been linked to the increased probability of developing testicular germ cell tumor [56]. Interestingly, HPGDS forms homodimers [57], and thus, the arbitrary generation of a PTS1 may not only cause mislocalization of this variant, but it might even co-import PTS-less HPGDS engaged in a dimer into peroxisomes. The effect would even further downregulate the cytosolic level of the enzyme, although the specific consequences of downregulation of HPGDS levels in human cells are difficult to predict based on the variety of different functions it performs.

### 3. Discussion

In this study, we examined the possibility of naturally occurring SNVs creating or abolishing a PTS1 through a non-synonymous change in the primary sequence of a protein. Each person's genome is estimated to contain around 6000 to 10,000 missense variants [58]. While the majority might be inconsequential to health, non-synonymous SNVs resulting in missense mutations can be deleterious to protein function and manifest as severe diseases at the phenotypic level. They have been associated with over 1000 diseases, and the Human Gene Mutation Database (HGMD) currently lists over 100,000 disease-associated missense variants [4]. In this study, in the context of protein localization in the peroxisome, we asked if missense variants found in a sequence variation database could result in the gain or loss of PTS1-mediated peroxisome targeting especially in disease-relevant genes. To this end, we mined gnomAD for SNVs occurring at the PTS1 motif "hotspot" focusing on variants located in the codons encoding the C-terminal tripeptide (Figure 1) of over 30,000 proteins and their isoforms. Over 35,000 unique missense variants were found to cause amino acid substitutions at the tripeptide end for more than 15,000 genes (Figure 2), and we searched for SNVs that possibly led to both gain and loss of function in terms of PTS1-mediated peroxisome localization.

For loss-of-function mutants, the functional consequences of mislocalization of peroxisomal enzymes should be the ablation of enzyme function because essential peroxisomal activity cannot be performed. The now cytosolic enzyme can no longer process educts within the peroxisome, and even if the educts could leave the peroxisomes, the enzyme occurs at a lower concentration in the cytosol and might even require cooperation with one of the peroxisomal enzymes not available in the cytosol to perform its function. In our study, we identified five PTS1-inactivating SNVs within four proteins (Figures 3,5A), which should cause a (nearly) complete loss of peroxisomal activity.

Unlike the LoF variants, the concept of disease-causing SNVs due to *de novo* generated targeting signals is new, and it is most interesting for the evaluation of proteins usually found in the cytosol. These proteins should be sensitive to such effects as soluble peroxisomal matrix proteins are transported from the cytosol into the organelle. Moreover, the position of PTS1 at the extreme C-terminus renders its free accessibility much higher than any internal sequence. Although such novel PTS1 are conceptually gain of function (GoF) mutants as they provide a novel faculty to the protein, genetically they should be loss of function because peroxisomal import should sequester the protein from its place of action, whereas the proteins of the other allele should remain in the cytosol. Thus, SNVs of the type described above should become disease-relevant only when both alleles are affected or when the gene dose is critical. However, in a subset of proteins such mutations might even generate a dominant phenotype because oligomeric proteins have been found to become imported in an oligomeric state [59]. Such mechanism has recently been demonstrated for the import of SOD1 [60] and was suggested as a driving force for the co-import of homo-oligomeric protein complexes in which a subfraction of proteins harbor a PTS1 due to reading through a stop codon [61].

In this study, we investigated the effect of annotated SNVs within the C-terminal tripeptide of five human cytosolic proteins, which were predicted to generate a PTS1 (Table 2). In four of these cases (except GLTP), the C-terminus was able to act as a PTS1 and to interact with the receptor PEX5, which suggested that the encoding proteins were also imported into peroxisomes (unless the C-terminus was embedded within the protein) (Figure 4, Figure 5B). None of these SNVs have been identified in a database compiling SNVs associated with diseases (ClinVar) [62]. The observation that none of these SNVs was found in a homozygous state in a database collecting SNVs of healthy (symptom-free) subjects was limited by the low frequency of these mutations.



We also showed that filtering based on a combination of predictions from computational tools, such as the PTS1 predictor and subcellular localization annotation data (see Method), were of value and could be applied to sieve the vast amount of sequence variation data to identify high-quality candidates for further analysis. Specifically, this proved useful when studying GoF variants, because unlike the LoF variants where we could focus on a smaller number of known PTS1-containing proteins; the entire proteome had to be searched to check for *de novo* generation of a PTS1. Filtering in this manner reduced the pool of candidates from over 30,000 to about 16 variants in 16 genes (Table S2), which would be a manageable size for testing even if all the variants were considered.

As there was some degree of success with using the PTS1-predictor to identify high-quality candidates in the GoF dataset, we decided to assess this tool and another computational method used previously in another study [26]. We sought to determine if they would be useful for future application to newly discovered SNVs to assess their functional outcome on PTS1-mediated transport. Hence, we evaluated the two prediction methods via comparison with the IFA results and the FRET affinity measurements. For the IFA comparison, the PTS1-predictor made correct predictions for ~60% of the tested SNVs, while FoldX had a hit rate of ~70%. Additionally, comparison with affinity measurements demonstrated that both FoldX-computed median  $ddG_{\text{bind}}$  (Figure 6A) and the PTS1-predictor generated score difference (Figure 6B) had statistically significant correlations with the experimental logarithmic (scaled) ratio of  $K_a^{\text{APP}}$ . The ratio explains the relative reduction or gain in affinity of the point mutant compared to the native protein (log ratio of >0 means that the MT have a lower affinity while if log ratio is <0, MT has greater affinity than the WT). These prediction values may act as a good surrogate estimate for the experimentally-derived ratio and may be beneficial in further prioritizing hits from categorical prediction outcomes of the PTS1 predictor. In addition, we noted that extreme values from these prediction methods could make a distinction between missense variants that induced a change in subcellular localization, and the variants that did not as shown by the segregation of the GoF (in green) and LoF (in red) mutants in the plots from Figure 6. However, further validation with a larger sample set containing a more diverse representation of tripeptide combinations must be performed to determine the robustness of these prediction methods across different situations.

Overall, we demonstrated that sequence variations occurring naturally in the human population have the potential to cause gain and loss in PTS1-mediated peroxisome targeting. Collaboratively, we showed that bioinformatics approaches could be used to search vast amounts of sequence data, and then filter and isolate an experimentally tractable number of candidates for further validation by bench experiments. As the search was limited to the gnomAD database which consists of germline variants depleted for human samples with severe genetic diseases, it would be interesting to perform a similar analysis in a database with somatic variants, such as the Catalogue of Somatic Mutations in Cancer (COSMIC) or The Cancer Genome Atlas (TCGA). Further work will also be required to gain a deeper mechanistic understanding of mistargeting of these proteins and their implications in disease, if at all, and a similar study could be extended to other targeting signals.

## 4. Materials and Methods

### 4.1. Obtaining CCDS Protein Coding Genomic Coordinates and Sequence Information

FASTA files containing protein sequence, DNA sequence, and protein coding coordinates for every CCDS transcript were downloaded from <ftp://ftp.ncbi.nlm.nih.gov/pub/CCDS/archive/Hs105/>. The relevant data were then extracted and consolidated from three files (“CCDS\_protein.current.faa”, “CCDS\_nucleotide.current.fna” and “CCDS\_protein.current.faa”) using a Python (v2.7.13) script. All transcripts with the CCDS status of “Withdrawn” were excluded from further analysis. For each unique protein coding transcript, its corresponding chromosome number and genomic coordinates for the last three amino acid coding codons preceding the stop codon at the C-terminus were obtained to create a query file for subsequent subsetting of gnomAD data.

### 4.2. Parsing gnomAD vcf Files to Obtain Relevant SNV Information

gnomAD v.2.1.1 variant call files (.vcf) for both the Exome and exome-calling intervals of the Genome call sets were downloaded from <https://gnomad.broadinstitute.org/downloads> along with their corresponding tabix-indexed files (.tbi). The query file from section 4.1 was used to subset each compressed vcf file separately using tabix v1.7-2 on the Windows Subsystem for Linux (WSL) to obtain all variants occurring at the positions indicated in the query file for each chromosome. A Python script was then created to consolidate, clean, and filter the vcf outputs from tabix. All other variants besides missense variants were excluded from further analysis. For all variants, the Variant Effect Predictor (VEP) annotation tool [63] (v96) was used to acquire variant-specific information such as allele frequencies, ClinVar annotations, and in silico pathogenicity predictions from tools such as PolyPhen2 [64] and SIFT [65].

#### 4.3. Obtaining Subcellular Localization Annotations from UniProt

Each unique CCDS transcript from section 4.1 was mapped to its corresponding UniProt ID, where available, using the UniProt mapping tool available at <https://www.uniprot.org/uploadlists/>. The mapped UniProt IDs were then used to query the UniProt database and data from the “Subcellular\_location\_CC” results column was extracted into a tab-delimited file (.tsv) for all queried UniProt IDs. These annotations were then merged with information from sections 4.1 and 4.2 with the help of a script utilizing the Pandas library (v0.23.4) in Python. Where available, isoform-specific annotations were retained if a gene had more than one isoform; otherwise the annotation of the canonical transcript of a gene will be extended to all its other isoforms.

#### 4.4. Obtaining PTS1 Predictor Results for Native Transcripts and their Corresponding Point Mutants

The PTS1 prediction algorithm developed by Neuberger et al. [25] was used to predict peroxisome localization of native proteins and their variants after SNV induced point mutation. The primary sequence of the protein for each CCDS transcript was consolidated, and the last 12 amino-acid sequences at the C-termini of the transcripts were used as inputs for the prediction algorithm (Metazoa option). The output from the algorithm was then combined with data from sections 4.1–4.3 into a single file (Table S3), and the file was used to search for SNVs relevant to this study.

#### 4.5. Prioritization of SNV before Final Selection for Experimental Testing

For the LoF variants, a curated list of 35 gene encoding proteins harboring a PTS1 and containing gnomAD-associated missense variants capable of mutating the PTS1 C-terminal tripeptide region was created (Table S1). Where available, each gene was annotated with disease annotations from OMIM and genes without these annotations were excluded from further analysis. From the narrowed gene list, further genes were excluded if their products were dually or partly localized in other cell compartments besides the peroxisome based on the UniProt subcellular localization data (see section 4.3). Specifically, the gene was included in our analysis only if its native product contained the keyword “Peroxisome” but did not have the keywords “Secreted,” “Nucleus,” “Mitochondrion,” “Endosome,” and “Cytoplasm” in its subcellular localization annotation data. Note that at this stage, AGXT (see Table S1) was included although it had the “Mitochondrion” keyword as the mutant form was annotated to be in the mitochondrion but the native protein is predominantly peroxisomal (Table S1). All SNVs found in these genes also must pass gnomAD variant quality filters (defined as a “PASS” in the “FILTER” column in either the gnomAD Genome or Exome .vcf call sets). The gnomAD associated quality filters are a collection of hard filters and a random forest classifier model that is applied on each variant to filter out low-quality variants. Some of these filters are based on defined threshold values for parameters, such as the inbreeding coefficient and the genotype quality [28] (further details at <https://macarthurlab.org/2018/10/17/gnomad-v2-1/>). The variant would be annotated “PASS” if it passes all filters otherwise it would be labeled with the filter that it has failed. To ensure that only high-quality variants likely to be real and not artefactually produced by sequencing errors are studied, all failed variants were not considered for further analysis in this study. After filtering programmatically as described above, we picked only SNVs resulting in mutant

tripeptides deviating from the traditional PTS1 motifs [18] in the WT tripeptide motifs for testing of PTS1 ablation.

For the GoF variants, three filters were programmatically applied on the consolidated data from section 4.3: (i) SNV must pass gnomAD associated quality filters similar to the LoF SNVs, (ii) UniProt subcellular localization annotation of the protein contains the keyword “Cytoplasm” but excludes the keywords “Secreted”, “Nucleus”, “Mitochondrion”, “Endosome”, and “Peroxisome”, and (iii) upon point mutation, PTS1 predictor results must change from “Twilight” or “Not targeted” to “Targeted”. If the SNV fell on a gene with more than one CCDS transcript, the longest transcript was chosen for analysis. After filtering programmatically, we picked 5 SNVs generating a PTS1-like WT tripeptide from non-PTS1 WT tripeptide motif in reasonably long proteins (<1000 amino acids) without transmembrane domains for testing.

#### 4.6. FoldX Prediction of Change in Free Binding Energy due to Point Mutations made at PTS1 Ligand-receptor Interface

A similar procedure as in Reference [26] was used. Using a previously crystallized structure of a PTS1 ligand–PEX5 interaction (pdb ID: 2c0l) [29] as a model, we first truncated the protein–ligand to 12 amino acids in length from the C-terminus. The complex was then energy minimized using the energy minimization function in Yasara Structure (v18.2.7) [66] with the AMBER15FB force field at default settings. The process included minimization of the backbone with short simulated annealing molecular dynamics simulations. Next, the RepairPDB function of the FoldX (v4.0) [33] plugin in Yasara was applied to prepare the complex for FoldX calculation. With the reference ligand being the wild-type protein sequence of each gene, we used the “Mutate residue” option in the FoldX plugin to introduce the point mutant at tripeptide (default settings, 5 runs). We calculated and analyzed the median change in free binding energy ( $\Delta G_{\text{bind}}$ , kcal/mol) of the complex upon mutation of the ligand at the PTS1 ligand–receptor interface. The illustration of the complex was created using Yasara.

#### 4.7. Statistical Analyses

From the affinity measurements of PTS1 quality using FRET, we calculated the  $\log_{10}$  ratio of relative affinity of the wild-type PTS1 (mean  $K_{\text{app}}$  WT) over its mutated variant (mean  $K_{\text{app}}$  MT), given by the formula:

$$\log K_a^{\text{app}} \text{ ratio} = \log_{10} \left( \frac{K_a^{\text{app}} \text{ WT}}{K_a^{\text{app}} \text{ MT}} \right) \quad (1)$$

The log-ratio above was then separately correlated with predicted values from the outputs of the i) FoldX-computed median  $\Delta G_{\text{bind}}$  (kcal/mol) values and ii) PTS1 predictor (see Section 4.4) score difference (MT Score–WT Score). Correlation analysis (Pearson’s) and hypothesis testing were performed using the base statistical package in R (v3.5.1) and similarly, all graphs were plotted using R.

#### 4.8. Cloning

EGFP-fusion proteins: plasmid EGFP-C3 (Clontech) was digested with restriction enzymes BglII and HindIII and then ligated with two oligos of interest (Table S4) annealed previously. Plasmids were sequenced, and large amounts of DNA were prepared via midi-preps.

#### 4.9. Cell Culture

Cells were cultivated in DMEM supplemented with 10% fetal calf serum (FCS) (Sigma-Aldrich, St. Louis, MO, USA), 2 mM L-glutamine, 50 U/mL penicillin, and 100  $\mu\text{g/mL}$  streptomycin ((all Lonza, Basel, Switzerland). The cells were transfected using Turbofect (Thermo-Scientific, Waltham, MA, USA) according to the manufacturer’s instructions using 1  $\mu\text{g}$  DNA and 1.25  $\mu\text{g}$  Turbofect. Then, 40 h after transfection, the cells were fixed for 15 min with 4% paraformaldehyde in PBS (phosphate

buffered saline). Cells were washed, permeabilized (5 min with 0.1% Triton X-100 in PBS) and blocked in blocking solution (PBS with 5% FCS and 0.5% bovine serum albumin (BSA, Roche, Basel, Switzerland)). After incubation with primary antibodies (rabbit:  $\alpha$ -PMP70; 1:2000, ABR, Golden CO, USA), the slides were washed with PBS several times and exposed to compatible secondary antibodies (Cy3-labelled donkey- $\alpha$ -rabbit IgG, 1:400, Jackson Immuno Research Laboratories, West Grove, PA, USA). Finally, the cells were mounted in PBS/glycerol (1:9) with 3% DABCO (Sigma, St. Louis, MO, USA). For microscopic analysis, an inverted microscope IX71 (Olympus, Wien, Austria) equipped with a CCD camera (CAM-XM10) and appropriate filter sets were used together with C-M-cell software (Olympus, Wien, Austria).

#### 4.10. FRET Affinity Measurements

In the approach, we determined the apparent interaction strength between two fluorescent proteins by measuring the energy transfer efficiency between an excited donor molecule (donor) and a proximally located acceptor molecule (acceptor). Thus, we expressed EGFP variants terminating in the different C-terminal PTS1 peptides as donors together with a fusion protein consisting of the PTS1-binding domain of PEX5 (PEX5<sup>TPR</sup>) fused to the C-terminus of mCherry (mCherry-PEX5<sup>TPR</sup>). Utilizing a cell line lacking peroxisomal matrix protein import due to a lack of endogenous full-length PEX5, all the donor proteins remained in the cytosol and their interaction strength with PEX5 could be determined. PEX5<sup>-/-</sup> cells were prepared at ~70% confluency in 24-well cell culture dishes. The cells were transfected with plasmids bearing the GFP-tagged PTS1 variants and mCherry-tagged PEX5<sup>TPR</sup> in 2 different ratios (0.4 + 0.6  $\mu$ g and 0.6 + 0.4  $\mu$ g), using TurboFect transfection Reagent (Thermo Scientific, Waltham, MA, USA) according to protocol. After incubation for 24 h, the cells were washed in PBS and detached with Trypsin (100  $\mu$ L per well). Detached cells were transferred to a 96 well v-bottom plate (combining the two different ratios) and then spun for 2 min at 300 g. Supernatant was removed, and the cells were resuspended in 200  $\mu$ L PBS. The cells were then immediately measured using FlowFRET. FlowFRET was performed on a Cytotflex S flow cytometer (Beckman Coulter, Brea, CA, USA). The donor channel was measured at a 488 nm excitation and 525/40 emission. FRET was measured at a 488 nm excitation and 610/20 emission. The acceptor channel was measured at a 561 nm excitation and 610/20 emission. Data were treated as described in our previous work [32].

**Supplementary Materials:** Supplementary materials can be found at [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1). Table S1: List of 35 genes curated for LoF prioritization, including disease annotation and subcellular localization data. Table S2: GoF SNVs shortlisted before final selection. Table S3: Complete SNV dataset for all missense variants found in the PTS1 “hotspot” region in two gnomAD call sets (G: Genomes, and E: Exomes). Table S4: Plasmids and oligonucleotides used.

**Author Contributions:** Conceptualization, S.M.S., M.K.(Markus Kunze), and C.S.C.; methodology, S.M.S., M.K.(Markus Kunze), and C.S.C.; validation, M.K.(Markus Kunze) and C.S.C.; formal analysis, S.M.S., M.K.(Markus Kunze), and C.S.C.; investigation, M.K.(Markus Kunze), M.K.(Martin Krenn), B.H., and C.S.C.; resources, M.K.(Markus Kunze), B.H., and J.B.; software, C.S.C. and S.M.S.; data curation, C.S.C. and S.M.S.; writing—original draft preparation, C.S.C., M.K.(Markus Kunze), and S.M.S.; writing—review and editing, all authors were involved; visualization, M.K.(Markus Kunze), B.H., and C.S.C.; supervision, S.M.S. and J.B.; project administration, C.S.C. and M.K.(Markus Kunze); funding acquisition, S.M.S. and J.B.

**Funding:** This research was partially funded under A\*STAR grant number IAF-PP6 (H18/01/a0/B14) and the Austrian Science Fund (FWF I2738-B26).

**Acknowledgments:** The authors would like to thank the Genome Aggregation Database (gnomAD) and the groups that provided exome and genome variant data to this resource. A full list of contributing groups can be found at <https://gnomad.broadinstitute.org/about>.

**Conflicts of Interest:** The authors declare no conflict of interests.

#### Abbreviations

gnomAD	Genome Aggregation Database
CCDS	Consensus Coding Sequence Project

UniProt	Universal Protein resource
PTS	Peroxisome targeting signal
PEX	Peroxin
SNV	Single nucleotide variant
dbSNP	The Single Nucleotide Polymorphism Database
vcf	Variant call file
tbi	Tabix-indexed file
GoF	Gain-of-function
LoF	Loss-of-function
EGFP	Enhanced green fluorescent protein
FRET	Fluorescence resonance energy transfer

## References

- Pinto, N.; Dolan, M.E. Clinically relevant genetic variations in drug metabolizing enzymes. *Curr. Drug Metab.* **2011**, *12*, 487–497.
- Landrum, M.J.; Lee, J.M.; Riley, G.R.; Jang, W.; Rubinstein, W.S.; Church, D.M.; Maglott, D.R. ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **2014**, *42*, D980–D985, doi:10.1093/nar/gkt1113.
- Merkulov, V.M.; Leberfarb, E.Y.; Merkulova, T.I. Regulatory SNPs and their widespread effects on the transcriptome. *J. Biosci.* **2018**, *43*, 1069–1075, doi:10.1007/s12038-018-9817-7.
- Stenson, P.D.; Mort, M.; Ball, E.V.; Shaw, K.; Phillips, A.; Cooper, D.N. The Human Gene Mutation Database: Building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum. Genet.* **2014**, *133*, 1–9, doi:10.1007/s00439-013-1358-4.
- Sherry, S.T.; Ward, M.H.; Kholodov, M.; Baker, J.; Phan, L.; Smigielski, E.M.; Sirotkin, K. dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res.* **2001**, *29*, 308–311.
- Lek, M.; Karczewski, K.J.; Minikel, E.V.; Samocha, K.E.; Banks, E.; Fennell, T.; O'Donnell-Luria, A.H.; Ware, J.S.; Hill, A.J.; Cummings, B.B.; et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **2016**, *536*, 285–291, doi:10.1038/nature19057.
- Kucukkal, T.G.; Petukh, M.; Li, L.; Alexov, E. Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. *Curr. Opin. Struct. Biol.* **2015**, *32*, 18–24, doi:10.1016/j.sbi.2015.01.003.
- Yates, C.M.; Sternberg, M.J. The effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on protein-protein interactions. *J. Mol. Biol.* **2013**, *425*, 3949–3963, doi:10.1016/j.jmb.2013.07.012.
- David, A.; Sternberg, M.J. The Contribution of Missense Mutations in Core and Rim Residues of Protein-Protein Interfaces to Human Disease. *J. Mol. Biol.* **2015**, *427*, 2886–2898, doi:10.1016/j.jmb.2015.07.004.
- Laurila, K.; Vihinen, M. Prediction of disease-related mutations affecting protein localization. *BMC Genom.* **2009**, *10*, 122, doi:10.1186/1471-2164-10-122.
- Fodor, K.; Wolf, J.; Erdmann, R.; Schliebs, W.; Wilmanns, M. Molecular requirements for peroxisomal targeting of alanine-glyoxylate aminotransferase as an essential determinant in primary hyperoxaluria type 1. *PLoS Biol.* **2012**, *10*, e1001309, doi:10.1371/journal.pbio.1001309.
- Schrader, M.; Fahimi, H.D. The peroxisome: Still a mysterious organelle. *Histochem. Cell Biol.* **2008**, *129*, 421–440, doi:10.1007/s00418-008-0396-9.
- Waterham, H.R.; Ferdinandusse, S.; Wanders, R.J. Human disorders of peroxisome metabolism and biogenesis. *Biochim. Biophys. Acta* **2016**, *1863*, 922–933, doi:10.1016/j.bbamcr.2015.11.015.
- Steinberg, S.J.; Dodt, G.; Raymond, G.V.; Braverman, N.E.; Moser, A.B.; Moser, H.W. Peroxisome biogenesis disorders. *Biochim. Biophys. Acta* **2006**, *1763*, 1733–1748, doi:10.1016/j.bbamcr.2006.09.010.
- Braverman, N.E.; Raymond, G.V.; Rizzo, W.B.; Moser, A.B.; Wilkinson, M.E.; Stone, E.M.; Steinberg, S.J.; Wangler, M.F.; Rush, E.T.; Hacia, J.G.; et al. Peroxisome biogenesis disorders in the Zellweger spectrum: An overview of current diagnosis, clinical manifestations, and treatment guidelines. *Mol. Genet. Metab.* **2016**, *117*, 313–321, doi:10.1016/j.ymgme.2015.12.009.
- Wanders, R.J.; Waterham, H.R. Peroxisomal disorders: The single peroxisomal enzyme deficiencies. *Biochim. Biophys. Acta* **2006**, *1763*, 1707–1720, doi:10.1016/j.bbamcr.2006.08.010.
- Gould, S.G.; Keller, G.A.; Subramani, S. Identification of a peroxisomal targeting signal at the carboxy terminus of firefly luciferase. *J. Cell Biol.* **1987**, *105*, 2923–2931, doi:10.1083/jcb.105.6.2923.

18. Gould, S.J.; Keller, G.A.; Hosken, N.; Wilkinson, J.; Subramani, S. A conserved tripeptide sorts proteins to peroxisomes. *J. Cell Biol.* **1989**, *108*, 1657–1664.
19. Lametschwandtner, G.; Brocard, C.; Fransen, M.; Van Veldhoven, P.; Berger, J.; Hartig, A. The difference in recognition of terminal tripeptides as peroxisomal targeting signal 1 between yeast and human is due to different affinities of their receptor Pex5p to the cognate signal and to residues adjacent to it. *J. Biol. Chem.* **1998**, *273*, 33635–33643.
20. Brocard, C.; Hartig, A. Peroxisome targeting signal 1: Is it really a simple tripeptide? *Biochim. Biophys. Acta* **2006**, *1763*, 1565–1573, doi:10.1016/j.bbamcr.2006.08.022.
21. Klootwijk, E.D.; Reichold, M.; Helip-Wooley, A.; Tolaymat, A.; Broeker, C.; Robinette, S.L.; Reinders, J.; Peindl, D.; Renner, K.; Eberhart, K.; et al. Mistargeting of peroxisomal EHHADH and inherited renal Fanconi's syndrome. *N. Engl. J. Med.* **2014**, *370*, 129–138, doi:10.1056/NEJMoa1307581.
22. Assmann, N.; Dettmer, K.; Simbuerger, J.M.B.; Broeker, C.; Nuernberger, N.; Renner, K.; Courtneidge, H.; Klootwijk, E.D.; Duerkop, A.; Hall, A.; et al. Renal Fanconi Syndrome Is Caused by a Mistargeting-Based Mitochondriopathy. *Cell Rep.* **2016**, *15*, 1423–1429, doi:10.1016/j.celrep.2016.04.037.
23. Leiper, J.M. Inhibition of alanine:glyoxylate aminotransferase 1 dimerization is a prerequisite for its peroxisome-to-mitochondrion mistargeting in primary hyperoxaluria type 1. *J. Cell Biol.* **1996**, *135*, 939–951, doi:10.1083/jcb.135.4.939.
24. Kunze, M. Predicting Peroxisomal Targeting Signals to Elucidate the Peroxisomal Proteome of Mammals. *Subcell. Biochem.* **2018**, *89*, 157–199, doi:10.1007/978-981-13-2233-4\_7.
25. Neuberger, G.; Maurer-Stroh, S.; Eisenhaber, B.; Hartig, A.; Eisenhaber, F. Prediction of Peroxisomal Targeting Signal 1 Containing Proteins from Amino Acid Sequence. *J. Mol. Biol.* **2003**, *328*, 581–592, doi:10.1016/s0022-2836(03)00319-x.
26. Hochreiter, B.; Chong, C.S.; Hartig, A.; Maurer-Stroh, S.; Berger, J.; Schmid, J.A.; Kunze, M. A novel FRET approach quantifies the interaction strength of peroxisomal targeting signals and their receptor in living cells. **2019**, submitted to *Nature Chemical Biology*.
27. Laurila, K.; Vihinen, M. PROlocalizer: Integrated web service for protein subcellular localization prediction. *Amino Acids* **2011**, *40*, 975–980, doi:10.1007/s00726-010-0724-y.
28. Karczewski, K.J.; Francioli, L.C.; Tiao, G.; Cummings, B.B.; Alfoldi, J.; Wang, Q.; Collins, R.L.; Laricchia, K.M.; Ganna, A.; Birnbaum, D.P.; et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* **2019**, 531210, doi:10.1101/531210.
29. Stanley, W.A.; Filipp, F.V.; Kursula, P.; Schuller, N.; Erdmann, R.; Schliebs, W.; Sattler, M.; Wilmanns, M. Recognition of a functional peroxisome type 1 target by the dynamic import receptor pex5p. *Mol. Cell* **2006**, *24*, 653–663, doi:10.1016/j.molcel.2006.10.024.
30. Schluter, A.; Real-Chicharro, A.; Gabaldon, T.; Sanchez-Jimenez, F.; Pujol, A. PeroxisomeDB 2.0: An integrative view of the global peroxisomal metabolome. *Nucleic Acids Res.* **2010**, *38*, D800–D805, doi:10.1093/nar/gkp935.
31. Skoulding, N.S.; Chowdhary, G.; Deus, M.J.; Baker, A.; Reumann, S.; Warriner, S.L. Experimental validation of plant peroxisomal targeting prediction algorithms by systematic comparison of in vivo import efficiency and in vitro PTS1 binding affinity. *J. Mol. Biol.* **2015**, *427*, 1085–1101, doi:10.1016/j.jmb.2014.12.003.
32. Hochreiter, B.; Kunze, M.; Moser, B.; Schmid, J.A. Advanced FRET normalization allows quantitative analysis of protein interactions including stoichiometries and relative affinities in living cells. *Sci. Rep.* **2019**, *9*, 8233, doi:10.1038/s41598-019-44650-0.
33. Guerois, R.; Nielsen, J.E.; Serrano, L. Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More Than 1000 Mutations. *J. Mol. Biol.* **2002**, *320*, 369–387, doi:10.1016/s0022-2836(02)00442-4.
34. Corvin, A.; McGhee, K.A.; Murphy, K.; Donohoe, G.; Nangle, J.M.; Schwaiger, S.; Kenny, N.; Clarke, S.; Meagher, D.; Quinn, J.; et al. Evidence for association and epistasis at the DAOA/G30 and D-amino acid oxidase loci in an Irish schizophrenia sample. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **2007**, *144B*, 949–953, doi:10.1002/ajmg.b.30452.
35. Chumakov, I.; Blumenfeld, M.; Guerassimenko, O.; Cavarec, L.; Palicio, M.; Abderrahim, H.; Bougueleret, L.; Barry, C.; Tanaka, H.; La Rosa, P.; et al. Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 13675–13680, doi:10.1073/pnas.182412499.

36. Caldinelli, L.; Sacchi, S.; Molla, G.; Nardini, M.; Pollegioni, L. Characterization of human DAAO variants potentially related to an increased risk of schizophrenia. *Biochim. Biophys. Acta* **2013**, *1832*, 400–410, doi:10.1016/j.bbadis.2012.11.019.
37. Shimizu, H.; Watanabe, S.; Kinoshita, A.; Mishima, H.; Nishimura, G.; Moriuchi, H.; Yoshiura, K.I.; Dateki, S. Identification of a homozygous frameshift variant in RFLNA in a patient with a typical phenotype of spondylocarpotarsal synostosis syndrome. *J. Hum. Genet.* **2019**, *64*, 467–471, doi:10.1038/s10038-019-0581-9.
38. Huyghe, S.; Mannaerts, G.P.; Baes, M.; Van Veldhoven, P.P. Peroxisomal multifunctional protein-2: The enzyme, the patients and the knockout mouse model. *Biochim. Biophys. Acta* **2006**, *1761*, 973–994, doi:10.1016/j.bbalip.2006.04.006.
39. Vilarinho, S.; Sari, S.; Mazzacuva, F.; Bilguvar, K.; Esendagli-Yilmaz, G.; Jain, D.; Akyol, G.; Dalgic, B.; Gunel, M.; Clayton, P.T.; et al. ACOX2 deficiency: A disorder of bile acid synthesis with transaminase elevation, liver fibrosis, ataxia, and cognitive impairment. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 11289–11293, doi:10.1073/pnas.1613228113.
40. Mitchell, J.; Paul, P.; Chen, H.J.; Morris, A.; Payling, M.; Falchi, M.; Habgood, J.; Panoutsou, S.; Winkler, S.; Tisato, V.; et al. Familial amyotrophic lateral sclerosis is associated with a mutation in D-amino acid oxidase. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 7556–7561, doi:10.1073/pnas.0914128107.
41. Ferdinandusse, S.; Denis, S.; Overmars, H.; Van Eeckhoudt, L.; Van Veldhoven, P.P.; Duran, M.; Wanders, R.J.; Baes, M. Developmental changes of bile acid composition and conjugation in L- and D-bifunctional protein single and double knockout mice. *J. Biol. Chem.* **2005**, *280*, 18658–18666, doi:10.1074/jbc.M414311200.
42. Houten, S.M.; Denis, S.; Argmann, C.A.; Jia, Y.; Ferdinandusse, S.; Reddy, J.K.; Wanders, R.J. Peroxisomal L-bifunctional enzyme (Ehhadh) is essential for the production of medium-chain dicarboxylic acids. *J. Lipid Res.* **2012**, *53*, 1296–1303, doi:10.1194/jlr.M024463.
43. Chen, G.I.; Tisayakorn, S.; Jorgensen, C.; D'Ambrosio, L.M.; Goudreault, M.; Gingras, A.C. PP4R4/KIAA1622 forms a novel stable cytosolic complex with phosphoprotein phosphatase 4. *J. Biol. Chem.* **2008**, *283*, 29273–29284, doi:10.1074/jbc.M803443200.
44. Ramos, F.; Villoria, M.T.; Alonso-Rodriguez, E.; Clemente-Blanco, A. Role of protein phosphatases PP1, PP2A, PP4 and Cdc14 in the DNA damage response. *Cell Stress* **2019**, *3*, 70–85, doi:10.15698/cst2019.03.178.
45. Sneider, A.; Hah, J.; Wirtz, D.; Kim, D.H. Recapitulation of molecular regulators of nuclear motion during cell migration. *Cell Adh. Migr.* **2019**, *13*, 50–62, doi:10.1080/19336918.2018.1506654.
46. Gay, O.; Gilquin, B.; Assard, N.; Stuelsatz, P.; Delphin, C.; Lachuer, J.; Gidrol, X.; Baudier, J. Refilins are short-lived Actin-bundling proteins that regulate lamellipodium protrusion dynamics. *Biol. Open* **2016**, *5*, 1351–1361, doi:10.1242/bio.019588.
47. Hashimoto, K.; Ochi, H.; Sunamura, S.; Kosaka, N.; Mabuchi, Y.; Fukuda, T.; Yao, K.; Kanda, H.; Ae, K.; Okawa, A.; et al. Cancer-secreted hsa-miR-940 induces an osteoblastic phenotype in the bone metastatic microenvironment via targeting ARHGAP1 and FAM134A. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 2204–2209, doi:10.1073/pnas.1717363115.
48. Satterfield, L.; Shuck, R.; Kurenbekova, L.; Allen-Rhoades, W.; Edwards, D.; Huang, S.; Rajapakshe, K.; Coarfa, C.; Donehower, L.A.; Yustein, J.T. miR-130b directly targets ARHGAP1 to drive activation of a metastatic CDC42-PAK1-AP1 positive feedback loop in Ewing sarcoma. *Int. J. Cancer* **2017**, *141*, 2062–2075, doi:10.1002/ijc.30909.
49. Li, J.P.; Liu, Y.; Yin, Y.H. ARHGAP1 overexpression inhibits proliferation, migration and invasion of C-33A and SiHa cell lines. *Onco Targets Ther.* **2017**, *10*, 691–701, doi:10.2147/OTT.S112223.
50. Urade, Y.; Hayaishi, O. Prostaglandin D synthase: Structure and function. *Vitam. Horm.* **2000**, *58*, 89–120.
51. Matsuoka, T.; Hirata, M.; Tanaka, H.; Takahashi, Y.; Murata, T.; Kabashima, K.; Sugimoto, Y.; Kobayashi, T.; Ushikubi, F.; Aze, Y.; et al. Prostaglandin D2 as a mediator of allergic asthma. *Science* **2000**, *287*, 2013–2017, doi:10.1126/science.287.5460.2013.
52. Mendez-Enriquez, E.; Hallgren, J. Mast Cells and Their Progenitors in Allergic Asthma. *Front. Immunol.* **2019**, *10*, 821, doi:10.3389/fimmu.2019.00821.
53. Rittchen, S.; Heinemann, A. Therapeutic Potential of Hematopoietic Prostaglandin D2 Synthase in Allergic Inflammation. *Cells* **2019**, *8*, 619, doi:10.3390/cells8060619.
54. Liu, M.; Eguchi, N.; Yamasaki, Y.; Urade, Y.; Hattori, N.; Urabe, T. Protective role of hematopoietic prostaglandin D synthase in transient focal cerebral ischemia in mice. *Neuroscience* **2009**, *163*, 296–307, doi:10.1016/j.neuroscience.2009.06.027.



55. Mohri, I.; Eguchi, N.; Suzuki, K.; Urade, Y.; Taniike, M. Hematopoietic prostaglandin D synthase is expressed in microglia in the developing postnatal mouse brain. *Glia* **2003**, *42*, 263–274, doi:10.1002/glia.10183.
56. Chung, C.C.; Kanetsky, P.A.; Wang, Z.; Hildebrandt, M.A.; Koster, R.; Skotheim, R.I.; Kratz, C.P.; Turnbull, C.; Cortessis, V.K.; Bakken, A.C.; et al. Meta-analysis identifies four new loci associated with testicular germ cell tumor. *Nat. Genet.* **2013**, *45*, 680–685, doi:10.1038/ng.2634.
57. Inoue, T.; Irikura, D.; Okazaki, N.; Kinugasa, S.; Matsumura, H.; Uodome, N.; Yamamoto, M.; Kumasaka, T.; Miyano, M.; Kai, Y.; et al. Mechanism of metal activation of human hematopoietic prostaglandin D synthase. *Nat. Struct. Biol.* **2003**, *10*, 291–296, doi:10.1038/nsb907.
58. Kumar, S.; Dudley, J.T.; Filipinski, A.; Liu, L. Phylomedicine: An evolutionary telescope to explore and diagnose the universe of disease mutations. *Trends Genet.* **2011**, *27*, 377–386, doi:10.1016/j.tig.2011.06.004.
59. Leon, S.; Goodman, J.M.; Subramani, S. Uniqueness of the mechanism of protein import into the peroxisome matrix: Transport of folded, co-factor-bound and oligomeric proteins by shuttling receptors. *Biochim. Biophys. Acta* **2006**, *1763*, 1552–1564.
60. Islinger, M.; Li, K.W.; Seitz, J.; Volkl, A.; Luers, G.H. Hitchhiking of Cu/Zn superoxide dismutase to peroxisomes—evidence for a natural piggyback import mechanism in mammals. *Traffic* **2009**, *10*, 1711–1721, doi:10.1111/j.1600-0854.2009.00966.x.
61. Schueren, F.; Lingner, T.; George, R.; Hofhuis, J.; Dickel, C.; Gartner, J.; Thoms, S. Peroxisomal lactate dehydrogenase is generated by translational readthrough in mammals. *eLife* **2014**, *3*, e03640, doi:10.7554/eLife.03640.
62. 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, doi:10.1093/nar/gkv1222.
63. McLaren, W.; Gil, L.; Hunt, S.E.; Riat, H.S.; Ritchie, G.R.; Thormann, A.; Flicek, P.; Cunningham, F. The Ensembl Variant Effect Predictor. *Genome Biol.* **2016**, *17*, 122, doi:10.1186/s13059-016-0974-4.
64. Adzhubei, I.; Jordan, D.M.; Sunyaev, S.R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **2013**, *76*, 7–20, doi:10.1002/0471142905.hg0720s76.
65. Ng, P.C.; Henikoff, S. Accounting for human polymorphisms predicted to affect protein function. *Genome Res.* **2002**, *12*, 436–446, doi:10.1101/gr.212802.
66. Krieger, E.; Joo, K.; Lee, J.; Lee, J.; Raman, S.; Thompson, J.; Tyka, M.; Baker, D.; Karplus, K. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* **2009**, *77* (Suppl. 9), 114–122, doi:10.1002/prot.22570.

