Molecular Vision 2023; 29:217-233 <http://www.molvis.org/molvis/v29/217>
Received 18 August 2022 | Accepted 23 October 2023 | Published 25 October 2023

Protein modeling and in silico analysis to assess pathogenicity of *ABCA4* variants in patients with inherited retinal disease

Senem Cevik,[1] Nutsuchar Wangtiraumnuay,[2] Kristof Van Schelvergem,[3] Mai Tsukikawa,[4] Jenina Capasso,[5] Subhasis B. Biswas,[1] Barry Bodt,[6] Alex V. Levin,[5] Esther Biswas-Fiss[1]

(The first three authors contributed equally to this study.)

[1]Department of Medical and Molecular Sciences, University of Delaware College of Health Sciences, Newark, DE; [2]Department of Ophthalmology, Queen Sirikit National Institute of Child Health, Bangkok, Thailand; [3]Department of Ophthalmology, Hôpitaux Iris Sud Bruxelles, Bruxelles, Belgium; [4]Department of Ophthalmology, Duke University, Durham, NC; [5]Departments of Ophthalmology and Pediatrics, Flaum Eye Institute and Golisano Children's Hospital, University of Rochester, Rochester, NY; [6]College of Health Sciences Biostatistics Core Facility, University of Delaware, Newark, DE

Purpose: The retina-specific ABCA transporter, ABCA4, plays an essential role in translocating retinoids required by the visual cycle. *ABCA4* genetic variants are known to cause a wide range of inherited retinal disorders, including Stargardt disease and cone-rod dystrophy. More than 1,400 *ABCA4* missense variants have been identified; however, more than half of these remain variants of uncertain significance (VUS). The purpose of this study was to employ a predictive strategy to assess the pathogenicity of *ABCA4* variants in inherited retinal diseases using protein modeling and computational approaches.

Methods: We studied 13 clinically well-defined patients with *ABCA4* retinopathies and identified the presence of 10 missense variants, including one novel variant in the *ABCA4* gene, by next-generation sequencing (NGS). All variants were structurally analyzed using AlphaFold2 models and existing experimental structures of human ABCA4 protein. The results of these analyses were compared with patient clinical presentations to test the effectiveness of the methods employed in predicting variant pathogenicity.

Results: We conducted a phenotype–genotype comparison of 13 genetically and phenotypically well-defined retinal disease patients. The in silico protein structure analyses we employed successfully detected the deleterious effect of missense variants found in this affected patient cohort. Our study provides American College of Medical Genetics and Genomics (ACMG)–defined supporting evidence of the pathogenicity of nine missense *ABCA4* variants, aligning with the observed clinical phenotypes in this cohort.

Conclusions: In this report, we describe a systematic approach to predicting the pathogenicity of *ABCA4* variants by means of three-dimensional (3D) protein modeling and in silico structure analysis. Our results demonstrate concordance between disease severity and structural changes in protein models induced by genetic variations. Furthermore, the present study suggests that in silico protein structure analysis can be used as a predictor of pathogenicity and may facilitate the assessment of genetic VUS.

Found primarily in the outer segments of vertebrate rod and cone photoreceptor cells, ABCA4 plays an essential role in phototransduction. A large number of variants in the *ABCA4* gene are known to be associated with a wide range of autosomal recessive inherited retinal disorders (IRDs), including Stargardt disease (STGD1, OMIM 248200), cone-rod dystrophy (CORD3, OMIM 604116), fundus flavimaculatus (FFM), and retinitis pigmentosa (RP, OMIM 268000). To date, more than 3,000 variants in the *ABCA4* gene have been reported (ClinVar; accessed April 23, 2023), with *ABCA4* representing the most frequently mutated gene associated with human retinal dystrophy [1]. *ABCA4* variants include missense, nonsense, frameshift, splicing, and structural mutations, with missense variants constituting 46% [2].

The *ABCA4* gene consists of 50 exons coding for a 2,273-amino-acid polypeptide chain. The ABCA4 protein consists of two tandem-arranged halves, each consisting of six membrane-spanning α helices, a large and evolutionary-conserved extracytoplasmic domain (ECD), a cytoplasmatic nucleotide-binding domain (NBD), and a regulatory domain (RD; Figure 1). Adenosine triphosphate (ATP) binding and hydrolysis occur in the two symmetric NBDs (NBD1 and NBD2), while ECD1 and ECD2 are suggested to play a role in retinal substrate binding [3,4].

Correspondence to: Esther E. Biswas-Fiss, University of Delaware College of Health Sciences, Department of Medical and Molecular Sciences, Ammon Pinizzotto Biopharmaceutical Innovation Center 590 Avenue 1743 • Newark, DE 19713; Phone: (302) 831-2912, FAX: (302) 831-4180; email: ebiswas@udel.edu

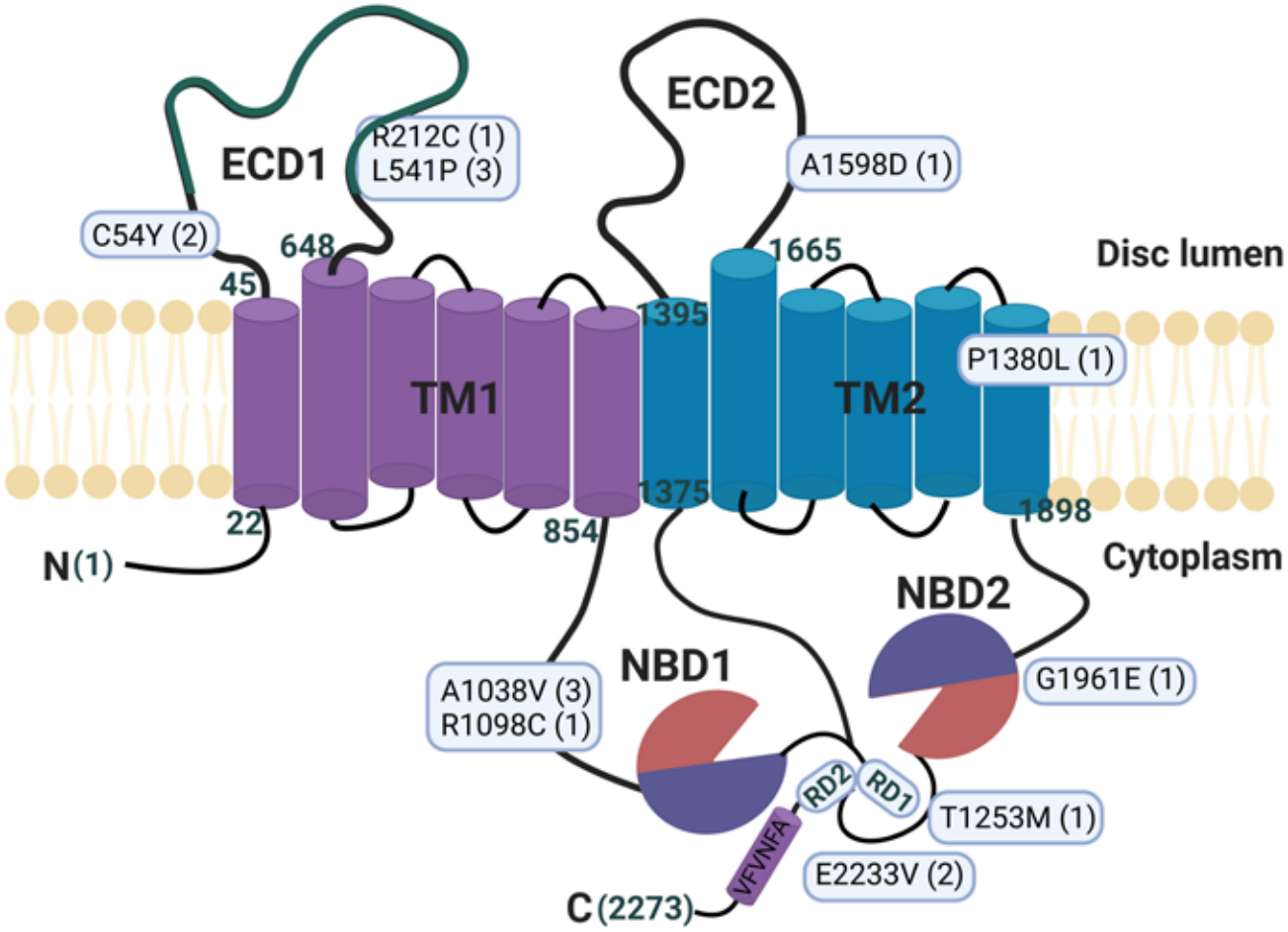

Figure 1. Topological organization of the ABCA4 protein. The pictorial representation shows the domains of ABCA4, as follows: transmembrane domain 1 (TMD1) aa 22–45/648–854, exocytoplasmic domain 1 (ECD1) aa 62–646, nucleotide-binding domain 1 (NBD1) aa 914–1152, regulatory domain 1 (RD1) aa 1153–1280, transmembrane domain 2 (TMD2) aa 1375–1395/1665–1898, exocytoplasmic domain 2 (ECD2) aa 1395–1665, nucleotide-binding domain 2 (NBD2) aa 1916–2163, and regulatory domain 2 (RD2) aa 2164–2237. Variations in our cohort are depicted in the structure. (Created with BioRender.com.).

Genetic variation databases such as ClinVar [2], Online Mendelian Inheritance in Man (OMIM) [5], the Human Gene Mutation Database (HGMD) [6], and the Leiden Open Variation Database (LOVD) [7] are widely used in clinical settings to infer a variant's phenotypic consequences. They help clinicians to predict disease prognosis and, possibly, to guide patients to suitable clinical trials when available. The jointly accepted standard terminology for interpreting the gene variants in Mendelian diseases is the five-tier classification: benign, likely benign, uncertain, likely pathogenic, and pathogenic [8]. At least half of the missense *ABCA4* variations have been designated as variants of uncertain significance (VUSs), and 12% of the remainder have conflicting interpretations (CIs; ClinVar; accessed April 23, 2023). Thus, elucidating *ABCA4* VUSs is a critical step in the progress of precision medicine in retinal diseases.

A remaining challenge in *ABCA4*-opathies is pinpointing the causative variants and assessing the combinatorial effects of biallelic variants [9-12]. Although some recurring variants of *ABCA4* have been characterized in terms of their effects on protein function and, thus, their predisposition to pathogenicity, experimental studies of the extensive and growing numbers of nonsynonymous *ABCA4* variants may be infeasible. Only a fraction of *ABCA4* variants have been functionally studied [13-20]. Genotype–phenotype correlation for *ABCA4*-related retinal degenerative disease has proved challenging due to the heterogeneity in both the phenotype and genotype and in the large number of reported sequence variations. Relatively few studies have reported on specific

gene variants that are correlated with specific phenotypes [9,17,21-24].

Thus, a robust platform is needed to predict the molecular consequences of *ABCA4* variants. In silico studies may provide a suitable predictive model for the role of *ABCA4* variants in retinal diseases. Although several pathogenicity prediction tools are available, most of them conduct their assessments based primarily on sequence features and/or evolutionary conservation and do not explain the possible functional effects at the protein level. The results of in silico studies can be enhanced using a structure-based pathogenicity analysis of *ABCA4* variants, which may lead to an improved predictive pathogenicity model.

A comprehensive study testing the reliability of three-dimensional (3D) protein structures to assess the pathogenic impact of missense variants concluded that structural damage could be accurately predicted using computational protein models [25]. Several studies have successfully used computational protein modeling to assess variations' pathogenic impact on various conditions, including developmental, neurologic, and metabolic disorders [26-33]. Our recent in silico study found concordance between predicted protein structural changes and known pathogenicity classification in a set of *ABCA4* clinical variants [34].

In this study, we further assessed the ability of computational protein structure analyses to predict the pathogenic impact of *ABCA4* variants in a phenotypically well-defined cohort of inherited retinal disease patients. Using AlphaFold2 (AF2) protein modeling [35] and available cryogenic electron microscopy (cryo-EM) structures of the human ABCA4 protein, we identified an alignment between severe clinical phenotypes and alterations in protein structure caused by *ABCA4* mutations.

METHODS

Patients: This retrospective study was conducted in accordance with Institutional Review Board approval (IRB#14–403E) of the Wills Eye Hospital and Thomas Jefferson University. A retrospective chart review was conducted for those patients with retinal disease–harboring mutations in the *ABCA4* gene. All 13 patients were seen in the Ocular Genetics Clinic at Wills Eye Hospital from August 2010 to August 2015 by an ocular genetics specialist (AVL). All patients who were found to have only a heterozygous *ABCA4* mutation or who had another suspected gene that might have modified the phenotype were excluded from this study. The medical records were reviewed, and the following data were collected: age of onset, age first seen, best-corrected visual acuity, and results of slit lamp examination and dilated fundus examination. Ophthalmic diagnostic investigations were reviewed, including fundus autofluorescence (AF), spectral-domain optical coherence tomography (SD-OCT), Goldmann visual field (GVF), and, if done, intravenous fluorescein angiography (IVFA). Full-field electroretinogram (ffERG) and multifocal electroretinogram (mfERG) were performed in accordance with the International Society of Clinical Electrophysiology and Vision standards [36]. Testing was customized for each patient as clinically indicated, based on age, visual acuity, and other factors. No clinical testing was done specifically for this study.

Genetic testing: The *ABCA4* gene was screened for variants using a range of molecular diagnostic methods, in isolation or in parallel with other genes, as determined from the phenotype and considering the patients' insurance coverage for testing (Appendix 1). When clinically indicated testing could not be performed—for example, due to a lack of insurance coverage—research testing was offered with the confirmation of detected variants performed in a Clinical Laboratory Improvement Amendments (CLIA)–certified laboratory.

Protein modeling and pathogenicity prediction: To predict the structural and functional consequences of missense *ABCA4* variants, protein structure analysis was performed on AF2-predicted protein models and the existing experimental structures of the human ABCA4 protein [4,37,38]. Full-length (FL) and domain-specific protein modeling was performed using AF2 [35] for wild-type (WT) proteins and each variant protein separately, as described previously [34]. We used AF2 through the ColabFold notebook in the Google Colaboratory [39], which is reported to have a level of accuracy that is almost identical to that of AlphaFold v2.3.1 locally for monomeric structures [36]. The default parameters on the ColabFold notebooks (AlphaFold2.ipynb and AlphaFold2_advanced.ipynb) were used, and the highest confidence models (rank 1) were selected among the generated models. The predicted WT and variant structures were refined for energy minimization using the Amber-Relax option in ColabFold, and the "repair object" feature in the FoldX plugin in YASARA View (free version: 22.5.22) [40,41]. The refinement process was repeated several times until no movement of the side chains was observed in each model. PyMOL2 (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC) was used for model visualization and structure analysis.

The structure analysis of variants was performed using two approaches. First, we compared the WT and variant AF2 models to assess 3D conformational changes, to calculate the root-mean-square deviation (RMSD) alignment scores, and to identify potential impairments in secondary

structure elements. Second, we evaluated the substitutions on the ABCA4 cryo-EM structures, considering various aspects, including stereochemical properties, binding, and molecular interactions (i.e., polar content, weak interactions, surface charge, and accessibility, cavities, and pockets). We measured the all-atom RMSD between the AF2-generated FL WT model and each variant, employing the "align" command without outlier rejection (cycles = 0) in PyMol2. We analyzed clashing interactions in available experimental structures and the refined AF models, as steric clashes resulting from the unnatural overlapping of newly positioned side chains with other residues can lead to conformational changes and structural destabilization due to a mutation.

To predict stability changes in protein variants, we used the FoldX plugin for the YASARA software. The ΔΔG free-energy change, calculated as the difference between the folded and unfolded free-energy states (ΔGmutant – ΔGwild-type), was determined using the "mutagenesis" feature of the software [40,41]. We reported the ΔΔG values obtained from all available ABCA4 experimental structures.

Furthermore, we evaluated the suitability of the AF2-predicted FL ABCA4 model for this stability analysis, by comparing the ΔΔG values per variant calculated using the AF2 model with those based on the experimentally determined structures. To facilitate this comparison, we conducted a Spearman's rank correlation using the JMP statistical software package (JMP®, Version 17, SAS Institute, Inc., Cary, NC, 1989–2021). It is important to note that in silico ΔΔG values are approximate indicators of potential stability changes rather than absolute, experimentally determined values.

For the informatics-based pathogenicity prediction of missense variants, we used Combined Annotation Dependent Depletion (CADD) [42], Polyphen-2 [43], REVEL [44], and ConSurf [45] for the evolutionary conservation of the amino acids. We also used Human Splicing Finder (HSF-Pro) Version 4 to predict the possible splicing effects of all variants in the study [46].

RESULTS

Patient characteristics: We enrolled 13 patients diagnosed with retinal dystrophy caused by variants in the *ABCA4* gene. Segregation analysis was performed for all patients to ascertain that the patients were biallelic for variants in the *ABCA4* gene. To our knowledge, the patients did not harbor any other pathogenic variants in other confounding retinopathy genes. Genotype–phenotype comparisons were conducted for the *ABCA4* variants occurring in these patients.

The clinical characteristics of the patient cohort are summarized in Table 1. The age first seen ranged from 6 to 65 years old, with a median age of 14 years old. The age of onset ranged from 6 to 15 years old, with a median age of 12 years. The best-corrected visual acuity varied from 20/30 to light perception. Two patients (A03M, A08M) had a visual acuity difference between the fellow eyes of more than two Snellen lines. In patients A02M and A05F, eccentric fixation was observed. The extraocular movements and applanation tonometry were normal in all patients. Nuclear sclerosis cataract was found in two patients (A02M and A04F).

OCT showed some degree of photoreceptor loss in 12 patients (92.3%), while 1 patient had a normal OCT. Seven patients were shown to have subretinal deposits with geographic atrophy, and five patients had only subretinal deposits. The foveal thickness in the right eye varied from 69 to 226 µm. Six patients had a normal GVF, three had a constricted visual field, and one had a central scotoma. For three patients, visual field testing was not performed. An ffERG was done in 12 patients, and the results were abnormal for 11 (91.66%). Seven patients had an abnormal response only in the photopic and four patients in both photopic and scotopic ERG. Following an isoelectric ERG, one patient had both abnormal scotopic and photopic responses. For all 12 patients who had an mfERG, abnormal results were produced. IVFA was done in 8 patients, and a silent choroid was found in 5 (62.5%).

There were 3 homozygous *ABCA4* mutation patients and 10 compound heterozygous *ABCA4* mutation patients. We identified 19 different *ABCA4* variants among the 13 patients: 3 frameshift variants, 1 nonsense variant, 5 splicing variants, and 10 missense variants (Table 2). We also identified one VUS, four CIs, and one novel *ABCA4* variant in this study, as reported in the ClinVar database (Table 3 and Table 4).

Protein modeling and structure analysis: To compare the deleterious impacts of the missense variants with the patients' disease severities, we used AF2 protein modeling and all existing human ABCA4 cryo-EM structures (Protein Data Bank identifiers [PDB IDs]: 7lkp, 7lkz, 7e7i, 7e7q, 7e7o, 7e7q, 7m1p, 7m1q). ABCA4 WT and variant models were obtained in high confidence (pLDDT > 80%) and with high structural similarity with the experimental structures (Appendix 2, Appendix 3 and Appendix 4).

Variants inside the open reading frame (ORF) can have various impacts at the protein level. Some variants have clearly devastating consequences, such as those that lead to the misfolding and mislocalization of the ABCA4 or truncating mutations, which result in the complete loss of function (LoF). However, LoF can also result from local effects

on the protein structure and stability that are more subtle and less obvious. Computational analysis aids in the evaluation of these types of effects. These types of local damage may lead to a spectrum of function loss, including affected intramolecular and ligand interactions. To interpret these types of functional damage, we used in silico protein analysis. For all missense variants, we examined the conformational changes, secondary structure elements, relative solvent accessibility, RMSD alignment scores, polar content alterations, clashing interactions, surface properties, and Gibbs ΔΔG free-energy changes. In silico analysis found a range of structural alterations in 10 missense *ABCA4* variants carried by affected individuals (Table 3, Figure 2).

Of the 10 missense variants analyzed, 9 resulted in substantial structural damage. These variants caused clashing interactions with neighboring residues (p.C54Y, p.L541P, p.P1380L, p.A1598D, and p.G1961E), disrupted secondary structure elements (p.L541P and p.T1253M), replaced a hydrophobic buried residue with a hydrophilic one (p.A1598D and p.G1961E), or affected molecular interactions (p.R212C, p.R1098C, p.P1380L, and p.E2233V). The remaining variant, p.A1038V, led to a local conformational

Table 1. Phenotypic characteristics of the patients with *ABCA4* mutations.

Pt. No./ Sex	Age of onset/ time from onset to first seen (years)	VA first seen OD-OS	Fundus findings		OCT	GVF	ffERG	mfERG	IVFA
			Flecks	Pig clumps					
A01F[a]	15 / 8	20/200 - 20/200	Mac	RPE alt	SD	Constrict	*Photopic>Scotopic	Iso	DC
A02M[a]	15 / 50	LP - LP	No	Mac	SD+GA	N/A	Iso Photopic&Scotopic	N/A	N/A
A03M	13 / 10	20/70 - 20/30	No	Mac, BB	SD	Normal	*Photopic	*Cen	No DC
A04F[b]	14 / 37	20/400 - CF	No	Diffuse	SD+GA	Normal	*Scotopic>Photopic	Iso	DC
A05F[b]	13 / 29	20/200 - 20/400	Mac	Mac	SD+GA	Constrict	*Photopic>Scotopic	Iso	No DC
A06F	8/0	20/30 - 20/25	Diffuse	CHRPE	SD	Normal	*Photopic	*Cen	N/A
A07M[c]	12/2	20/125 - 20/100	Mac	RPE alt	SD	Normal	*Photopic	Iso	N/A
A08M[c]	14 / 3	20/60 - 20/200	Mac	RPE alt	SD+GA	Constrict	Normal	*Cen	DC
A09M	6/0	20/80 - 20/200	No	No	SD+GA	N/A	*Photopic	Iso	No DC
A10M	10/0	20/100 - 20/125	Diffuse	Mac, BB	SD+GA	N/A	*Photopic	*Cen	DC
A11F	7/1	20/80 - 20/80	No	No, BB	SD	Normal	Borderline photopic	*Cen	N/A
A12F	9/2	20/125 - 20/200	Diffuse	No, BB	SD+GA	Central sco	*Photopic	*Cen	N/A
A13M	11/2	20/40 - 20/30	No	No, Bulls	Normal	Normal	N/A	*Cen	DC

[a]Consanguinity in the family.[b] A04F and A05F are siblings. [c] A07M and A08M are siblings. Abbreviation Pt: Patient, BB: "beaten bronze" appearance, CF: count fingers, CHRPE: Congenital Hypertrophy of the Retinal Pigment Epithelium, Dif: diffuse, IVFA: Intravenous fluorescein angiography, Mac: Macular, OCT: Optical coherence tomography, GA: geographic atrophy, SD: subretinal deposits, N/A: not applicable, LP: Light perception GVF: Goldmann visual field, Constrict: constrict visual field, Central sco: Central scotoma, ffERG: full-field electroretinography, *P: Abnormal photopic, *Photopic>Scotopic: Abnormal photopic more affected than scotopic, IsoPhotopic&Scotopic: Isoelectric photopic and scotopic, *Scotopic>Photopic: Abnormal scotopic more affected than photopic, mfERG: multifocal electroretinography, *Central: Abnormal central part of mfERG, Iso: isoelectric mfERG.

Table 2. Genotypic characteristics of the patients with *ABCA4* mutations.

Patient No./ Sex	Allele 1		Allele 2	
	ABCA4 Variation	Location (Domain)	ABCA4 Variation	Location (Domain)
A01F[a]	c.4793C>A:p(A1598D)	ECD2	c.4793C>A:p(A1598D)	ECD2
A02M[a]	c.5018+2 T>C	sECD2	c.5018+2 T>C	sECD2
A03M	c.5714+5G>A	sNBD2	c.3758C>T:p(T1253M);c.5882G>A:p(G1961E)	RD1, NBD2
A04F[b]	c.161G>A:p(C54Y)	ECD1	*c.6698A>T:p(E2233V)	RD2
A05F[b]	c.161G>A:p(C54Y)	ECD1	*c.6698A>T:p(E2233V)	RD2
A06F	c.1622T>C:p(L541P);c.3113C>T:p(A1038V)	ECD1, NBD1	c.4234C>T:p(Q1412*)	ECD2
A07M[c]	c.1622T>C:p(L541P);c.3113C>T:p(A1038V)	ECD1, NBD1	c.5714+5G>A	sTMD2
A08M[c]	c.1622T>C:p(L541P);c.3113C>T:p(A1038V)	ECD1, NBD1	c.5714+5G>A	sTMD2
A09M	c.634C>T:p(R212C)	ECD1	c.768G>T:p(V256=)	ECD1
A10M	c.4139C>T:p(P1380L)	TMD2	c.4139C>T:p(P1380L)	TMD2
A11F	c.4539+2028C>T	sECD2	c.859-13T>C	sECD1
A12F	c.850_857delATTCAAGA:p(I284fs)	ECD1	c.6184_6187delGTCT:p(V2062fs)	NBD2
A13M	c.3292C>T:p(R1098C)	NBD1	c.664delG:p(A222fs)	ECD1

[a]Consanguinity in the family.[b] A007F and A008F are siblings. [c] A014M and A015M are siblings *Novel variant, FS – Frameshift. NBD: Nucleotide Binding domain, ECD: Exocytoplasmic Domain, TDM: Transmembrane Domain, RD: Regulatory Domain. Reference genome assembly, GRCh38:Chr1:83457325–104273917, Reference Transcript: NM_000350.3

change in the backbone and reduced stability (Figure 2, Table 3). The stability analysis suggested that 6 of the missense variants (p.C54Y, p.L541P, p.A1038V, p.P1380L, p.A1598D, and p.G1961E) were particularly destabilizing with ΔΔG > 5 (Table 3 , Appendix 4).

We examined how the AF2 model determined ΔΔG values compared to those calculated using the ABCA4 cryo-EM structures. The ΔΔG profiles over 10 missense variants (and 2 complex alleles) were formed to compare these measures. Seven values were missing because the Arg-212 residue was unavailable in the 7lkp, 7m1p, and 7lkz structures, and the Gly-1961 and Glu-2233 residues were absent in the 7mlq structure. Appendix 5 shows agreement in the relative ΔΔG values for each missense variant among the experimental structures and the AF2 model. We formalized the similarity seen with a pairwise correlation analysis using Spearman's rank correlation. Unadjusted p values showed 26 of 28 possible paired correlations to be significant (maximum p value = 0.0114); the ΔΔGs calculated based on AF2 specifically correlated with those done on the experimentally determined structures at approximately 0.89 or above for all except 7lkz (ρ = 0.75; Appendix 6). Based on these findings, ΔΔG predictions from the AF2 model can be considered an alternative to those obtained from experimentally determined structures.

In silico protein structure analysis and phenotypic association: The missense variants were distributed across the ABCA4 protein sequence: there were 3 in ECD1, 3 in NBD1, 1 in ECD2, 1 in transmembrane domain-2 (TMD2), and 2 in NBD2 (Figure 1, Table 3). The variants identified in our cohort are described in terms of the clinical phenotype and assessment of the impact of the variant on protein structure.

The p.C54Y variant was identified in severely affected siblings, A04F and A05F, with ages of onset of 14 and 13 years, respectively. The two patients had the same biallelic combination, with p.E2233V on the opposite allele (in trans). The visual acuity of these patients was poor, ranging from 20/200 to 20/400. The predicted protein model and experimental structures showed disulfide bond breakage and clashing interactions (Figure 2A), along with a poor alignment score (RMSD = 1.26) and destabilization as reflected by the ΔΔG (+ 41.68 kcal/mol), supporting the deleterious effect of this variant (Table 3). The informatics tools also predicted the variant to be pathogenic (Table 3, Supplementary Table S2). The observed severity of this variant, based on the

Table 3. Structure analysis and overall pathogenicity assessments of the missense *ABCA4* variants found in the study.

Domain	ABCA4 variants	Clinical significance	Functional in vitro studies	Pathogenicity prediction	RMSD	In silico ΔΔG exp. (kcal/mol)	Protein structural changes
ECD1	c.161G>A:p(C54Y)	P/LP	-	P	1.263	41.7	Broken disulfide bond, clashes
	c .634C>T:p(R212C)	P/LP	↓ Expression[20, 62], ↓ATP binding[20], Basal&Stim. ATPase[62]	P	0.971	1.26	Broken H bond
	c .1622T>C:p(L541P)	CI	Mislocalization, ¯ Basal&Stim. ATPase[17,20]	P	1.271	7.65	Buried Pro introduced, broken helix
NBD1	c.3113C>T:p(A1038V)	P/LP	↓ Basal&Stim. ATPase[17], Normal subcellular localization[23]	B	1.248	6.6	Slight conformation change
	c .3292C>T:p(R1098C)	P/LP	-	P	1.217	2.49	Lost salt bridge with Asp-2242
RD1	c.3758C>T:p(T1253M)	CI	-	P	1.392	1.54	Premature B-sheet, loss of H-bond
TMD2	c.4139C>T:p(P1380L)	P/LP	↓ Basal&Stim. ATPase [19]	P	1.264	12.2	Clashes, loss of the Pro-induced kink
ECD2	c.4793C>A:p(A1598D)	P/LP	↓ Expression, ↓ Basal&Stim. ATPase [62]	P‡	1.197	13.56	Clashes, buried hydrophilic introduced
NBD2	c.5882G>A:p(G1961E)	CI	↓ Basal&Stim. ATPase [17,20]	P	0.985	31.68	Clashes, buried hydrophilic introduced
RD2	c.6698A>T:p(E2233V)	Novel	-	P	1.35	0.71	Loss of H-bond: RD2-NBD1 interaction *

Reference genome assembly, GRCh38:Chr1:83457325–104273917, Reference Transcript: NM_000350.3. Informatics-based pathogenicity prediction is based on CADD, PolyPhen-2, REVEL, and Consurf amino acid conservation scores [45]. ǂ When there is no absolute consensus, given based on the majority's prediction. (See supp. Table 2). *The novel variant c.6698A>T:p(E2233V) was also predicted to impact splicing with significant alteration of auxiliary sequences ESE / ESS motifs ratio (−10; HSF-Pro [46]). The wild-type (WT) and variant protein models were predicted in the AlphaFold2 [35] through the ColabFold AlphaFold2_advanced notebook running in Google Colaboratory [39]. The in silico stability change prediction is based on the mean of ΔΔGs calculated from all ABCA4 experimental structures (exp.) using the FoldX plugin for Yasara [40,41]. ΔΔG is the measure of the free energy change due to an amino acid substitution. ΔΔG ≥1, the substitution is defined as destabilizing. For ΔΔG values based on each protein structure, and their correlations, see supplementary files: Figure S3 and Table S3. All other structural analyses were performed in PyMOL2. Root-mean square displacement (RMSD; all-atom) is calculated to determine the structural alignment scores. VUS: Variants of Uncertain significance, CI: Conflicting interpretation, NBD: nucleotide-binding domain, ECD: Exocytoplasmic Domain, TMD: Transmembrane Domain, RD: Regulatory Domain.

patient's phenotype (Table 1), aligned well with the protein structural change.

Variant p.R212C was identified in a 6-year-old patient (A09M) with a rapidly progressing disease. The patient's visual acuity at the time of the study ranged from 20/80 to 20/200, with constricted visual fields. OCT showed subretinal deposits and geographic atrophy (Table 1). The p.R212C variant was found in trans with the c.768G>T:p(V256=) synonymous variant that was predicted to be pathogenic due to its likely effect on splicing (HSF-Pro; Table 4) [46]. Given that one allele was predicted to be completely compromised (due to coding for a severely truncated polypeptide), the pathogenicity level of the p.R212C allele was expected to be directly related to the patient's disease severity. The p.R212C variant was predicted to be pathogenic by the informatics

Table 4. Putative Loss of Function (pLoF) and intronic *ABCA4* variants in the study.

ABCA4 pLOF and intronic variants	Predicted effect	Predicted effect on protein	ClinVar Asses.
c.664delG	New Acceptor site▪	p.A222Qfs*19	P
c.768G>T	Splicing (Exon-intron boundary) ▪	p.V256=fs*12	P/LP
c.850_857delATTCAAGA	Splicing (Broken WT donor site) ▪	p.I284Vfs*33	P
c.859-13T>C	No impact on splicing▪	?	VUS
c.4234C>T	Nonsense	p.Q1412*	P
c.4539+2028C>T	No impact on splicing▪ 345-nt pseudoexon insertion	? p.R1514Lfs*36[56]	CI
c.5018+2T>C	Splicing (Broken WT donor site) ▪	p.T1675Vfs*30	P/LP
c.5714+5G>A	Splicing (Broken WT donor site) ▪	Skipping of exon 40[65]	P/LP
c.6184_6187delGTCT	Frameshift	p.V2062Tfs*51	P

▪HSF-Pro: Human Splicing Finder, version 4 [46], P: Pathogenic, LP: Likely Pathogenic, VUS: Variants of Uncertain Significance, CI: Conflicting interpretation, NP: Not provided.

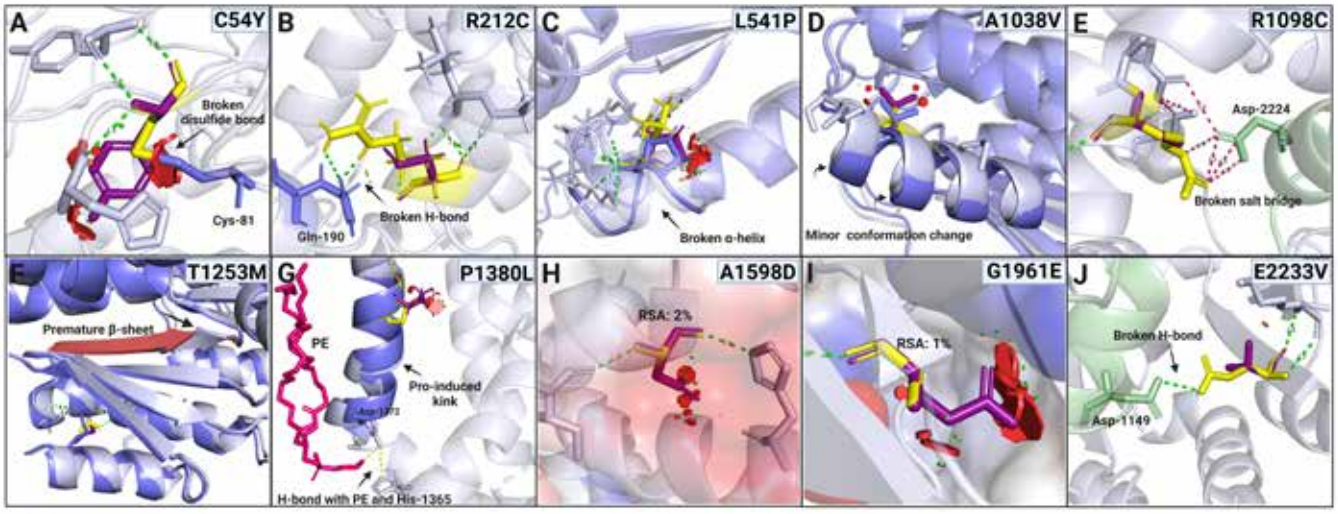

Figure 2. In silico structure analysis of the missense *ABCA4* variants. The variant models (blue) are superimposed and compared with the wild-type (WT) model or experimental structures (gray). The WT residues are shown as yellow sticks, and substitutions are shown as purple. The red discs show van der Waals (vdW) overlaps or steric clashing with the substitution. They are given only when every possible side-chain conformation resulted in clashes in experimental and predicted structures, but only the highest possibility conformation is shown here. A. C54Y results in disulfide bond breakage and clashing interactions. B. R212C leads to an intra-domain H-bond breakage. C. L541P disrupts an α-helix and produces steric clashes. D. A1038V causes slight conformational change. E. R1098C causes the loss of a salt bridge affecting NBD1–RD2 domain–domain interaction; the RD2 is colored light green. F. T1253M inside the RD1 leads to a loss of the H-bond and a premature β-sheet. G. P1380L results in clashing interactions with neighboring residues as well as changes in the swivel angle of the transmembrane helix and indirectly breaks H-bonds with His-1365 and phosphatidylethanolamine (PE; shown in pink). This proline-induced kink in the WT transmembrane domain may have a functional role in substrate transport across the membrane. The in silico variant (blue) model is superposed onto the cryo-EM structure of the human ABCA4: 7e7o (gray) to show the distortion. H. A1598D results in steric clashes and introduces a buried hydrophilic residue with an expected destabilizing effect. I. G1961E causes clashing interactions and replaces a buried hydrophobic residue with a hydrophilic amino acid. J. The E2233V novel variant in the RD2 results in H-bond breakage with a residue in the NBD1 domain (light green). Visualized and analyzed in PyMOL2.

tools (Supplementary Table S2), and structure analysis revealed a missing H-bond with Gln-190 due to the substitution (Figure 2B, Table 3). However, p.R212C is in the part of the ECD1 domain where the experimental structures have not been well resolved. This residue might be important for the structural flexibility of the domain.

Complex allele p.L541P/p.A1038V was detected in 3 patients (A06F, A07M, and A08M) in our study. The siblings A07M and A08M had the same biallelic combination with the c.5714+5G>A splicing variant on the trans allele, while the third patient had a p.Q1412* nonsense mutation in trans (Table 2). All three patients exhibited fundus flecks, subretinal deposits, and pigmentary clumps. The age of onset, visual acuity, and other findings were comparable in the siblings (Table 1). Protein structure analysis revealed that p.L541P had a more deleterious impact than did p.A1038V by causing significant impairment to the protein structure. This substitution is located within an α-helical region and is predicted to disrupt the α-helix structure and cause clashing interactions, leading to conformational changes (RMSD = 1.27; Figure 2C, Table 3). Although we did not observe a noticeable structural difference between the WT and p.A1038V variant, it showed an overall poor alignment score (RMSD = 1.25) and ΔΔG (+ 6.56 kcal/mol), indicating that it can be a destabilizing mutation (Figure 2D). The informatics-based tools suggested that the p.L541P variant is pathogenic while p.A1038V is benign (Table S2). For the complex allele L541P/A1038V, the computational analysis of the variant on the ABCA4 protein structure aligned well with patient phenotype.

Variant p.R1098C was found in patient A13M, who had an age of onset of 11. It was in trans with a frameshift mutation (Ala222Glnfs*19), which leads to a complete loss of function. The Arg1098 residue, localized inside the NBD1, was found to have an essential role in the NBD1–RD2 interdomain interaction by forming a salt bridge, which is disrupted by the Cys substitution (Figure 2E). Prediction tools classified p.R1098C as pathogenic (Table 3, Supplementary Table S2). For this variant, the computational analysis of the effects on the ABCA4 protein structure coincided with the disease phenotype.

Complex allele p.T1253M/p.G1961E was identified in patient A03M, with the c.5714+5G>A splicing variant in trans, which is predicted to affect the splicing pattern according to HSF-Pro [46]. The patient presented with a "beaten bronze" appearance in fundus imaging, with lipofuscin deposits centrally and a mild surrounding ring of hyperpigmentation (Table 1). Remarkably, 10 years after the diagnosis, the patient's vision remained relatively preserved compared to the other patients in our cohort. In silico analysis predicted that both missense variants individually cause protein structural changes. The p.T1253M variant, located within RD1, led to a major structural alteration with a degenerated secondary structure element (Figure 2F). While, the p.G1961E variant replaced a hydrophobic buried residue with a hydrophilic amino acid, which was predicted to destabilize the structure, as indicated by the ΔΔG value (Table 3) and as shown in Figure 2I. The prediction tools unanimously supported the pathogenicity of both variants (Supplementary Table S2).

The p.P1380L variant was identified in a 10-year-old homozygous patient (A010M) with a highly progressive *ABCA4* retinopathy. The patient's severe phenotypic findings (Table 1) are consistent with protein modeling that indicates the Pro → Leu substitution within the transmembrane domain has significant structural consequences. The substitution causes steric clashes with neighboring residues in all possible positions, leading to altered protein conformation (RMSD = 1.26). Additionally, the variant protein model superimposed with the substrate-bound cryo-EM structure suggests that this proline residue likely plays a role in a kinked transmembrane helix, with implications of a functional role in substrate translocation (Figure 2G).

The p.A1598D variant, located in the ECD2 domain, was identified in patient A01F, who presented with macular flecks and alterations in the retinal pigment epithelium (RPE) subretinal deposits, with a visual acuity of 20/200 in both eyes. The patient was homozygous for the p.A1598D variant, with no other *ABCA4* variants detected. In silico analysis revealed that by introducing a hydrophilic residue to a buried location, the p.A1598D variant resulted in destabilization (ΔΔG = + 13.56 kcal/mol) and caused steric clashes (Figure 2H). However, the conventional pathogenicity assessment of this variant resulted in inconclusive predictions. PolyPhen-2 and REVEL both classified it as possibly damaging, with scores of 0.57 and 0.685, respectively. In contrast, the CADD score of 16.3 was below the threshold of 20, indicating a lower likelihood of pathogenicity. Ala-1598 is not evolutionarily conserved [45], but the variant has a low allele frequency (2.63×10^{-5}; Supplementary Table S2) [47]. Overall, the patient's severe phenotype was consistent with the substantial changes observed in the protein structure predicted by structure analysis.

Variant p.E2233V was identified in siblings A04F and A05F, who had comparable age of onset and visual acuity (Table 1). A04F presented with large retinal pigmentary clumps involving the macula and mild periphery, choroidal attenuation, but no fleck lipofuscin deposits. A05F showed geographic atrophy in the macular area with mid-periphery pigmentary clumping, with several subretinal flecks in the

macula. A grayish elevation in the superior macula could represent either a choroidal nevus or subretinal neovascular membrane. Both patients had relatively healthy optic nerves, with minimal pallor as well as mild retinal vascular attenuation (Table 1). This novel variant is located in RD2. Computational analysis revealed a loss of an H-bond with a residue in the NBD1 domain, significantly impairing the protein structure (RMSD = 1.35; Figure 2J). This variant was also predicted to affect splicing by altering the auxiliary sequences' ESE/ESS motifs ratio (−10; HSF-Pro [46]). Although the true consequence of this variant is not yet fully understood, we saw an agreement between the computational analysis of the missense effects on the ABCA4 protein structure and the patient phenotype.

Our cohort had three patients who did not have a missense *ABCA4* variation. Patient A02M carried the c.5018+2T>C variant in a homozygous state. This intronic variant is predicted to affect the splicing pattern by causing a broken WT splice site (HSF-Pro [46],), resulting in translated product lacking the last quarter of the protein. Patient A11F was compound heterozygous with the c.859–13T>C and c.4539+2028C>T intronic variants, present in the ClinVar database as a CI and VUS, respectively. This variant was also predicted to affect splicing by altering the ratio of Exonic Splicing Enhancer (ESE) to Exonic Splicing Silencer (ESS) motifs, with a score of −10 as predicted by the HSF-Pro [46]. The c.4539+2028C>T deep intronic variant has previously been shown to cause 345-nt pseudoexon insertion with a likely effect on the protein level (p.R1514fs) [48]. Unless there are other undetermined variations in the patient, the phenotype suggests that these intronic variants are pathogenic. Patient A12F carried two different frameshift variants—p.I284Vfs*33 and p.V2062Tfs*51—both with high levels of pathogenicity.

DISCUSSION

The *ABCA4* gene, with more than 3,000 reported variants, is known to be associated with a wide range of autosomal recessive inherited retinal disorders. Unfortunately, many of these variants remain uncharacterized in terms of their disease propensity, making it difficult to determine disease prognosis when detected in patients. With the recent advancement of computational tools such as pathogenicity prediction software and protein modeling platforms, it is now possible to carry out detailed in silico investigations of specific variants and to correlate them with available clinical findings, even in the absence of an in vitro functional analysis. In this study, we investigated the use of the in silico analysis of the ABCA4 protein as a tool to augment the prediction of disease severity and the clinical significance of sequence variants.

To evaluate this paradigm, we performed a patient-based study of individuals with confirmed *ABCA4* retinopathy. Subsequently, we performed a computational protein structure analysis to see how well the predicted structural changes aligned with the observed pathogenicity of the missense variants based on the patient phenotypes. The phenotypic outcome of *ABCA4* retinopathies depends on the collective effect of the specific variants in the patient [12,49,50]. Therefore, we included all types of variants with their predicted effects when discussing the relation between a patient's genotype and phenotype. Furthermore, we applied a combinatorial approach to increase accuracy in predicting pathogenicity and explaining the pathogenicity of disease-associated variants, as no single type of analysis is sufficient to explain the deleterious effect of a genetic variant at the protein level (Figure 3).

We used AF2, a highly reliable way to predict protein structures [35,51,52], along with recently resolved experimental structures of ABCA4 to analyze the structural impacts of the variants. In addition to physically observable conformational changes, we evaluated bonding, interaction, surface characteristics, and protein stability in the 3D protein models to gain insight into the effects of variants on the ABCA4 protein. We used two numerical measures: the RMSD calculation to compare the conformational alignment scores between the WT and mutant predicted protein structures, and the ΔΔG calculation to compare thermodynamic stability changes upon amino acid substitution. We observed the consistency between the in silico ΔΔG values determined by the AF2 model and the ABCA4 cryo-EM structures (Supplementary Figure S3, Supplementary Table S3). Our findings are in line with a recent study by Akdel et al. [52], supporting the notion that AF2 models can serve as a reliable alternative for predicting the destabilization effects of variants based on protein structure, which can be particularly helpful for proteins with no available experimental structure.

In the present study, we used manual approaches to detect structural alterations; however, with the help of this initial study and with larger data sets, more sophisticated, automated, or machine learning–based strategies such as support vector machines (SVMs) can be used. It should be noted that although the variants alter the protein structure, it does not follow that they will necessarily alter its function [53]. We also acknowledge that the structural changes we mention here are insufficient to explain all the in vivo effects and all aspects of the mutation consequences, including mRNA-level and post-transcriptional alterations.

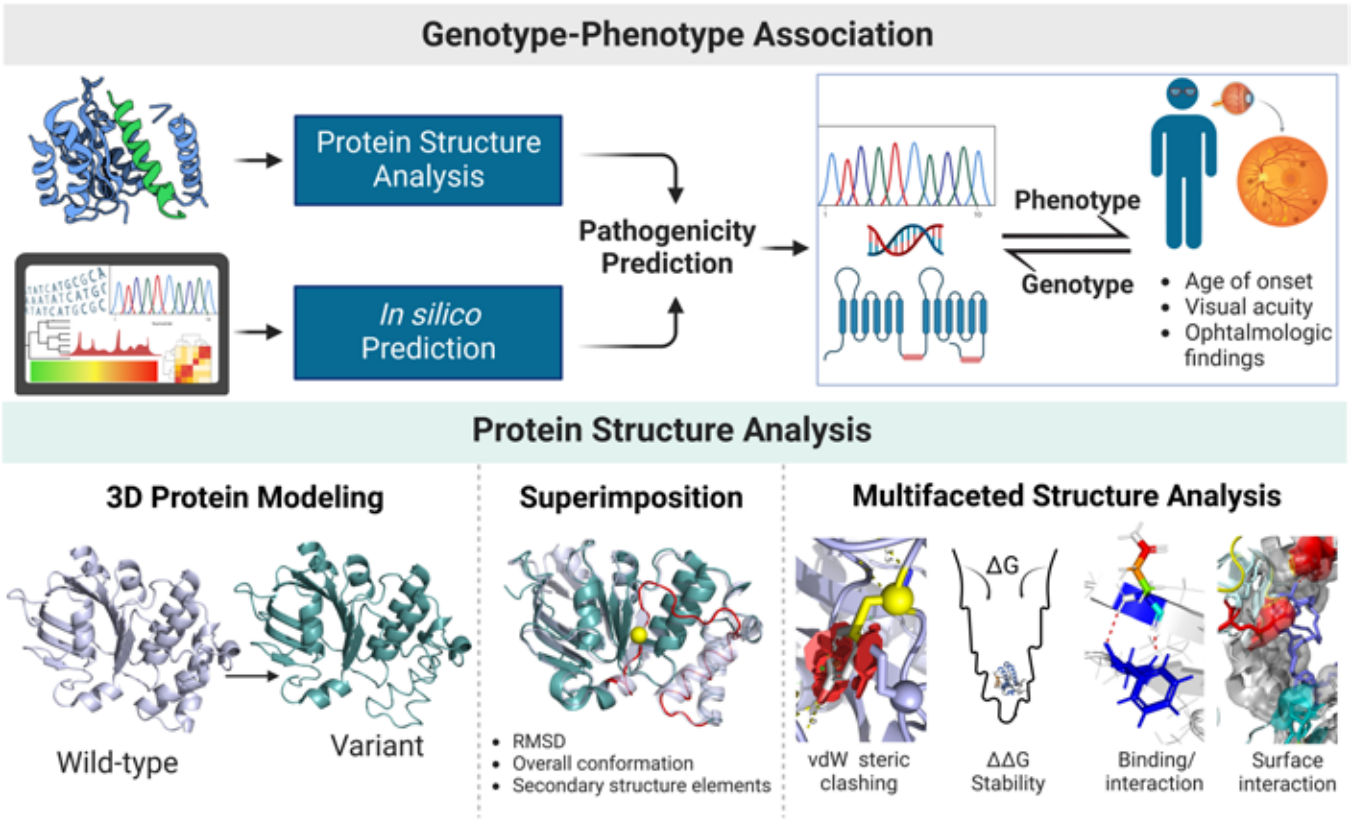

Figure 3. General workflow of the study. The genetic and phenotypic information collected from a well-defined retrospective cohort of *ABCA4* retinopathy patients provided a list of *ABCA4* missense variants, allowing us to test the protein structure–based computational pathogenicity prediction platform. Our analysis methods included physically comparing WT and mutant proteins based on observable structural changes between the AF2 WT and variant models, evaluating thermodynamic stability, examining changes in inter- and intramolecular binding/interaction, surface properties, and solvent accessibility, mainly using experimental structures. In conjunction with other in silico pathogenicity prediction methods, these analyses can aid in understanding the genotype–phenotype association in *ABCA4*-related inherited retinal diseases and in classifying VUS.

Functionality and localization analyses can be performed to support the findings of this study.

According to American College of Medical Genetics and Genomics / Association for Molecular Pathology (ACMG/AMP) guidelines [8], being detected in the trans to another pathogenic variant provides at least a moderate level of evidence of pathogenicity (PM3) for genetic variants in recessive disorders. Six missense variants in our cohort (c.161G>A:p(C54Y), c.634C>T:p(R212C), c.3292C>T:p(R1098C), c.4139C>T:p(P1380L), c.4793C>A:p(A1598D), and c.6698A>T:p(E2233V)) were known to be alone in the corresponding allele (i.e., no other variants in cis; Figure 4), and all of them were structurally damaging (Figure 2). Our findings from the protein structure analysis support the expected pathogenicity of these six missense variants identified in trans to a pathogenic allele in affected patients of a recessive disorder.

In addition to being identified in the affected individuals who presented with severe phenotypes, six of the missense variants in our study (p.R212C, p.L541P, p.A1038V, p.P1380L, p.A1598D, and p.G1961E) have previously been characterized in functional studies [17,19,20,23,54] and presented with impairments in protein expression, localization, or enzymatic function (PS3; Table 3, Figure 4); this provided an excellent opportunity to assess the employed in silico approach. We found severe structural defects and destabilizing effects in relation to the variants p.L541P, p.P1380L, p.A1598D, and p.G1961E (Figure 2, Table 3), consistent with previous functional studies. Specifically, p.L541P and p.P1380L exhibited mislocalization and a significant reduction in basal and retinal stimulated ATPase activities [17,19]. The variants p.A1598D and p.G1961E, in contrast, showed almost a complete loss of retinal-stimulated ATPase activity [17,20]. These findings

suggest that in silico protein structure analysis can predict the functional effects of *ABCA4* missense variants.

We had one novel variant, c.6698A>T:p(E2233V), in our cohort. We used the ACMG and AMP guidelines, which provide a series of standard criteria for classifying the potential pathogenicity of variants [8]. Our study provided different degrees of pathogenicity evidence for nine missense *ABCA4* variants, including this novel variant (Figure 4). The p.E2233V variant was found in affected siblings in our cohort (supporting evidence of pathogenicity: PP1 and PP4, ACMG/AMP Guidelines). It was alone on the allele and in trans to another pathogenic variant (PM3). The standard informatics tools predicted that its effect is probably damaging (Supplementary Table S2), and protein modeling demonstrated substantial impairment in the structure (PP3). We determined the p.E2233V variant as likely pathogenic based on the supporting evidence, and we deposited it into the ClinVar database (Accession: SCV002526665.1).

Identifying disease-causing variants in inherited diseases is undoubtedly crucial in clinical management, including

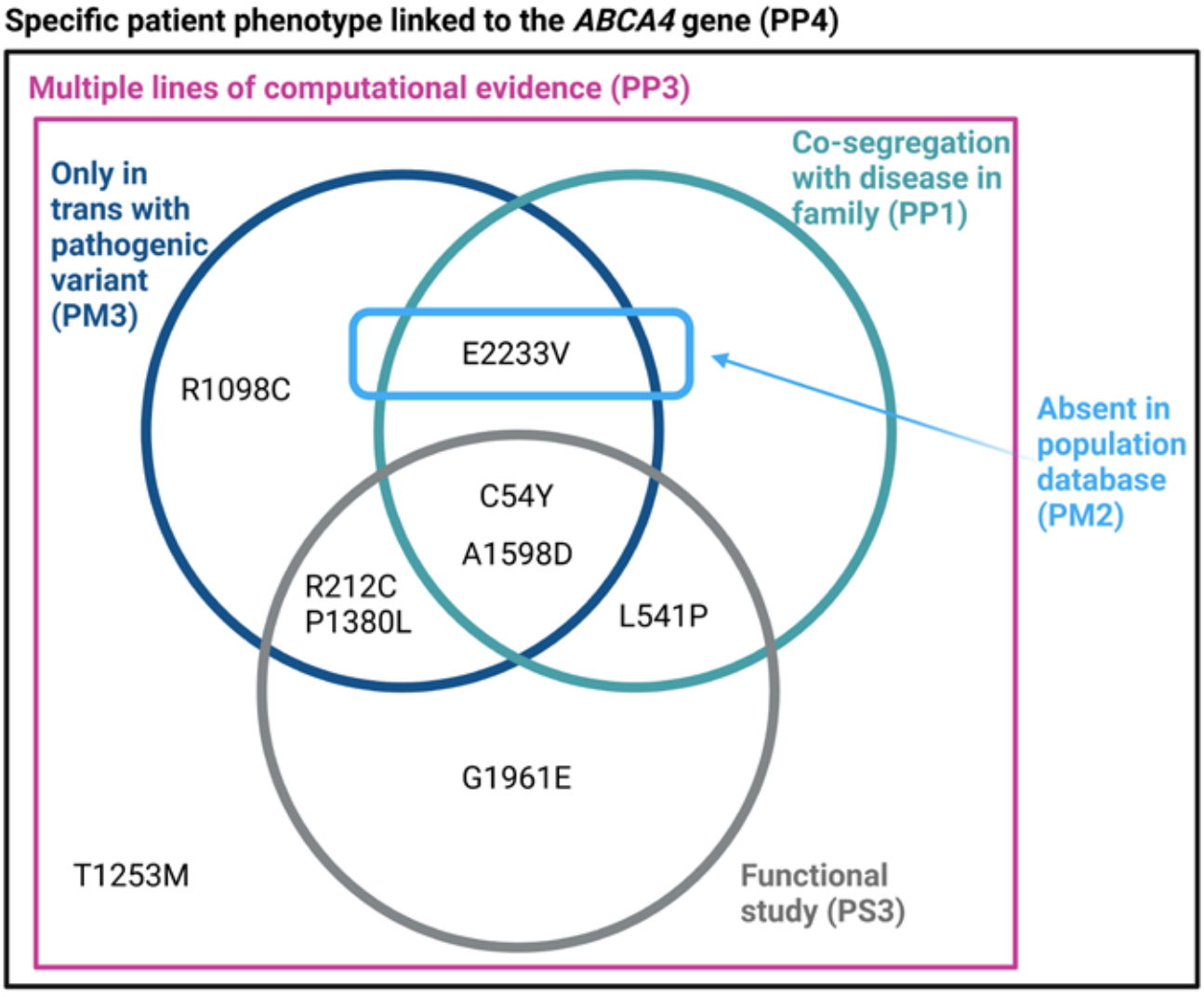

Figure 4. Pathogenicity evidence of the *ABCA4* missense variants in the study following the American College of Medical Genetics and Genomics / Association for Molecular Pathology (ACMG/AMP) guidelines. Our study provides varying degrees of moderate and supporting evidence on nine missense *ABCA4* variants. According to the latest ACMG/AMP guidelines, multiple lines of computational evidence provide a third degree of supporting evidence. In addition, the genotypic and phenotypic information obtained from our retrospective patient cohort provided further evidence of pathogenicity. Note that the functional evidence given here is derived from a literature search and is cited in Table 3.

future clinical trials. In our study, three missense *ABCA4* variants had conflicting interpretations in ClinVar. While p.L541P was one of them, it can be considered pathogenic/likely pathogenic, as the conflict arose from only 1 submission with uncertain significance (ClinVar submissions: 13 pathogenic, 2 likely pathogenic, and 1 uncertain significance). The other variants with conflicting interpretations were p.T1253M and p.G1961E, which were found together as a complex allele in patient A03M in our cohort, precluding allelic assessment. However, in silico analysis predicted both variants to be structurally defective, and conventional prediction tools suggested that both were likely pathogenic (PP3). Furthermore, patient A03M carried this complex allele in trans to the pathogenic c.5714+5G>A splicing variant, which was also found in two other patients in our cohort in trans with the p.L541P/p.A1038V complex allele (A07M and A08M), who presented with poorer outcomes (Table 1). These findings suggest that the p.L541P/p.A1038V complex allele may have a worse clinical outcome than does the p.T1253M/p.G1961E complex allele.

Homozygous patients offer a unique opportunity to investigate the genotype–phenotype relationship in autosomal recessive diseases. In our study, we had three homozygous patients. The p.P1380L biallelic variant was found with early onset (i.e., 10 years old) disease presenting with the rapid deterioration of sight. The patient showed a "beaten bronze" appearance on examination, with diffuse flecks and pigmentary clumps. OCT showed subretinal deposits and geographic atrophy. In silico analysis found that the p.P1380L caused severe structural alterations, including conformational change, reduced stability, and likely intermolecular interactions (Table 3, Figure 2G). Biallelic p.A1598D led to a mid-adolescence onset but a progressive disease phenotype with 20/200 visual acuity in both eyes, presenting with flecks and subretinal deposits in the macular area. Similarly, this variant led to decreased protein stability, as indicated by the structure analysis and the ΔΔG value (Table 3, Figure 2H). The last biallelic variant was the intronic c.5018+2T>C variant, found in a patient with onset at 15 years old. Over a disease duration of 50 years, the patient lost his vision to the degree of light perception (Table 1). This variant was predicted to affect splicing by the HSF-Pro web server (Table 4) [46].

In this study, we illustrated the predicted structural consequences of genetic variations found in patients with *ABCA4*-retinopathies. It is promising to see that the protein modeling and in silico analysis used in this study successfully identified structural defects in variants found in affected individuals, thus allowing for the identification of variants that were expected to be pathogenic. In our recent study, we also demonstrated the accuracy of this in silico approach by assessing known pathogenic and known benign *ABCA4* missense variants retrieved from clinical databases [34]. However, it is important to test this in silico approach on a large number of variants found in patients with diverse clinical presentations to observe whether the predicted structural changes are proportional to disease severity. As a transporter protein, ABCA4's function is closely related to its dynamic properties [55,56], and some sequence variants can impact the dynamic properties of the protein. In this study, we did not assess these potential molecular dynamics–related impacts. Also, we did not conduct deep intronic or *ABCA4* promoter sequencing, so the effects of these potential variants, if present, in cis with the known variants as well as the effect of variants in other genes remains unknown.

Our study shows that the types of structural damage caused by the *ABCA4* variants found in inherited retinal disease patients are consistent with their affected phenotypes, the assessments of genetic variation databases, previous functional studies, and, for the most part, conventional in silico pathogenicity prediction programs. Our computational protein structure method of analysis helps predict the functional roles of individual amino acids in the ABCA4 protein and provides possible explanations for the pathogenesis of missense mutations. The promising results of this small-scale study suggest that this methodology can be extended to larger and more diverse cohorts. Additional studies with biochemical and molecular focuses can be conducted to better understand all aspects of the variants studied here. We conclude from the results of this study that computational protein structure analysis may be a helpful adjunct in predicting the pathogenicity of *ABCA4* VUS and in disease prognosticating.

Conclusion: In silico methods are essential elements of biomedical research. Computational protein structure analysis is a practical tool used to understand protein structure and function, as well as to infer the pathogenic impacts of sequence variations. This study illustrated the likely structural consequences of 10 missense variants. Our study demonstrated that a disease severity that is closely defined by the age of onset and by visual acuity is well associated with in silico–predicted protein structural changes caused by missense *ABCA4* variants. Determining disease-causing variants is a crucial part of clinical management, and we believe that our research will contribute to elucidating the molecular pathology of variants associated with *ABCA4* retinal diseases.

APPENDIX 1. SUPPLEMENTARY TABLE S1. GENETIC TESTING FOR PATIENTS IN THIS STUDY.

To access the data, click or select the words "Appendix 1." CEI: Casey Eye Institute Molecular Diagnostics Laboratory, BCM: Medical Genetics Laboratories at Baylor College of Medicine, Denver: Denver Genetic laboratories, Emory: Emory Genetics Laboratories, arRP: autosomal recessive Retinitis Pigmentosa, RDS: Peripherin 2 (PRPH2), ELOVL4: Elongation of Very Long Chain Fatty Acids-Like 4.

APPENDIX 2. SUPPLEMENTARY FIGURE S1.

To access the data, click or select the words "Appendix 2." Comparison of the Alphafold2 (AF2) generated WT full-length and domain-specific ABCA4 models with the experimental structures. A. AF2-predicted full-length (FL) ABCA4 protein model colored based on the pLDDT score. B. Superimposition of AF2 domain-specific models ECD1 (green), TMD1 (yellow), NBD1 (cyan), R1 (purple), ECD2 (magenta), TMD2 (blue), NBD2 (salmon), R2 (pink) on human ABCA4 cryo-EM structures (gray), demonstrating the high structural similarity between AF2-predicted models and the available cryo-EM structures of human ABCA4 protein. Visualized in PyMol2.

APPENDIX 3. SUPPLEMENTARY FIGURE S2.

To access the data, click or select the words "Appendix 3." Plot showing the residual pLDDT quality scores of the AlphaFold2-generated ABCA4 WT and variant protein models, indicating the confidence of the modeling. The overall pLDDT score for each model is given in Supplementary Table S2.

APPENDIX 4. SUPPLEMENTARY TABLE S2.

To access the data, click or select the words "Appendix 4." Detailed structure analysis and pathogenicity assessments of the missense ABCA4 variants in the study. Reference genome assembly: GRCh38:Chr1:83457325-104273917. Reference Transcript: NM_000350.3. ConSurf Score: A measure of amino acid evolutionary conservation based on the ConSurf web server analysis. The score ranges from 1-9 with increasing conservation [45]. Allele frequencies: Based on The Genome Aggregation Database (gnomAD) v3.1.2 [47]. AF2 Residual pLDDT: Based on the full-length WT ABCA4 AF2 model and corresponding WT residue. FL-Variant pLDDT: For the entire full-length ABCA4 variant AF2 models. RMSD (Å): The root-mean-square deviation. Based on the structural comparison of the full-length AF2 predicted variant and WT ABCA4 models. Destabilizing variants: Predicted using Gibbs free energy change calculation in the FoldX plugin for YASARA [40,41]. Side chain steric clashes: Reported only when all possible rotamers lead to clashing interactions. ECD (Exocytoplasmic domain), NBD (Nucleotide-binding domain), TMD (Transmembrane domain), and IH (intracellular α-helix). PDB IDs of the cryo-EM structures of the human ABCA4 used in the analyses: 7lkp, 7lkz, 7e7i, 7e7q, 7e7o, 7e7q, 7m1p, 7m1q). HSF Pro Version 4: Used to interpret potential splicing effects [46].

APPENDIX 5. SUPPLEMENTARY TABLE S3.

To access the data, click or select the words "Appendix 5." Measure (Variable) by Measure Spearman Pairwise Correlations. Unadjusted p-values indicate that 26 out of 28 possible paired correlations are significant (maximum p-value = 0.0114). The ΔΔGs calculated based on AF2 highly correlate ($\rho \geq 0.89$) with the experimentally determined structures, except for 7lkz ($\rho = 0.75$).

APPENDIX 6. SUPPLEMENTARY FIGURE S3.

To access the data, click or select the words "Appendix 6." ΔΔG variant profiles from eight measurement sources. The plot shows the in silico ΔΔG values calculated by the FoldX plugin in YASARA software [40,41], using the mutagenesis option for 10 missense variants in the study based on the seven available cryo-EM structures of human ABCA4 and AlphaFold2-generated full-length ABCA4 wild-type protein model. The complex alleles found in the patient cohort have been analyzed both separately for each variant and for the variant protein having the two variants together to detect possible combinatorial effects of these variants; however, no alleviating or augmenting effect was observed in this analysis. Seven ΔΔG values are missing due to the Arg-212 residue not being available in the 7lkp, 7m1p, and 7lkz structures and the Gly-1961 and Glu-2233 in the 7mlq structure.

ACKNOWLEDGMENTS

Financial support: Supported in part by the Foerderer Fund (A.L.), the Robison D. Harley, MD Endowed Chair in Pediatric Ophthalmology and Ocular Genetics (A.L.), Adeline Lutz - Steven S.T. Ching, M.D. Distinguished Professorship in Ophthalmology (A.L.), an unrestricted grant from the Flaum Eye Institute by Research to Prevent Blindness, and grants from the National Institutes of Health (NEI, R15EY013113; E.B.-F.), and the Foundation Fighting Blindness (BR-GE-0623-0860-UDEL E.B.-F.). We gratefully acknowledge the fellowship grant to S.C. by the Republic of Turkiye Ministry of National Education. The authors thank Dr. Shawn Polson

for his valuable discussions related to bioinformatics and Jocelyn Korth for her contributions. Ethics approval: This retrospective study was conducted in accordance with Institutional Review Board approval (IRB#14–403E) of the Wills Eye Hospital and Thomas Jefferson University.

REFERENCES

1. Stone EM, Andorf JL, Whitmore SS, DeLuca AP, Giacalone JC, Streb LM, Braun TA, Mullins RF, Scheetz TE, Sheffield VC, Tucker BA. Clinically Focused Molecular Investigation of 1000 Consecutive Families with Inherited Retinal Disease. Ophthalmology 2017; 124:1314-31. [PMID: 28559085].
2. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, Karapetyan K, Katz K, Liu C, Maddipatla Z, Malheiro A, McDaniel K, Ovetsky M, Riley G, Zhou G, Holmes JB, Kattman BL, Maglott DR. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res 2018; 46:D1D1062-7. [PMID: 29165669].
3. Biswas-Fiss EE, Kurpad DS, Joshi K, Biswas SB. Interaction of extracellular domain 2 of the human retina-specific ATP-binding cassette transporter (ABCA4) with all-trans-retinal. J Biol Chem 2010; 285:19372-83. [PMID: 20404325].
4. Xie T, Zhang Z, Fang Q, Du B, Gong X. Structural basis of substrate recognition and translocation by human ABCA4. Nat Commun 2021; 12:3853-[PMID: 34158497].
5. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Res 2015; 43:D789-98. [PMID: 25428349].
6. Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. 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. [PMID: 24077912].
7. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. Hum Mutat 2011; 32:557-63. [PMID: 21520333].
8. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL. ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015; 17:405-24. [PMID: 25741868].
9. Holtan JP, Aukrust I, Jansson RW, Berland S, Bruland O, Gjerde BL, Stokowy T, Bojovic O, Forsaa V, Austeng D, Rødahl E, Bredrup C, Knappskog PM, Bragadóttir R. Clinical features and molecular genetics of patients with ABCA4-retinal dystrophies. Acta Ophthalmol 2021; 99:e733-46. [PMID: 33258285].
10. Zhang N, Tsybovsky Y, Kolesnikov AV, Rozanowska M, Swider M, Schwartz SB, Stone EM, Palczewska G, Maeda A, Kefalov VJ, Jacobson SG, Cideciyan AV, Palczewski K. Protein misfolding and the pathogenesis of ABCA4-associated retinal degenerations. Hum Mol Genet 2015; 24:3220-37. [PMID: 25712131].
11. Shroyer NF, Lewis RA, Yatsenko AN, Wensel TG, Lupski JR. Cosegregation and functional analysis of mutant ABCR (ABCA4) alleles in families that manifest both Stargardt disease and age-related macular degeneration. Hum Mol Genet 2001; 10:2671-8. [PMID: 11726554].
12. Cremers FPM, Lee W, Collin RWJ, Allikmets R. Clinical spectrum, genetic complexity and therapeutic approaches for retinal disease caused by ABCA4 mutations. Prog Retin Eye Res 2020; 79:100861[PMID: 32278709].
13. Biswas-Fiss EE. Functional analysis of genetic mutations in nucleotide binding domain 2 of the human retina specific ABC transporter. Biochemistry 2003; 42:10683-96. [PMID: 12962493].
14. Patel MJ, Biswas SB, Biswas-Fiss EE. Functional significance of the conserved C-Terminal VFVNFA motif in the retina-specific ABC transporter, ABCA4, and its role in inherited visual disease. Biochem Biophys Res Commun 2019; 519:46-52. [PMID: 31481235].
15. Suárez T, Biswas SB, Biswas EE. Biochemical defects in retina-specific human ATP binding cassette transporter nucleotide binding domain 1 mutants associated with macular degeneration. J Biol Chem 2002; 277:21759-67. [PMID: 11919200].
16. Wangtiraumnuay N, Capasso J, Tsukikawa M, Levin A, Biswas-Fiss E. Novel ABCA4 mutation leads to loss of a conserved C-terminal motif: implications for predicting pathogenicity based on genetic testing. Eur J Ophthalmol 2018; 28:123-6. [PMID: 28885670].
17. Garces F, Jiang K, Molday LL, Stöhr H, Weber BH, Lyons CJ, Maberley D, Molday RS. Correlating the Expression and Functional Activity of ABCA4 Disease Variants With the Phenotype of Patients With Stargardt Disease. Invest Ophthalmol Vis Sci 2018; 59:2305-15. [PMID: 29847635].
18. Molday LL, Wahl D, Sarunic MV, Molday RS. Localization and functional characterization of the p.Asn965Ser (N965S) ABCA4 variant in mice reveal pathogenic mechanisms underlying Stargardt macular degeneration. Hum Mol Genet 2018; 27:295-306. [PMID: 29145636].
19. Garces FA, Scortecci JF, Molday RS. Functional Characterization of ABCA4 Missense Variants Linked to Stargardt Macular Degeneration. Int J Mol Sci 2020; 22:185-[PMID: 33375396].
20. Sun H, Smallwood PM, Nathans J. Biochemical defects in ABCR protein variants associated with human retinopathies. Nat Genet 2000; 26:242-6. [PMID: 11017087].
21. Lee W, Schuerch K, Zernant J, Collison FT, Bearelly S, Fishman GA, Tsang SH, Sparrow JR, Allikmets R. Genotypic spectrum and phenotype correlations of ABCA4-associated

disease in patients of south Asian descent. Eur J Hum Genet 2017; 25:735-43. [PMID: 28327576].

22. Westeneng-van Haaften SC, Boon CJF, Cremers FPM, Hoefsloot LH, den Hollander AI, Hoyng CB. Clinical and genetic characteristics of late-onset Stargardt's disease. Ophthalmology 2012; 119:1199-210. [PMID: 22449572].

23. Wiszniewski W, Zaremba CM, Yatsenko AN, Jamrich M, Wensel TG, Lewis RA, Lupski JR. ABCA4 mutations causing mislocalization are found frequently in patients with severe retinal dystrophies. Hum Mol Genet 2005; 14:2769-78. [PMID: 16103129].

24. Zernant J, Lee W, Nagasaki T, Collison FT, Fishman GA, Bertelsen M, Rosenberg T, Gouras P, Tsang SH, Allikmets R. Extremely hypomorphic and severe deep intronic variants in the *ABCA4* locus result in varying Stargardt disease phenotypes. Cold Spring Harb Mol Case Stud 2018; 4:a002733[PMID: 29848554].

25. Ittisoponpisan S, Islam SA, Khanna T, Alhuzimi E, David A, Sternberg MJE. Can Predicted Protein 3D Structures Provide Reliable Insights into whether Missense Variants Are Disease Associated? J Mol Biol 2019; 431:2197-212. [PMID: 30995449].

26. Kalay E, Uzumcu A, Krieger E, Çaylan R, Uyguner O, Ulubil-Emiroglu M, Erdol H, Kayserili H, Hafiz G, Başerer N, Heister AJGM, Hennies HC, Nürnberg P, Başaran S, Brunner HG, Cremers CWRJ, Karaguzel A, Wollnik B, Kremer H. MYO15A (DFNB3) mutations in Turkish hearing loss families and functional modeling of a novel motor domain mutation. Am J Med Genet A 2007; 143A:2382-9. [PMID: 17853461].

27. Kueppers F, Andrake MD, Xu Q, Dunbrack RL Jr, Kim J, Sanders CL. Protein modeling to assess the pathogenicity of rare variants of SERPINA1 in patients suspected of having Alpha 1 Antitrypsin Deficiency. BMC Med Genet 2019; 20:125-[PMID: 31307431].

28. Ou L, Przybilla MJ, Whitley CB. Phenotype prediction for mucopolysaccharidosis type I by in silico analysis. Orphanet J Rare Dis 2017; 12:125-[PMID: 28676128].

29. Sallah SR, Sergouniotis PI, Barton S, Ramsden S, Taylor RL, Safadi A, Kabir M, Ellingford JM, Lench N, Lovell SC, Black GCM. Using an integrative machine learning approach utilising homology modelling to clinically interpret genetic variants: CACNA1F as an exemplar. Eur J Hum Genet 2020; 28:1274-82. [PMID: 32313206].

30. Rhoades R, Henry B, Prichett D, Fang Y, Teng S. Computational Saturation Mutagenesis to Investigate the Effects of Neurexin-1 Mutations on AlphaFold Structure. Genes (Basel) 2022; 13:789-[PMID: 35627176].

31. Buonfiglio PI, Bruque CD, Lotersztein V, Luce L, Giliberto F, Menazzi S, Francipane L, Paoli B, Goldschmidt E, Elgoyhen AB, Dalamón V. Predicting pathogenicity for novel hearing loss mutations based on genetic and protein structure approaches. Sci Rep 2022; 12:301-[PMID: 34997062].

32. Bruque CD, Delea M, Fernández CS, Orza JV, Taboas M, Buzzalino N, Espeche LD, Solari A, Luccerini V, Alba L, Nadra AD, Dain L. Structure-based activity prediction of CYP21A2 stability variants: A survey of available gene variations. Sci Rep 2016; 6:39082-[PMID: 27966633].

33. Caswell RC, Owens MM, Gunning AC, Ellard S, Wright CF. Using Structural Analysis *In Silico* to Assess the Impact of Missense Variants in MEN1. J Endocr Soc 2019; 3:2258-75. [PMID: 31737856].

34. Cevik S, Biswas SB, Biswas-Fiss EE. Structural and Pathogenic Impacts of *ABCA4* Variants in Retinal Degenerations-An In-Silico Study. Int J Mol Sci 2023; 24:7280-[PMID: 37108442].

35. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. Nature 2021; 596:583-9. [PMID: 34265844].

36. McCulloch DL, Marmor MF, Brigell MG, Hamilton R, Holder GE, Tzekov R, Bach M. ISCEV Standard for full-field clinical electroretinography (2015 update). Doc Ophthalmol 2015; 130:1-12. [PMID: 25502644].

37. Liu F, Lee J, Chen J. Molecular structures of the eukaryotic retinal importer ABCA4. Elife 2021; 10:10-[PMID: 33605212].

38. Scortecci JF, Molday LL, Curtis SB, Garces FA, Panwar P, Van Petegem F, Molday RS. Cryo-EM structures of the ABCA4 importer reveal mechanisms underlying substrate binding and Stargardt disease. Nat Commun 2021; 12:5902-[PMID: 34625547].

39. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nat Methods 2022; 19:679-82. [PMID: 35637307].

40. Krieger E, Vriend G. YASARA View - molecular graphics for all devices - from smartphones to workstations. Bioinformatics 2014; 30:2981-2. [PMID: 24996895].

41. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. Nucleic Acids Res 2005; 33:suppl_2W382-8[PMID: 15980494].

42. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 2019; 47:D1D886-94. [PMID: 30371827].

43. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods 2010; 7:248-9. [PMID: 20354512].

44. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D, Cannon-Albright LA, Teerlink CC, Stanford JL, Isaacs WB, Xu J, Cooney KA, Lange EM, Schleutker J, Carpten JD, Powell IJ, Cussenot O, Cancel-Tassin G, Giles GG, MacInnis

RJ, Maier C, Hsieh C-L, Wiklund F, Catalona WJ, Foulkes WD, Mandal D, Eeles RA, Kote-Jarai Z, Bustamante CD, Schaid DJ, Hastie T, Ostrander EA, Bailey-Wilson JE, Radivojac P, Thibodeau SN, Whittemore AS, Sieh W. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet 2016; 99:877-85. [PMID: 27666373].

45. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res 2016; 44:W1W344-50[PMID: 27166375].

46. Desmet FO, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 2009; 37:e67[PMID: 19339519].

47. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Neale BM, Daly MJ, MacArthur DG, Ardissino D, Atzmon G, Barnard J, Beaugerie L, Benjamin EJ, Boehnke M, Bonnycastle LL, Bottinger EP, Bowden DW, Bown MJ, Chambers JC, Chan JC, Chasman D, Cho J, Chung MK, Cohen B, Correa A, Dabelea D, Daly MJ, Darbar D, Duggirala R, Dupuis J, Ellinor PT, Elosua R, Erdmann J, Esko T, Färkkilä M, Florez J, Franke A, Getz G, Glaser B, Glatt SJ, Goldstein D, Gonzalez C, Groop L, Haiman C, Hanis C, Harms M, Hiltunen M, Holi MM, Hultman CM, Kallela M, Kaprio J, Kathiresan S, Kim B-J, Kim YJ, Kirov G, Kooner J, Koskinen S, Krumholz HM, Kugathasan S, Kwak SH, Laakso M, Lehtimäki T, Loos RJF, Lubitz SA, Ma RCW, MacArthur DG, Marrugat J, Mattila KM, McCarroll S, McCarthy MI, McGovern D, McPherson R, Meigs JB, Melander O, Metspalu A, Neale BM, Nilsson PM, O'Donovan MC, Ongur D, Orozco L, Owen MJ, Palmer CNA, Palotie A, Park KS, Pato C, Pulver AE, Rahman N, Remes AM, Rioux JD, Ripatti S, Roden DM, Saleheen D, Salomaa V, Samani NJ, Scharf J, Schunkert H, Shoemaker MB, Sklar P, Soininen H, Sokol H, Spector T, Sullivan PF, Suvisaari J, Tai ES, Teo YY, Tiinamaija T, Tsuang M, Turner D, Tusie-Luna T, Vartiainen E, Vawter MP, Ware JS, Watkins H, Weersma RK, Wessman M, Wilson JG, Xavier RJ, Neale BM, Daly MJ, MacArthur DG, Genome Aggregation Database C. Genome Aggregation Database Consortium. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020; 581:434-43. [PMID: 32461654].

48. Albert S, Garanto A, Sangermano R, Khan M, Bax NM, Hoyng CB, Zernant J, Lee W, Allikmets R, Collin RWJ, Cremers FPM. Identification and Rescue of Splice Defects Caused by Two Neighboring Deep-Intronic ABCA4 Mutations Underlying Stargardt Disease. Am J Hum Genet 2018; 102:517-27. [PMID: 29526278].

49. Al-Khuzaei S, Broadgate S, Foster CR, Shah M, Yu J, Downes SM, Halford S. An Overview of the Genetics of *ABCA4* Retinopathies, an Evolving Story. Genes (Basel) 2021; 12:1241-[PMID: 34440414].

50. Khan M, Cremers FPM. ABCA4-Associated Stargardt Disease. Klin Monbl Augenheilkd 2020; 237:267-74. [PMID: 32016942].

51. Fowler NJ, Williamson MP. The accuracy of protein structures in solution determined by AlphaFold and NMR. Structure 2022; 30:925-933.e2. [PMID: 35537451].

52. Akdel M, Pires DEV, Pardo EP, Jänes J, Zalevsky AO, Mészáros B, Bryant P, Good LL, Laskowski RA, Pozzati G, Shenoy A, Zhu W, Kundrotas P, Serra VR, Rodrigues CHM, Dunham AS, Burke D, Borkakoti N, Velankar S, Frost A, Basquin J, Lindorff-Larsen K, Bateman A, Kajava AV, Valencia A, Ovchinnikov S, Durairaj J, Ascher DB, Thornton JM, Davey NE, Stein A, Elofsson A, Croll TI, Beltrao P. A structural biology community assessment of AlphaFold2 applications. Nat Struct Mol Biol 2022; 29:1056-67. [PMID: 36344848].

53. Prabantu VM, Naveenkumar N, Srinivasan N. Influence of Disease-Causing Mutations on Protein Structural Networks. Front Mol Biosci 2021; 7:620554[PMID: 33778000].

54. Curtis SB, Molday LL, Garces FA, Molday RS. Functional analysis and classification of homozygous and hypomorphic ABCA4 variants associated with Stargardt macular degeneration. Hum Mutat 2020; 41:1944-56. [PMID: 32845050].

55. Jones PM, George AM. Mechanism of ABC transporters: a molecular dynamics simulation of a well characterized nucleotide-binding subunit. Proc Natl Acad Sci U S A 2002; 99:12639-44. [PMID: 12237398].

56. Khalili-Araghi F, Gumbart J, Wen P-C, Sotomayor M, Tajkhorshid E, Schulten K. Molecular dynamics simulations of membrane channels and transporters. Curr Opin Struct Biol 2009; 19:128-37. [PMID: 19345092].

Articles are provided courtesy of Emory University and the Zhongshan Ophthalmic Center, Sun Yat-sen University, P.R. China. The print version of this article was created on 25 October 2023. This reflects all typographical corrections and errata to the article through that date. Details of any changes may be found in the online version of the article.