The Monogenic Architecture of Retinal and Neurological Diseases

Winston Lee

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2023

Abstract

The Monogenic Architecture of Retinal and Neurological Diseases

Winston Lee

Monogenic diseases, or single-gene disorders, are clinical manifestations that can be traced to genetic variation in a single gene that alters the biologically intended (wildtype) function of its protein (or mRNA) product. Although the causal gene and its function are well-understood in many monogenic diseases, this knowledge alone often does not fully encapsulate the extensive clinical spectrum of phenotypes seen in patients. This is due in part to the numerous types of pathogenic variants that can arise in a single gene, all of which can have distinct effects on disease expression. Understanding the relationship between the vast number of possible genotypes and corresponding disease phenotypes defines a gene's monogenic disease architecture—an important but poorly understood concept that can yield informative mechanistic and clinical insight. This doctoral dissertation integrates traditional sequencing approaches with in-depth characterization of patient phenotypes to elucidate the monogenic disease architecture of three etiologically distinct disorders: retinal degeneration caused by autosomal recessive variation in *ABCA4* and neurodevelopmental disease entities caused by autosomal dominant variants in *CERT1* and *PUM1*. Genetic modifiers are identified as a significant factor in the penetrance of the major disease-causing allele of *ABCA4* and several other genetic inconsistencies are resolved to create a coherent genotype-phenotype model for the disease. Insight from this model is then applied to demonstrate the effect of allele differences in disease progression and evaluation of treatment efficacy in patients. A large cohort of affected

individuals with *CERT1* variation is assembled to (1) validate the causal role of *CERT1* in disease, (2) delineate the precise mechanism of CERT protein dysfunction in sphingolipid metabolism and (3) demonstrate therapeutic efficacy of an inhibitor compound for a newly described syndrome. Finally, the mutational spectrum of *PUM1* is expanded to previously unattributed variant classes with unexpected pathophysiological consequences to patients. Not only do the findings in this dissertation advance the prospects of delivering personalized, precision medicine to patients, the overall impact underscores the importance of this integrated approach in reconciling knowledge gaps between observations at the molecular and organismal level.

Table of Contents

List of Figures and Tables

List of Abbreviations

ABCA4: ATP-binding cassette, sub-family A, member 4 (human)

AMD: age-related macular degeneration

ANOVA: Analysis of Variance

AO: Age of onset

BCVA: Best-corrected visual acuity

BEM: Bull's eye maculopathy

BP: Base pair

BPS: Bone-spicule pigment

CADD: Combined annotation dependent depletion

CER: Ceramide

CERT1: Ceramide transporter 1

CerTra: Ceramide Transporter syndrome

CF: Counting fingers

CI: circularity index

CNV: Copy number variant

DA: Dark-adapted

DDAF: Definitely decreased autofluorescence

EEG: Electroencephalogram

ER: Endoplasmic reticulum

ExAC: Exome Aggregation Consortium

FET: Fisher's exact test

ffERG: Full-field eletroretinogram

GOF: Gain-of-function

gnomAD: Genome Aggregation Database

h-DDAF: Heterogeneous- definitely decreased autofluorescence

HM: Hand motion

IQR: Interquartile range

LA: Light-adapted

LCA: Leber's congenital amaurosis

LOF: Loss-of-function

MAF: Minor allele frequency

MIM#: Mendelian Inheritance in Man #

MRI: Magnetic resonance imaging

NGS: Next-generation sequencing

NMD: Nonsense-mediated decay

OD: Oculus dexter (right eye)

ORT: Outer retinal tubulation

OS: Oculus sinister (left eye)

PADDAS: PUM1-associated developmental delay and seizures

PRCA: Pumilio1-related cerebellar ataxia

PTC: Premature truncation codon

PUM1: Pumilio homolog 1 (human)

PVS-1: very strong evidence of pathogenicity, tier 1 (denoting complete loss of function variant)

QDAF: Questionably decreased autofluorescence

ROC: Rapid-onset chorioretinopathy

ROH: Runs of homozygosity

RP: Retinitis pigmentosa

RP19: Retinitis pigmentosa 19

RPE: Retinal pigment epithelium

SD: Standard deviation

SD-OCT: Spectral domain-optical coherence tomography

SM: Sphingomyelin

SNV: Single nucleotide variant

STGD1: Stargardt disease 1

SW-AF: Short wavelength-autofluorescence

TREC+: Total receptor+

WES: Whole exome sequencing

Acknowledgements

First and foremost, I would like to thank Prof. Rando Allikmets, Ph.D. who has not only supervised my research these past 10+ years, but who has also been like a father to me. You have influenced me more than anyone else in my life, both professionally and personally. I'm looking forward to many more years of work and friendship.

I also owe a big debt of gratitude to my second graduate supervisor, Prof. Vincenzo A. Gennarino, Ph.D. Thank you for giving me a chance to step out of my comfort zone and learn new things. I will never forget how you discovered *PUM1* disease as a post-doc. This story inspires me every day to think independently and work relentlessly.

None of the work on retinal disease in this dissertation would have been possible without the immense contribution of Jana Zernant, Ph.D., a legend in the ABCA4 field, my academic partner in crime and a very close friend.

I am also extremely grateful to Prof. Stephen H. Tsang, M.D., Ph.D. for the innumerable ways you have influenced me. You are the reason I first stepped foot into the retinal field and a big part of why I hope to dedicate my life to working on inherited retinal diseases. Thank you for sharing a small slice of your encyclopedic knowledge me.

The availability of patients studied in this dissertation would not have been possible without collaborations with Dr. Gerald A. Fishman, M.D., Dr. Stanley Chang, M.D., Dr. Peter Gouras, M.D., among others. Although it goes without saying, I feel an immense amount of gratitude towards all patients and family members who I have interacted with over the past several years. Knowing and talking to all of them has taught me more than anything I could ever read out of a clinical textbook.

I would like to show immense appreciation to members of both the Allikmets lab and Gennarino lab who along with both P.I.'s have essentially been family (away from home). I must especially thank Takayuki Nagasaki, Ph.D. for the advice, encouragement and bioinformatics wisdom. Likewise, the immense *CERT1* project would not have been possible without the collaboration of Prof. Giovanni d'Angelo, Ph.D. and Vicky Brandt.

Lastly (but certainly not least), I have to acknowledge my committee members, Prof. Angela Christiano, Ph.D., Prof. Xin Zhang, Ph.D. and Prof. Krzysztof Kiryluk, M.D. for their time, expertise and guidance through this long process. Finally, I want to sincerely thank both Prof. Mimi Shirazu-Hiza, Ph.D. and Jooyeon Kim for being there for me during a difficult period these past 4.5 years. I'm not sure I would have pulled through without both of you.

Dedication

To my mom and my wife.

Chapter 1 Introduction

1.1 Overview

The intricacies of disease can be traced from the observable scale of the human body all the way down to the precise sequence of nucleotides in DNA encased within the nucleus of a cell. This ability to understand disease at such a fundamental level is the basis for modern medicine; however, progress on this front can often move at a slow and incremental pace. A reason for this may be attributed to the fact that questions in basic science are often tackled without a more nuanced understanding of disease manifestation at the clinical level. Monogenic disorders, for instance, are often naïvely characterized by the overly simplistic concept of "one gene -- one disease" whereas in reality, dysfunction of a single gene can lead to a vastly heterogeneous spectrum of disorders that differ in clinical etiology and treatment response. The gap between the existing mechanistic knowledge of a disease and its clinical manifestation in patients can be quite large and, as with most things in nature, there is almost never a straight line between cause and effect. For rare, monogenic disorders, correlative studies not only fill this gap but also provide a detailed map to navigate the vast genetic complexities found in almost all diseases.

1.2 Monogenic disease architecture

Monogenic diseases, or single-gene disorders, are clinical manifestations that can be traced to genetic variation (or mutation) in a single gene that alters the biologically intended (wildtype) function of its protein (or mRNA) product. Understanding the biological context of a gene provides important details—cell type or tissue of origin, biological pathways, etiological consequences, etc.—all of which furthers our understanding of its pathogenic consequences. Although the causal gene is unknown for many inherited disorders, there are numerous

monogenic diseases for which the precise and fundamental cellular function is well-established. Often, however, this fundamental understanding does not track with the resulting clinical phenotypes seen in patients. A common discrepancy is the breadth of variability across individuals with the same monogenic disorder. This is in part attributable to the numerous possible variants (and types of variation) that can occur in each gene, all of which can have a distinct effect on the corresponding phenotype. For autosomal recessive diseases that result from pathogenic variation on both alleles, the variability in possible phenotypic outcomes can potentially increase by an additional order of magnitude. Understanding a disorder by its "monogenic disease architecture" —that is, the relationship between all possible pathogenic variation and resulting clinical phenotypes—provides a global perspective that yields both mechanistic and clinical insight about a gene that may otherwise be unapparent at a local level (*see Appendix H*). Furthermore, understanding a monogenic disease from this perspective will help elucidate genetic irregularities or inconsistencies (e.g., genetic modifiers, auxiliary gene functions and pathways, phenotypic discordances, etc.) that may not be explainable by underlying genotype alone. This doctoral dissertation is devoted to understanding and untangling the complex monogenic disease architecture of three mechanistically well-characterized genes: *ABCA4*, *CERT1* and *PUM1*.

1.3 *ABCA4*: protein function, pathophysiology, and *in vivo* models

Pathogenic variation in the *ABCA4* gene is the underlying cause of the most common monogenic inherited retinal dystrophy in the world.[1,2] The disease is predominantly characterized by progressive loss of central vision due to the degeneration of photoreceptors and retinal pigment epithelium (RPE) cells beginning in the macula (**Figure 1**).[2] In the more than two decades since

ABCA4 was first cloned by Rando Allikmets, Ph.D. in 1997,[1] the scientific community at large

has gained an unprecedented understanding of its disease pathophysiology. The large 6.8 kb

ABCA4 gene encodes the ATP-binding cassette (ABC) transporter protein that is expressed in

both rod[3,4] and cone[5] photoreceptors—the specialized neurons in the retina responsible for

capturing light in the initial step of visual phototransduction--and is localized to the rim and

incisures of outer segment disc membranes.[6]
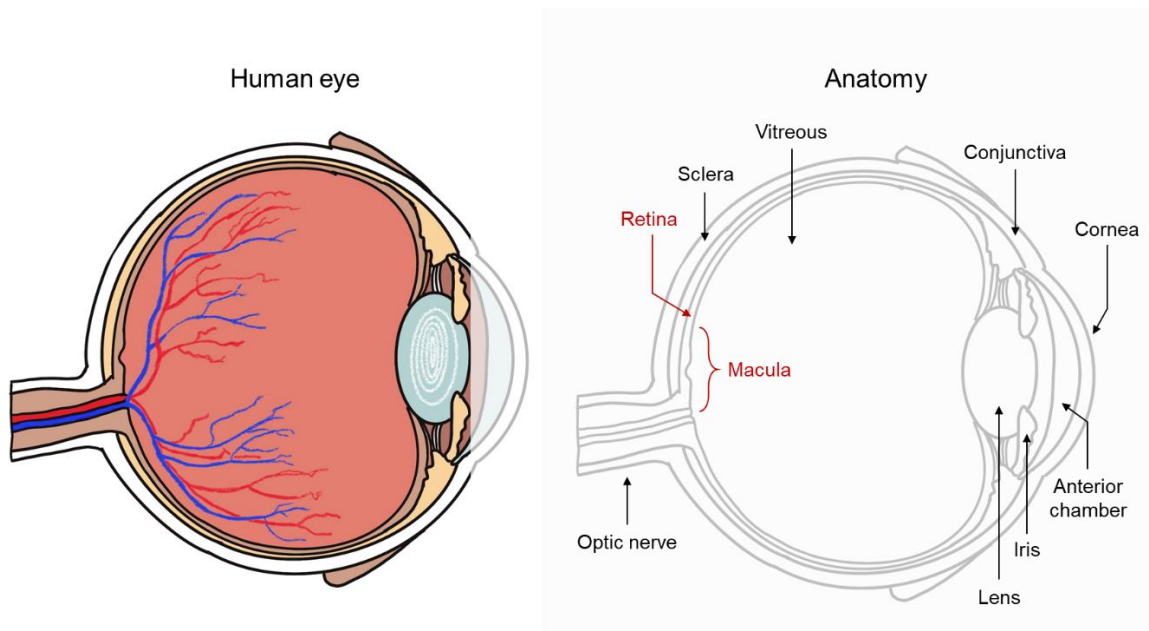
Figure 1. Anatomical illustration of the human eye.

The process of visual phototransduction relies on the regeneration of the vitamin A derivative

molecule, 11-*cis*-retinal, which occurs in a series of biochemical reactions in the visual cycle

(**Figure 2A**). In brief, 11-*cis*-retinal is a chromophore bound to an outer segment membrane-

bound G-protein-coupled receptor--rhodopsin (in rods) and photopsin (in cones)--to form the

light-sensitive complex responsible for triggering visual phototransduction. The absorption of

light results in the photoisomerization of 11-*cis*-retinal to all-*trans*-retinal and consequent release

from rhodopsin (or cone opsin) into the photoreceptor disc lumen. The released all-*trans*-retinal

is then bound by ABCA4 and "flipped"[7,8] across the disc membrane out into the outer segment

cytosol where it is converted to all-*trans*-retinol by RDH8. All-*trans*-retinol is transported out of

photoreceptors into the adjacent RPE by the aptly named interphotoreceptor retinoid-binding

protein (IRBP).[9] Within the RPE, all-*trans*-retinol is esterified by lecithin retinol acyltransferase

(LRAT)[10] and converted to 11-*cis*-retinol by the retinoid isomerohydrolase 65 (RPE65).[11-13]

Finally, 11-*cis*-retinol is then oxidized into the 11-*cis*-retinal by retinoid dehydrogenase 5

(RDH5)[14] and shuttled back into the photoreceptor outer segment disc lumen by IRBP to re-enter

the visual cycle.
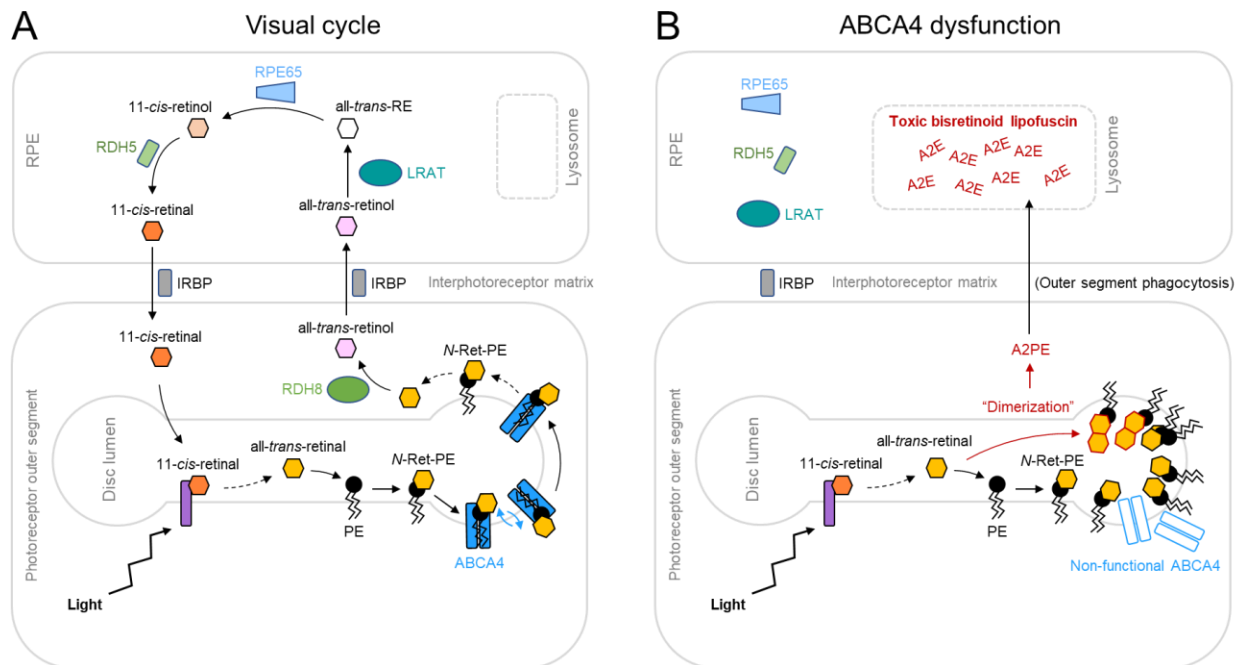
Figure 2. *Canonical pathway of the visual cycle with (A) functional and (B) non-functional ABCA4.*

The dysfunction of ABCA4 results in the accumulation of all-*trans*-retinal (**Figure 2B**) and

subsequent irreversible formation of cytotoxic bisretinoids comprising lipofuscin.[15-17] Over time,

the constituent components of lipofuscin, mostly the all-*trans*-retinal dimer N-retinyl-N-

4

retinylidene ethanolamine (A2E), become cytotoxic and trigger degeneration resulting in vision loss.[15,18,19] Although this pathophysiological process begins in photoreceptors, RPE degeneration precedes photoreceptor degeneration in the disease in most cases due to the regular process of RPE phagocytizing outer segment tips.[20,21] Pathologic A2E/lipofuscin accumulation has been visualized *in vivo* and quantitatively measured in both patients with Stagardt/ABCA4 disease[16,22-24] and knockout (KO) mice (*Abca4[-/-]*).[25,26] It should be noted that mice do not recapitulate the retinal degeneration phenotype due difference in retinal anatomy, specifically the absence of a macular region. More recently, a naturally occurring canine model (Labrador retriever) homozygous for two complete loss-of-function (LOF) alleles of *ABCA4*, c.4176insC, p.Ph1393Leufs*1395, was described.[27] Canine retinas have a cone-rich streak analogous to the human macula and unsurprisingly, diffuse degeneration was found in conjunction with accumulation of RPE lipofuscin accumulation in this region.[28,29]

1.4 Challenges Facing Genotype-Phenotype Correlation Studies

1.4.1 Scale of Heterogeneity

Despite this abundance of fundamental knowledge about the visual cycle described above, we remain far from understanding how ABCA4 dysfunction precisely gives rise to the large breadth of clinical phenotypes in patients. The first hurdle in tackling genotype-phenotype correlations is an issue of sheer scale. While the consequence of having no functional ABCA4 protein in the visual cycle is clear, disease manifestation varies considerably depending on the type of variant and combination of alleles in an individual. To date, more 2,300 pathogenic variants have reported across the 50 exons and within the deep, non-coding intronic regions of the *ABCA4* locus.[2,30] Mathematically, this number already allows over 5 million possible allelic combinations ($2,300^{2\ \text{alleles}} = 5,290,000$ possible genotypes). Virtually all mutation classes are represented in *ABCA4* including missense, nonsense, indels, canonical and noncanonical splice site changes and splice-affecting deep-intronic variants.[31-35] Large structural variants

include multi-exonic deletions, duplications and even Alu insertions.[36,37] The effect of protein-truncating or complete LOF variants is clear; however, they make up only ~23% of total variants/alleles and 33% of unique variants/alleles. Missense variants, on the other hand, which vary considerably in severity, make up 50% of the unique variants/alleles and 61% of total variant/allele composition.[2] Correspondingly, a conservative list of clinical phenotypes attributable to *ABCA4* includes: Stargardt disease,[38] fundus flavimaculatus,[39,40] cone-rod dystrophy, optical gap or occult macular dystrophy,[41,42] bull's eye maculopathy, generalized choriocapillaris dystrophy,[43,44] early-onset sever retinal dystrophy,[45] late-onset pericentral macular dystrophy,[46,47] early-onset pattern dystrophy, vitelliform macular dystrophy (see **Appendix E**),[48] among others (**Figure 3**).
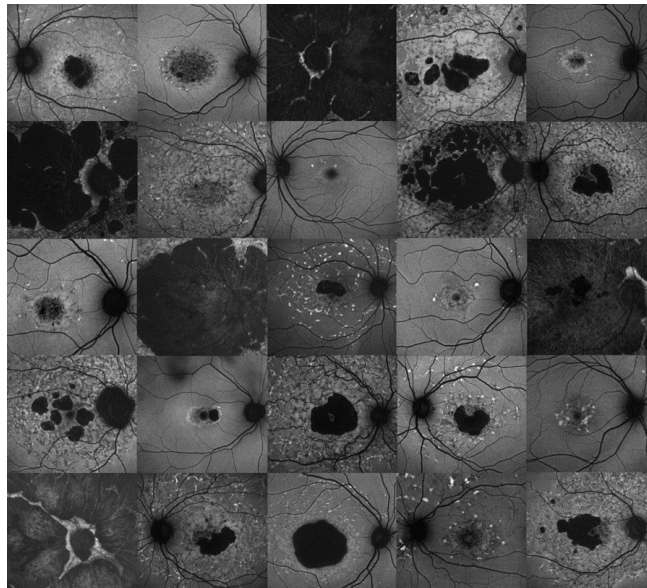
Figure 3. *Phenotypic spectrum of ABCA4 disease.*

1.4.2 Inconsistencies

While ABCA4 disease[a] is complicated enough under the standard assumptions of human genetics, a significant fraction of the disease exists outside of this understanding. The most

---

[a] "ABCA4 disease" collectively refers to all clinical entities caused by pathogenic variation in the *ABCA4* gene: Stargardt disease 1, Cone-rod dystrophy, ROC, bull's eye maculopathy, etc.

prominent example is the recent discovery of conditionally penetrant, "extremely" hypomorphic alleles,[b] most notably the p.(Asn1868Ile) variant.[46,49,50] The pathogenicity of this variant as a causal allele has been puzzling due to the contradicting observations that its allele frequency in the general population is too high to be pathogenic (~7% among non-Finnish Europeans), yet somehow more frequent in cases versus controls in several studies.[51-54] While the disease association proposed by these studies was largely due to linkage disequilibrium (LD) with other highly penetrant pathogenic variants,[53] our lab uncovered the precise conditions of pathogenicity, which occurs only when this allele is present in combination with a severe *ABCA4* allele in *trans*.[46] The resulting phenotype is a late-onset form of Stargardt disease characterized by preserved central vision due to sparing of the fovea, an anatomical phenomenon that is not fully understood.[46,47]

Another long-standing unresolved issue is the penetrance of the p.(Gly1961Glu) variant, the most common pathogenic allele in *ABCA4* found in ~20% of affected cases of European descent[55] and up to 50% of cases of South Asian descent.[56] Correspondingly, the frequency of p.(Gly1961Glu) is also exceedingly high in certain healthy populations (~0.4% in non-Finnish Europeans and 5.65% in the Horn of Africa).[55,57] This allele also gives rise to discordant phenotypes among individuals with the same genotypes (**Figure 4**), pointing to the broader issue impeding genotype-phenotype correlations, specifically the existence of genetic modifiers. Modifiers are defined here as cryptic variants that significantly influence phenotypic expression (disease heterogeneity or severity) but have no singular effect themselves. Modifiers can either

---

[b] Note: The use of the term "hypomorph" or "hypomorphic" here and throughout the text specifically denotes a mild autosomal recessive allele that is pathogenic in patients only when inherited in *trans* to a severe or complete loss-of-function allele as defined by Zernant et al. (2017) *J Med Genet*. Jun;54(6):404-412), this is in contrast to the historical usage to denote any mutation that causes a partial loss of gene function.

be intragenic (*cis*-modifiers) or extragenic (*trans*-modifiers)[c] and indeed examples of both have been found in ABCA4 disease.[46,58-60] Accounting for both incomplete penetrance and variable expression is therefore crucial, especially for the most frequent pathogenic allele in a disease.
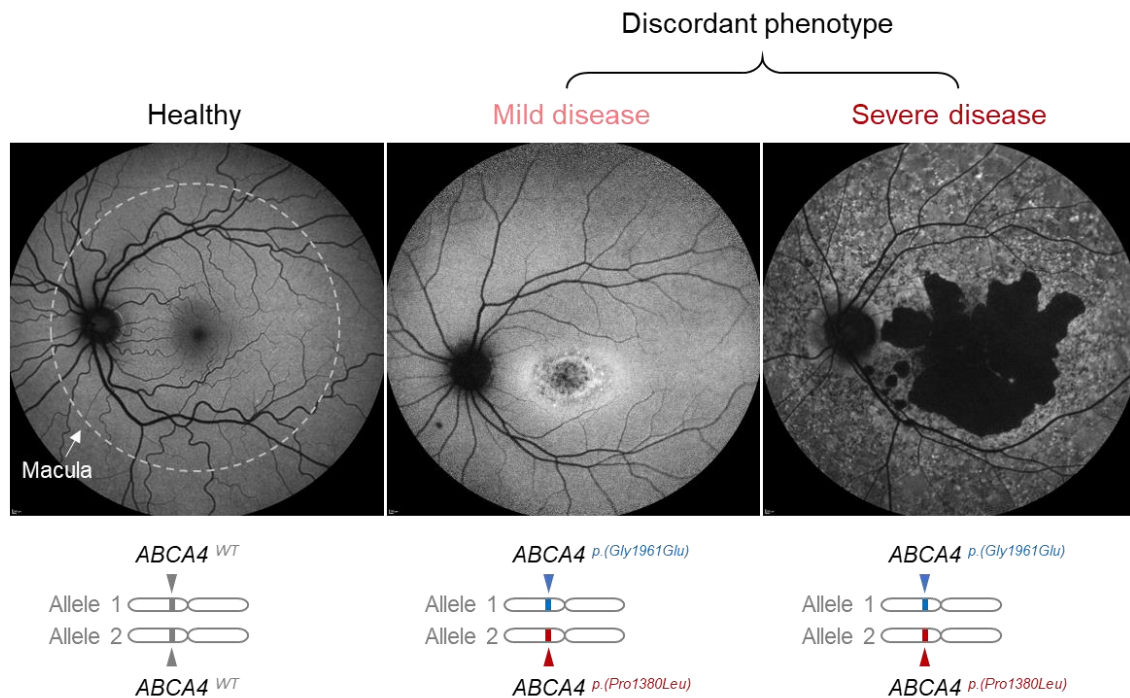
Figure 4. *Phenotypic discordance in ABCA4 disease as evidence of genetic modifiers.*

1.4.3 Paucity of Critical Details

Unlike ABCA4 disease, which suffers from "too much known complexity", there is deep clinical knowledge gap for many other genes that are otherwise functionally well-characterized and potentially treatable. One example is the ceramide transport 1 gene (*CERT1*) (also known as *COL4A3BP*). *CERT1* (NM_005713.2), which encodes the important inter-organelle lipid

---

[c] Note: Here, *cis-* and *trans-* are used to differentiate modifier variants within (*cis*) (i.e., coding, non-coding, regulatory sequences) and outside (*trans*) of a gene, respectively. In other parts of the text, *cis* and *trans* are used to denote the gametic phase of variants in a genotype (i.e., two variants on the same allele are in *cis* and two variants on opposite alleles are in *trans* relatively to one another.

transporter CERT that is responsible for transferring ceramide (Cer) from the endoplasmic reticulum (ER) to the trans-Golgi for conversion to sphingomyelin (SM) during sphingolipid metabolism.[61,62] Unlike the transport of all-*trans*-retinal by ABCA4 in the visual cycle, sphingolipid metabolism is a strictly regulated process. Once sufficient rates of SM production are reached, CERT is phosphorylated and rendered inactive to avoid a detrimental overproduction of SM.[63,64] The consequence of a complete depletion of CERT is unknown, however, conditional KO mice have been generated (*Col4a3bp*[tm1a(KOMP)Wtsi]) that do not appear to have significant phenotypes (https://www.mousephenotype.org/data/genes/MGI:1915268, accessed February 2023). CERT contains several domains with specific functional roles and knowing the type of variation that causes disease and its spatial distribution relative to these domains would be highly informative of its pathophysiology. Therapeutically, confirming a GOF (dominant negative) mechanism could open the door to pharmacological use of known CERT inhibitors such as N-(30hydroxy-1-hydroxymethyl-3-phenylpropyl)dodecamide (HPA-12).[65]

While the disease etiology caused by*ABCA4* and *CERT1* variation can be roughly distinguished as either a LOF and possible GOF problem, respectively, there are genes such as *PUM1* in which both an overabundance and depletion is known to be problematic.[66,67] The *PUM1* gene encodes Pumilio-1 (PUM1), a member of the PUMULIO/FBF (PUF) family[68] that is evolutionarily conserved from humans (*H. sapiens*) to *D. melanogaster*.[69] PUM1 is a ubiquitously expressed RNA-binding protein (RBP) that post-transcriptionally regulates specific mRNA targets[70] in a manner that is crucial to the maintenance and function of neurons in the brain.[67] Elevated levels of PUM1 reduce dendritic arborization (branching) in both induced pluripotent stem cell (iPSC)-

9

derived neurons[71] and HEK293T cells.[72] On the other hand, completely knocking out *PUM1* leads to mitotic errors during cell division.[73-75]

The connection between pathogenic variation in *PUM1* and disease has been clear since it was first discovered by Vincenzo A. Gennarino, Ph.D. Homozygous KO (*Pum1*[-/-]) are born at low Mendelian ratios while heterozygous KO mice (*Pum1*[+/-]) are relatively small in size, have seizures and motor dysfunction.[67] This haploinsufficiency model is also reflected in the human disease phenotype. The most common types of variants are heterozygous *de novo* genomic deletions spanning the *PUM1* locus which result in a clinical phenotype called P̲umilio1-A̲ssociated D̲evelopmental D̲isability, A̲taxia and S̲eizure syndrome (PADDAS).[72] Biallelic patients have not been reported. Four cases heterozygous for the *de novo* p.(Arg1147Trp) missense variant have been described as having features consistent with PADDAS.[72,76-78] This variant falls outside of the functional RNA-binding domain (RBD). Another variant inside the RBD, specifically the p.(Thr1035Ser) variant, was found to segregate with a condition distinct from PADDAS in a three-generation family with multiple affected individuals.[72] This phenotype was described as a late-onset pure cerebellar ataxia, termed P̲UM1-r̲elated c̲erebellar a̲taxia (PRCA). Even more intriguing, *in vitro* studies revealed that this etiological discordance was attributable to only a 25% difference in PUM1 levels. These perplexing observations raise numerous questions. How can such a small difference in gene dosage have such etiologically distinct consequences of a single gene and what other types of pathogenic variants exist, and do they cause disease phenotypes other than PADDAS or PRCA?[d]

---

[d] These are the central hypotheses addressed in Chapter 8.

1.5 Study Rationale and Goals

The primary goal of this doctoral dissertation is to address critical knowledge gaps in 3

etiologically distinct monogenic diseases attributed to variation in *ABCA4*, *CERT1* and *PUM1* by

analyzing genotype-phenotype correlations based on in-depth characterization of disease features

and underlying genetic variation identified by sequencing in patients. Defining these

relationships would have direct and immediate implications for treatment for 2 of the diseases.

There are quite a few treatment options in the clinical trial pipeline for ABCA4 disease

(clinicaltrials.gov), however, the disease heterogeneity impedes our ability to stratify which types

of patients can be treated and how to reliable assess the efficacy of such treatments. An existing

inhibitor compound may be repurposed as an intervention for a monogenic disease caused by

overactive CERT function, however, both the existence of such a disease and this precise

mechanism remains to be confirmed. Resolving the Mendelian complexities would allow us to

establish coherent variant-level knowledge so that treatment can be properly designed for

individual patients.

Chapter 2 Assembly of Study Cohorts, Genetic Methods, and Clinical Phenotyping

2.1 Recruitment of study subjects

All prospective study procedures related to human subjects were performed in accordance with the established tenets of the Declaration of Helsinki. Prior to subject enrollment in studies performed at the Department of Ophthalmology (*Chapters 3-6)*, written informed consent (or assent from minors below 18 years of age) was obtained from all subjects (both patients and family members when available) in accordance with protocols #IRB-AAAI9906 and #IRB-AAAB6560 that have been approved by the Institutional Review Board of Columbia University Irving Medical Center. The study in *Chapter 3* also included subjects enrolled from the Pangere Center for Inherited Retinal Diseases at the Chicago Lighthouse who were also consented under protocol #IRB-AAAI9906. All subjects enrolled and studied in *Chapters 7 and 8* were consented under protocol #IRB-AAAS7401 which has also been approved by the Institutional Review Board of Columbia University Irving Medical Center. Enrollment of these subjects in *Chapters 7 and 8* was primarily conducted through GeneMatcher (*Appendix A*). All correspondence and personally identifiable data obtained (clinical photographs, MRI, etc.) were stored on encrypted devices and transmitted via the secure CUIT email server.

2.2 Molecular Analysis of *ABCA4*

Genetic confirmation of ABCA4 disease entails the identification of at least 2 pathogenic variants using a combination of direct sequencing approaches. In a typical cohort diagnosed with Stargardt disease, next-generation sequencing (NGS) of the *ABCA4* gene uncovers 2 pathogenic variants in 60-70% of cases, 1 pathogenic variant in 15-20% of cases and no *ABCA4* variants in the remaining cases. Our direct sequencing approach starts with the amplification of

all 50 exons and flanking intronic regions with tagged PCR primers using an amplicon tagging

protocol (Access Array; Fluidigm, South San Francisco, CA;

http://www.fluidigm.com/products/access-array.html). This was followed by NGS on the Roche

454 platform (Roche Applied Science, Penzberg, Upper Bavaria, Germany). Sequences of the

barcoded samples were then analyzed using the NextGENE software (SoftGenetics, State

College, PA) to map reads against the reference genome (GRCh37/hg19) and to identify all

variants for further processing. The remaining 1 and 0 *ABCA4* variant cases were then further

assessed for the presence of non-coding intronic variants by Illumina Truseq Custom Amplicon

target enrichment strategy and sequencing (Illumina, San Diego, CA). This design targeted the

genomic region chr1:94,456,700-94,591,600 which spans the entire *ABCA4* locus as well as

~4.9kb and ~1.7kb of the 5' and 3'UTR sequences, respectively. Further technical details for

both the NGS and locus sequencing protocol can be found provided in two previously published

articles from our lab.[34,79]

2.3 Whole Exome Sequencing

In *Chapter 2*, whole exome sequencing (WES) was outsourced to either Macrogen, Psomagen

(Rockville, MD) or the Baylor College of Medicine Human Genome Sequencing Center (BCM-

HGSC).   Library preparation, exome capture and sequencing platforms differed across each

center, however, the standard workflow is as follows: Standard exome capture libraries were

generated using the Agilent SureSelectXT Low Input Target Enrichment (Magrogen), Agilent

SureSelect Human All Exon V8 (Psomagen) or the Baylor custom array HGSCCORE designed

based on NimbleGen VCRome 2.1 (Roche NimbleGen, Madison, http://sequencing.roche.com/).

Depending on the specific array employed, typical exome capture covered 30-70Mb of target

regions. NGS sequencing of enriched libraries at both centers was performed using either the

Illumina HiSeq or NovaSeq platforms (Illumina, San Diego, http://www.illumina.com). In all

samples, sequencing achieved a minimum mean read depth of 10x. Raw sequencing reads were

then filtered to remove low-quality readers and artifacts. To standardize across different

sequencing methods, all data pre-processing, variant calling, alignment and calling steps were

performed according to the Institute of Genomic Medicine (IGM) pipeline at Columbia

University.[80] Processed reads were converted to FASTQ format and mapped against the human

reference genome (hg19) with the short-read aligner software, the Burrow-Wheeler Aligner

(BWA).[81] Single nucleotide variants (SNV) and small indels were called from the aligned data

using GATK according to best practice recommendations.[82] Finally, all resulting variants were

compiled into the standard variant call format (VCF)[83] for annotation and further analysis. VCF

files and/or other variants lists are annotated with ANNOVAR[84] variant-based and gene-based

datasets available on https://annovar.openbioinformatics.org/.

2.4 Variant prioritization

The pathogenicity of all autosomal recessive variants was determined by standard parameters

such as minor allele frequency in the general population, prior disease association, variant

class/type, familial segregation and *in silico* predictions. For *ABCA4* variants, we also consider

its enrichment in our large genetic database of >800 genetically confirmed ABCA4 disease cases

relative to the general population in gnomAD under the assumption that variants significantly

enriched in patient populations may be more likely to be pathogenic.

2.4.1 Minor allele frequency

The minor allele frequencies (MAF) of all SNV and small indels are derived from gnomAD

database (v2.1.1) which contains the sequenced exomes of 125,748 and genomes of 15,708 (total

= 141,456) presumably healthy individuals, respectively (https://gnomad.broadinstitute.org/).

When applicable, the available ethnicity-specific allele frequencies are used for specific ethnic patient subgroups although the overwhelming majority of our cohort is of non-Finish European (NFE) descent. As a standard rule, we set the MAF threshold for pathogenicity in autosomal recessive variants at ≤0.005 (or 0.5%) based on the generalization that highly penetrant variants that cause Mendelian (monogenic) disease are very rare in the healthy population. We do not, however, consider this a fixed MAF threshold, given the existence of incompletely penetrant alleles in *ABCA4*[46,47] and *PUM1*,[72] as well as other diseases with much higher MAF's. Adjustments to the MAF threshold are made on a gene-by-gene basis.

2.4.2 Prior disease association

A variant's prior disease association of a variant is primarily verified using several aggregate clinical databases. ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) is used for initial cross-referencing. For variants in *ABCA4*, we also reference our own internal database of >800 genetically-confirmed ABCA4 disease cases, as well as the Leiden Open Variation Database (LOVD) (https://databases.lovd.nl/shared/genes/ABCA4), which contains >2,300 known pathogenic variants in *ABCA4* as of 2023.[30] For variants associated with neurological disease, we primarily reference Decipher (https://www.deciphergenomics.org/) and the internal database of the Human Genome Sequencing Center of Baylor College of Medicine (private access). Further clinical and genetic databases referenced are provided in the Methods sections of the respective study chapters of this dissertation.

2.4.3 Variant classification

Based on the general assumption that most disease-causing variants affect protein coding regions of a gene, the priority ranking for pathogenicity of variant class/types is as follows (from least likely to most likely to be pathogenic): synonymous (silent) → missense → frameshift/non-

frameshift → nonsense (stop gain/loss) → canonical splice site substitutions/changes (±1 or 2

position) → large genomic deletions/insertions. The latter 4 variant classes are assumed to result

in null or complete LOF allele and using the "PVS1" acronym defined by according to the

standards and guidelines for the interpretation of sequence variants by the American College of

Medical Genetics and Genomics and the Association for Molecular Pathology. The PVS1

acronym denotes the criterion for variants with <u>v</u>ery <u>s</u>trong evidence of <u>p</u>athogenicity (tier <u>1</u>) (see

Table 2 in *Chapter 5*).[85]

2.4.4 Familial segregation

One of the most important pathogenicity criteria considered for both the autosomal recessive and

autosomal dominant diseases in our studies is the segregation of alleles and phenotypes in

families. Whenever possible, first degree relatives (parents, siblings, and children) are fully

examined and genetically screened along with the proband to verify the gametic phase of

variants identified after sequencing (see **Appendix Figure E7**). In the absence of relatives,

phasing may be statistically inferred based on observed variant co-occurrence in gnomAD

(https://gnomad.broadinstitute.org/variant-cooccurrence). Individual allele and haplotype

frequencies are obtained and the relative coefficient of linkage disequilibrium (*D*) is calculated

for both variants (see full calculation in **Appendix Figure E8**). For hypothetical variants C>T

and G>A, individual haplotype counts presented in a 2x2 contingency table:

	A	G
T	α	β
C	γ	δ

The coefficient of linkage disequilibrium (D) is defined as the difference between the observed

and expected haplotype frequencies, where $-1 \leq D \leq 1$. A negative value of D indicates that the

two alleles are less likely to be inherited together than expected by chance, while a positive value

of D indicates that they are more likely to be inherited together than expected by chance.

$$D = P(\alpha) - P(A)P(T)$$

Since the range of possible values of D depends on the frequencies of the alleles it refers to,

comparisons between different pairs of alleles may be unreliable. As such, normalizing D by

dividing it by the theoretical maximum difference between the observed and expected haplotype

frequencies to determine D' addresses this issue:

$$D' = \frac{D}{D_{max}}$$

where,

$$D_{max} = \begin{cases} max\{-P(A)P(C), -(1 - P(A))(1 - P(C))\} & when\ D < 0 \\ min\{P(A)(1 - P(C)), (1 - P(A))P(C)\} & when\ D > 0 \end{cases}$$

The range of D' is $0 \leq D' \leq 1$ such that a value of 0 indicates complete linkage disequilibrium

indicating that the variants are likely allelic, whereas a value of 1 indicates complete linkage

disequilibrium indicating that the variants are biallelic.

In addition, we subsequently calculate a correlation coefficient (R^2) between the two variants as

follows:

$$R^2 = \frac{D^2}{P(A)P(G)P(T)P(C)}$$

Statistical significance is determined by calculating a X^2 statistic ($df = 1$):

$$X^2 = \sum \frac{N(\alpha \times \delta - \gamma \times \beta)^2}{(A)(G)(C)(T)}$$

For calculations with very small sample sizes (e.g., number of cases is no greater than 15 (and consequently $N_{alleles} < 30$), one should apply Yates's correction for continuity to the X2 test, where 0.5 is subtracted from the numerical difference between the observed and expected frequencies:

$$X^2 = \sum \frac{|(N_{alleles}(\alpha \times \delta - \gamma \times \beta)| - 0.5)^2}{(A)(G)(C)(T)}$$

This statistical approach is applied in certain cases, however, the family-based approach is nevertheless preferred. Familial segregation is additional factor in the analysis of the autosomal dominant diseases in *Chapter 7* (*CERT1*) and *Chapter 8* (*PUM1*) since most pathogenic variants in these cases are not expected to be inherited but instead, occur *de novo* in the proband. Exclusion of inheritance is important however we cannot strictly rule out pathogenicity of inherited variants given the possible existence of non- or incompletely penetrance alleles.

2.4.5 *In silico* prediction

The pathogenicity of missense variants is comparatively more challenging to assess. For these variants, we consider a combination of factors based on the predicted effect of biochemical changes and the position of the altered amino acid relative to functionally and/or structurally critical domains of the protein. The biological importance of variant positions is based on quantitative estimates such as the ratio of nonsynonymous to synonymous variants missense

tolerance ratios observed across the general population (dN/dS)[e][86,87] in proteins or the level of DNA sequence conservation using phyloP.[88] Pathogenicity based on the chemical and structural impact on the protein are predicted using several composite deep learning algorithms: CADD (PHRED scale 0 to 48),[89,90] REVEL (likelihood ratio 0 to 1),[91] Eigen (probability score -4.09 to 6.31),[92] and M-CAP (percentile range 0 to 1)[93]. As a general guideline, we abide by the recommended pathogenicity thresholds of 0.025 for M-CAP, 0.5 for REVEL, 0.5 for Eigen and 20 for CADD-PHRED. For deep intronic variants or those adjacent to canonical splice sites ($> \pm$ 2), possible effects on splicing are predicted using tools such as such as MaxEntScan[94] and Human Splicing Finder,[95] SPIDEX,[96] SpliceRover,[97] EX-SKIP[98] and SpliceAI.[99]

As with allele frequency thresholds, these pathogenicity prediction thresholds may not reflect some genetic phenomena such as incomplete penetrance. As such, prediction scores are adjusted on a gene-by-gene basis. For autosomal dominant diseases caused by *de novo* variation, we developed a custom analysis pipeline that calibrates pathogenicity thresholds against the distribution of predicted scores of singleton[f] missense variants for a respective gene (i.e., missense variants with an allele count =1 in the gnomAD database) (**Figure 5**). It has been estimated that up to 5% of the general population harbors variants associated with autosomal dominant disease but will not or has not yet developed symptoms;[100] hence their inclusion in databases of healthy, unaffected individuals. Given this observation, considering the distribution of singleton missense variants provides a more conservative cut-off for *de novo*

---

[e] Specifically, the dN/dS ratio quantifies selective pressures by comparing the rate of substitutions at silent/synonymous sites (dS), which are presumed to be neutral, against the rate of substitutions at non-silent/non-synonymous sites (dN), which likely undergo selection.
[f] Variants found in one (a single) individual (allele count = 1).

variants exhibiting incomplete penetrance, such as the p.(Thr1035Ser) variant in *PUM1*.[72] More

details and codes for this calibration procedure are available in **Appendix A**.[101,102]
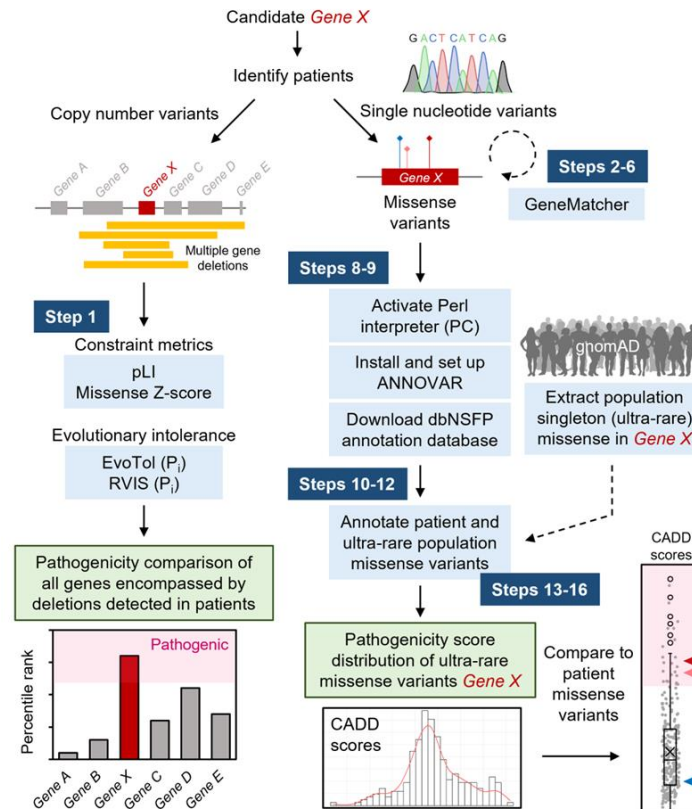
Figure 5. *Pipeline for gene-specific calibration of pathogenicity score thresholds.*[g]

2.5 Retinal and Fundus Imaging

The retinal phenotyping protocol includes three principal fundus imaging modalities: color

photography, fundus autofluorescence imaging and optical coherence tomography (OCT).

Pupils were maximally dilated to a minimum of 7-8 mm in diameter using topical 1%

tropicamide and 2.5% phenylephrine prior to all image acquisition. Color and red-free

[g] See Appendix A for full protocol. The text and figures from this protocol have been recently published (Lee et al. *STAR Protocols*. 2022 Feb 2;3(1):101150).

photographs of the fundus of 30° x 30° and 50° x 50° field magnification were captured using the FF450[plus] Fundus Camera (Carl Zeiss Meditec).

The evaluation of *ABCA4*-related disease features in our studies are primarily based on autofluorescence imaging, especially short wavelength-autofluorescence (SW-AF) which has an excitation wavelength of 488-nm and emits in the range of 500-nm to 680-nm.[103] This specific mode of autofluorescence predominantly emanates from the complex mixture of bisretinoids in photoreceptor and RPE lipofuscin.[15] While SW-AF provides high resolution images of most retinal diseases affecting the photoreceptor and RPE layers, it is especially informative as a tool for studying ABCA4 disease. The absolute levels of autofluorescence can be precisely quantified[103] and indeed, elevated SW-AF, as well as the long wavelength, near-infrared-autofluorescence (NIR-AF, 787-nm, emission >830-nm), have been measured *in vivo* in patients and mouse models of ABCA4 disease.[22-26]

SW-AF images of 30° x 30° and 55° x 55° field magnification were acquired using the confocal scanning laser ophthalmoscope Spectralis HRA+OCT (Heidelberg Engineering, Heidelberg, Germany). NIR-AF images of 30° x 30° and 50° x 50° field magnification were acquired on the confocal scanning laser ophthalmoscope HRA2 (Heidelberg Engineering, Heidelberg, Germany). Both SW-AF and NIR-AF images were acquired as "high-resolution" (1536x1536 pixels) and pixel histogram-normalized (histogram-stretch) in order to better visualize the distribution of autofluorescence intensity throughout the image. In **Chapter 4**, green autofluorescence images (532-nm excitation) and pseudo-color fundus photographs were acquired with the Optos 200Tx (Optomap Daytona, United Kingdom) instrument (*see **Chapters 3 and 4** Methods for additional specifications*).

2.6 Spectral domain-optical coherence tomography

Optical coherence tomography (OCT) renders *in vivo* cross sections of the retina providing near cellular level visualizations of tissue architecture at a resolution of about 10-15 micrometers.[104] OCT differs from other tomographic scanning modalities such as ultrasound in that the latter measures acoustic waves whereas the former measures light. The manner of detection has varied since the inception of OCT ranging from Time and Fourier domains. The OCT data acquired in this study use spectral domain-OCT (SD-OCT), which captures at a speed of approximately 27,000-70,000 A-scans per second (axial resolution ~10μm, transverse resolution ~20μm) to diminish the likelihood of movement artifacts when imaging live patients.[105] The SD-OCT scans in these studies are also captured simultaneously alongside corresponding infrared-reflectance (IR-R) images of the fundus using the Spectralis HRA+OCT (Heidelberg Engineering, Heidelberg, Germany). The SD-OCT capture protocol for each eye of each subject includes a 9 mm horizontal line scan (normalized, high-resolution) aligned and 19-line raster scan aligned at the anatomical fovea. Custom raster scan patterns used are specified in the Methods section of each individual study.

The range and depth of a standard SD-OCT scan includes all cellular layers of the retina from the vitreoretinal interface to the RPE/choroid interface (**Figure 6**). The thickness of various retinal layers can be assessed by segmenting each of these layers directly in each scan. Both manual and automatic segmentations exist, however, we use a custom "semi-automated" approach of automatic segmentation and manual correction (if needed).[106] The cellular layers quantified in our studies are total receptor+ (TREC+) which corresponds to the total length of photoreceptors and apical tips of RPE (**Figure 6**).[107] For segmentation, TREC+ is defined as the distance

between the Bruch's membrane/choroid interface and the inner nuclear layer (INL)/outer

plexiform layer (OPL) boundary (**Figure 6**).
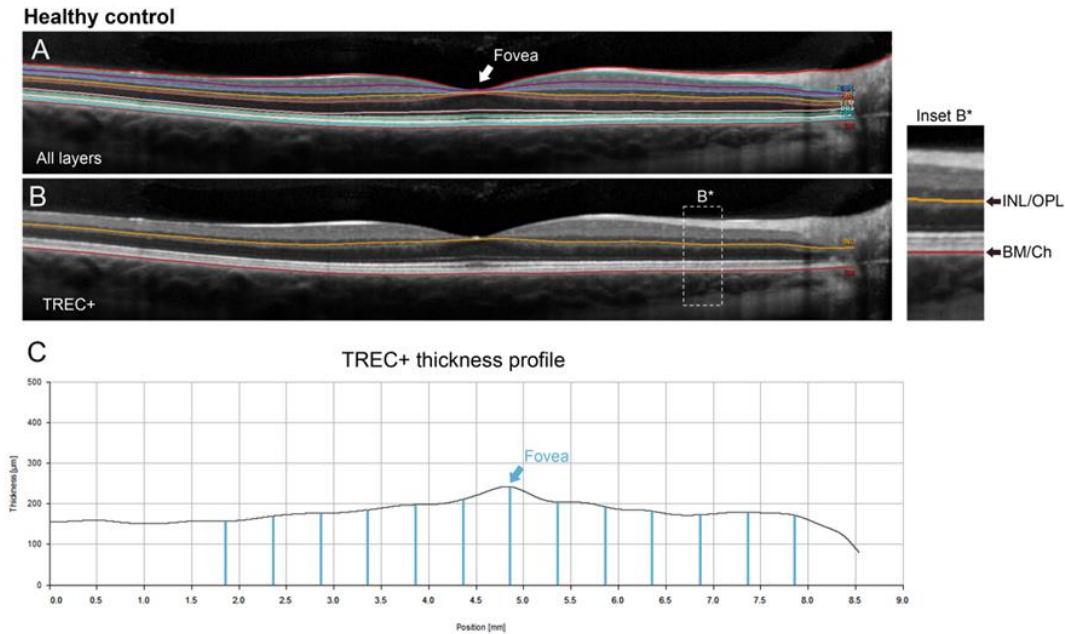
Figure 6. Semi-automated segmentation of spectral domain-optical coherence optical coherence tomography (SD-OCT) scans and thickness quantification of the total receptor+ (TREC+) layer in a healthy retina.[h]

2.7 Electroretinography

Full-field electroretinograms (ffERG) were recorded using the Diagnosys Espion (E^3)

Electrophysiology System (Diagnosys LLC, Littleton, MA, USA) in accordance with the

recommended protocol of the International Society for Clinical Electrophysiology of Vision

(ISCEV).[108] A total of six evoked responses were recorded in all patients (rod-specific dark-

adapted 0.01, mixed rod and cone dark-adapted 3.0, dark-adapted 10.0, oscillatory potentials,

---

[h] (A) Automated segmentation of all retinal layers visible on SD-OCT was performed by the HEYEX software (Heidelberg Engineering). (B) The TREC+ layer spanning photoreceptor attributable layers are demarcated by the inner nuclear layer/outer plexiform (INL/OPL, orange line)(Inset B*) interface and Bruch's membrane/choroidal (BM/Ch, red line) interface boundaries. Mis-segmented regions were manually corrected as necessary. (C) Thicknesses were obtained at 0.5 mm interval positions (blue line markers) out to 3 mm eccentricity from the fovea.

single flash light-adapted 3.0 and light-adapted 30-Hz flicker); however only four were

specifically evaluated for our studies of ABCA4 disease. Pupils were maximally dilated to a

minimum of 7-8 mm in diameter using topical 1% tropicamide and 2.5% phenylephrine prior to

testing. Dawson-Trick-Litzkow (DTL) electrodes (silver impregnated fiber) over the cornea with

ground electrodes attached to the forehead were used in almost all cases; any data recorded with

Burian-Allen (BA) contact lenses are specified in the *Methods* section of the individual study.

Patients were dark-adapted for a minimum of 40 minutes prior to performing the scotopic

protocol and light-adapted for at least 10-15 minutes prior to performing the photopic protocol.

The scotopic protocol was performed using the rod-specific dark-adapted dim flash 0.0 cd·s/m^2

(DA 0.01) to evoke the mass response of ON-bipolar cells of the rod system (b-wave amplitude

in µV) and the dark-adapted (maximal) bright flash 3.0 cd·s/m^2 (DA 3.0) for recording the mixed

rod and cone responses (b-wave minus a-wave amplitude in µV). The photopic protocol was

performed using the light-adapted 3.0 cd·s/m^2 rapid 30 Hz "flicker" (30-Hz Flicker) and the

light-adapted 3.0 cd·s/m^2 single flash to assess generalized cone pathway function. Implicit time

delays (in milliseconds), particularly in response to the 30-Hz Flicker are indicative of

generalized cone dysfunction. All evoked responses are presented as flash strength in candela-

seconds (cd·s) per meter squared (m^2) or cd·s/m^2. During acquisition, 30-50 single traces are

recorded for each stimulus in each eye (**Figure 7A**). During post-acquisition processing, outlier

traces (if any) are manually removed. Remaining single traces are averaged and extracted for

analysis (**Figure 7B**).
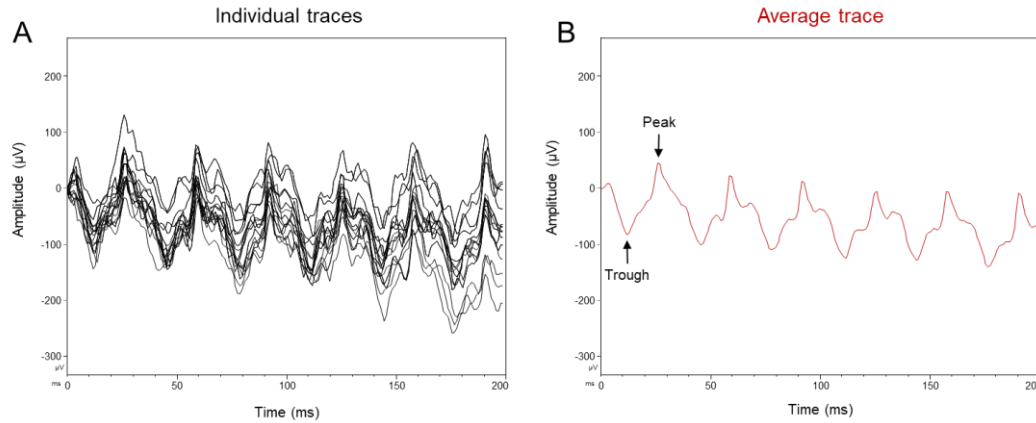
Figure 7. *Representative post-processing of ffERG recordings (light-adapted 3.0 cd·s/m² rapid 30 Hz flicker).*

ffERG classifications in **Chapter 5** were assigned according to electrophysiological attributes described by Lois et al.[109] Group 1 is characterized by no detectable loss in scotopic or photopic function; Group 2 is characterized by photopic loss, but normal scotopic function; and Group 3 exhibits deterioration of both scotopic and photopic function. Normative amplitude and implicit time ranges for each stimulus were provided by the Diagnosys system software.

Chapter 3 *Cis*-acting modifiers in the *ABCA4* locus contribute to the penetrance of the major disease-causing variant in Stargardt disease

3.1 Introduction

Variation in the *ABCA4* gene[1] causes a spectrum of recessive macular disease phenotypes, ranging from severe, rapid-onset chorioretinopathy (ROC)[110] to late-onset maculopathies resembling age-related macular degeneration (AMD).[46] This phenotypic heterogeneity is attributable to the >2,300 known pathogenic variants in coding sequences[30] and by many hypomorphic alleles in the *ABCA4* locus in both exons and (deep) introns.[32,47] As such, the ABCA4 locus represents a model for extensive genetic complexity underlying a Mendelian disorder.[32,47] Some of the hypomorphic alleles are very frequent (minor allele frequency (MAF) up to 7%) in certain populations and manifest only under certain conditions; i.e. the strength of the allele *in trans* becomes a deciding factor.[46,47] More recently, an emerging group of variants has been determined to be modifiers, which do not result in pathogenicity alone, but instead increase the penetrance of some alleles in certain allelic configurations.[32,46,50]

The c.5882G > A p.(Gly1961Glu) variant is the most frequent disease-causing allele in Stargardt/ABCA4 disease found in ~ 20% of patients of European descent,[55] and >50% of cases in South Asia.[56] Its origin can be traced to the 'Horn of Africa', where its MAF is estimated to be ~5.65% based on a genotyping data from a small ethnic diaspora group.[55,57] Historical migration out of this region has spread p.(Gly1961Glu) throughout other continents but the allele frequency has dropped significantly.[55] The MAF for this variant is 0.004 in non-Finnish Europeans and much lower, or almost nonexistent, in many racial and ethnic groups and geographical locations, including East Asia, West Africa and Native American populations, suggesting that the variant is causal in all, or at least most, populations (https://gnomad.broadinstitute.org/; accessed February

2023). Due to its high population frequency, the pathogenicity of the p.(Gly1961Glu) variant

has been questioned;[57] however, it appears mostly as a fully penetrant allele and the most

frequent *ABCA4* disease-associated allele in almost all populations studied. The suggested

pathogenicity is also supported by the fact that the population frequency of p.(Gly1961Glu) has

dropped significantly during population migration, suggesting evolutionary pressure. One

question has dominated the discussion—there are fewer than expected homozygous

p.(Gly1961Glu) cases detected in STGD1 cohorts (based on population frequencies) suggesting

low penetrance.[30] At the same time, the clinical and genetic data are not consistent with the

functional analyses, where the p.(Gly1961Glu) amino acid change has profound effects on the

ATPase and transport activities of the ABCA4 protein.[111,112]

It is also evident that some additional variants, both coding and non-coding, in the ABCA4 locus

have occurred on chromosomes carrying p.(Gly1961Glu); however, the haplotype spanning the

3' end of the gene is the same in all individuals, suggesting a single founder occurrence of the

mutation (**Figure 7**).[55] While in several cases other coding missense variants have been

identified on the same chromosome with p.(Gly1961Glu),[55] the analysis of possible disease-

associated intronic variants in *cis* with the mutation has not been comprehensively performed.

We had initiated the analysis of the entire ~140-kb *ABCA4* locus, including all exons, introns,

and 5 and 3′ regulatory regions, several years ago.[34] The analysis revealed many possibly

disease-causing alleles in non-canonical splice sites and deep intronic regions, which have been

functionally investigated.[32,49,113] One of these potentially disease associated variants is c.769-

784C > T, which was described as a hypomorphic (i.e. very mild) variant.[32,50] It was suggested

that this variant, similar to the c.5603A > T p.(Asn1868Ile) and c.4253 + 43G > A variants,[46,47]

expresses only when in trans from a deleterious ABCA4 variant. It is unlikely, however, that the

variant is disease-causing, even as a hypomorph, for two reasons. First, the c.769-784C > T variant, with one exception, has not been identified 'alone'; i.e. without any other known disease-causing allele in *cis*.[32,35] In these studies the c.769-784C > T variant was always in *cis* with p.(Asn1868Ile) (7), which has been shown to be pathogenic when "alone".[46] Second, c.769-784C > T resulted in the insertion of a pseudo-exon in ~ 30% of *ABCA4* mRNA derived from the relevant allele in iPSC-derived photoreceptor precursor cells, which is close to the threshold for *ABCA4* pathogenicity.[32,33] Nevertheless, the question about whether the c.769-784C > T variant could be a modifier, remained unresolved.

We analyzed 12 homozygous p.(Gly1961Glu) cases in 2012 and determined that half of these had another coding variant in *cis* on at least one of the two alleles.[55] At the time only coding sequences of *ABCA4* were analyzed, leaving the information of intronic variation obscure. In this chapter, we performed the analysis of a large cohort of patients with the p.(Gly1961Glu) variant in heterozygosity and homozygosity for the entire *ABCA4* locus to reassess the possibility of cis-modifiers influencing the penetrance of the p.(Gly1961Glu) allele.

Text and figures from this chapter have been recently published: Lee et al. *Hum Mol Genet*

. 2021 Jun 26;30(14):1293-1304.

3.2 Results

3.2.1 The deep intronic variant, c.769-784C>T, is enriched among p.(Gly1961Glu) alleles

Of the 644 biallelic cases, 150 were compound heterozygous or homozygous for the p.(Gly1961Glu) variant (**Supplemental Table 1**). Of these, 23 also harbored the c.769-784C > T variant, of which 22 were heterozygous and one homozygous for both p.(Gly1961Glu) and c.769-784C > T variants, comprising a complex allele c.[769-784C > T;5882G > A]

(**Supplemental Table 2**, **Figure 8 and Figure 9**). Both variants were always on the same chromosome, suggesting that c.769-784C > T appears on the p.(Gly1961Glu) haplotype in ~15% of p.(Gly1961Glu) alleles (**Figure 8**).
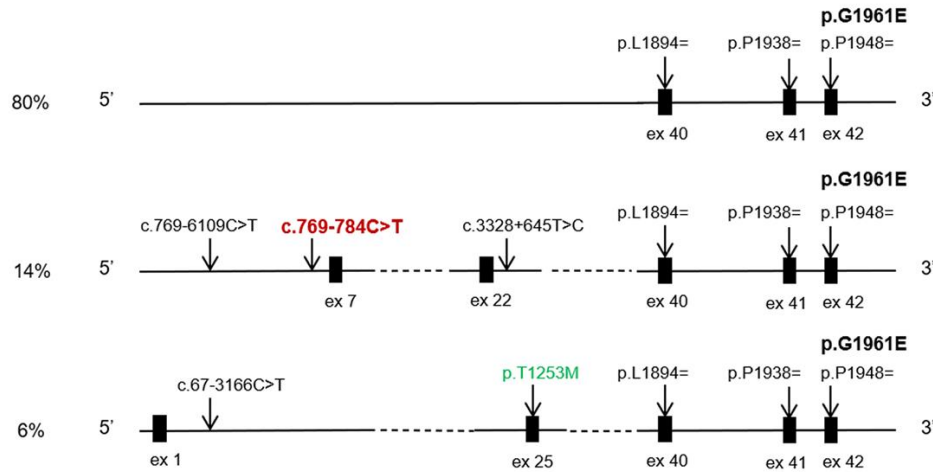
Figure 8. Most frequent haplotypes in STGD1 patients carrying the c.5882G > A, p.(Gly1961Glu) variant.[i]

In addition to the c.769-784C > T variant, two other variants, c.769-6109C > T and c.3328 + 645T > C, defined a specific haplotype. The c.769-784C > T variant is suggested to moderately affect splicing by *in silico* analysis and has shown to result in pseudo-exon insertion in ~30% of mRNA from the relevant allele.[50] The c.769-6109C > T and c.3328 + 645 T > C variants are not predicted to affect splicing by *in silico* analysis (http://www.interactive-

biosoftware.com), therefore, pending functional confirmation, these variants are considered haplotype-tagging and not pathogenic.

The c.769-784C > T variant was found almost exclusively in *cis* with the p.(Gly1961Glu) variant. An earlier study from the Netherlands reported this variant in *cis* with the p.(Asn1868Ile) variant.[32] Of particular interest is the observation that the haplotype containing both c.769-784C > T and p.(Gly1961Glu) variants was present in 6/10 cases homozygous for p.(Gly1961Glu) (**Figure 9, Supplemental Table 2**).
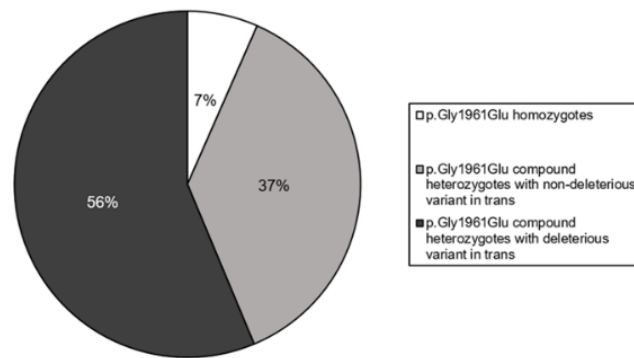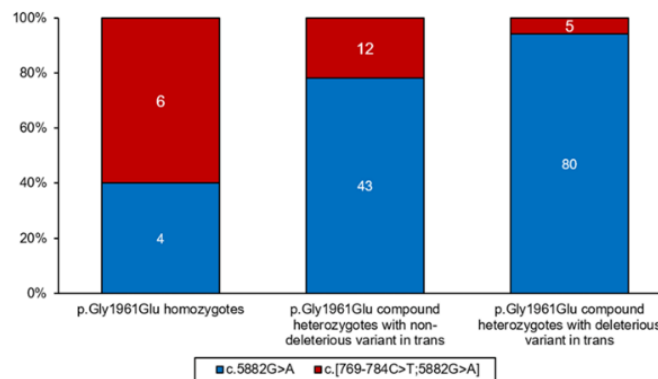
Figure 9. *Summary of genotypes in the p.(Gly1961Glu) study cohort.*[j]

---

[j] Out of 644 total biallelic patients, 150 (34.7%) harbored the p.(Gly1961Glu) variant.[j] (A) Within this group, 10 (7%) patients were homozygous while the remaining are compound heterozygotes. In more than half of the compound heterozygotes (n = 85, 56%), the mutation in *trans* was a known or expected complete LOF allele (PVS1) as compared to the remaining 56 (37%) compound heterozygotes whose allele in *trans* is likely not a complete LOF

In five cases, one of the two alleles contained the c.769-784C > T variant in addition to the p.(Gly1961Glu); and in one case, both alleles contained the complex allele. Two cases did not have the c.769-784C > T or any other variant on either allele; three cases had another exonic variant on the same allele with the p.(Gly1961Glu). These include the previously described p.([Thr1253Met;Gly1961Glu]) allele[55] in one homozygous patient and five compound heterozygous patients of Ashkenazi Jewish descent, the p.([Tyr850Cys;Gly1961Glu]) allele in one homozygous and one compound heterozygous patient, and the p.([Gly1748Arg;Gly1961Glu]) allele in one homozygous patient (**Supplemental Table 2**). The contribution of the c.769-784C > T variant for enhanced penetrance in homozygous p.(Gly1961Glu) cases is also supported by allele segregation data in 58 families. The c.[769-784C > T;5882G > A] complex allele segregated with the disease in all 12 families. The single c.5882G > A; p.(Gly1961Glu) allele segregated in 44 out of 46 families. The two remaining families had one unaffected individual each who were homozygous for the p.(Gly1961Glu) variant without any *cis*-modifiers (**Supplemental Figure 1**).

3.2.2 Harboring the additional c.769-784C>T variant on the p.(Gly1961Glu) allele resulted in increased clinical severity

Demographic, clinical and genetic characteristics of all homozygous p.(Gly1961Glu) cases are summarized in **Supplemental Table 2**. The mean age of onset (AO) of all homozygous p.(Gly1961Glu) patients (n = 10) is 42.7 years (median = 43 years, range = 21–60 years, SD = 12.3) which is significantly delayed compared to other compound heterozygous p.(Gly1961Glu) patients (mean = 22.1, median = 20 years, range = 5–60 years, SD = 11.4)

---

allele. (B) The deep intronic c.769-784C > T variant was in *cis* with p.(Gly1961Glu) in 6 out of 10 homozygotes. Among compound heterozygotes, a significantly smaller fraction of those with deleterious alleles in trans had the c.769-784C > T variant on the p.(Gly1961Glu) allele (5.9% vs 21.4% in non-deleterious alleles; FET, P = 0.006).

(P < 0.005). Best-corrected visual acuities at initial examination ranged from 20/20 to 20/80 in both eyes. Fundoscopic examinations were largely unremarkable with the exception of mild changes in the central macula. Multimodal imaging of the retina identified more distinguishable characteristics of retinal degeneration. Patient 1, a 61-year-old woman, exhibited a relatively larger lesion of chorioretinal atrophy in only one eye (OS) and was the only patient to retain sparing of the fovea late in life. The remaining patients presented with either round bull's eye maculopathy (BEM) lesions or diffuse foveal atrophy surrounded by small, nascent autofluorescent flecks at the edge of atrophy. Normal photopic and scotopic recordings on full-field electroretinogram (ffERG) testing were detected in all homozygous p.(Gly1961Glu) patients, indicating no generalized dysfunction of the cone and rod systems, respectively (**Supplemental Table 2**).

While the disease phenotypes of affected homozygous p.(Gly1961Glu) patients were relatively similar, differences in lesion morphology were observed between homozygous patients with and without the c.769-784C > T variant. Homozygous p.(Gly1961Glu) patients with the c.769-784C > T variant consistently exhibited lesions of diffuse foveal atrophy with no discrete borders and nascent autofluorescent flecks (**Figure 10E–10G, insets, 10H**), whereas patients without c.769-784C > T variant, including Patient 7 (homozygous for the complex p.[(Tyr850Cys;Gly1961Glu)] allele), exhibited apparent BEM lesions with a round, well-delineated border, and a notable absence of flecks (**Figure 10A–10C, insets, 10D**).
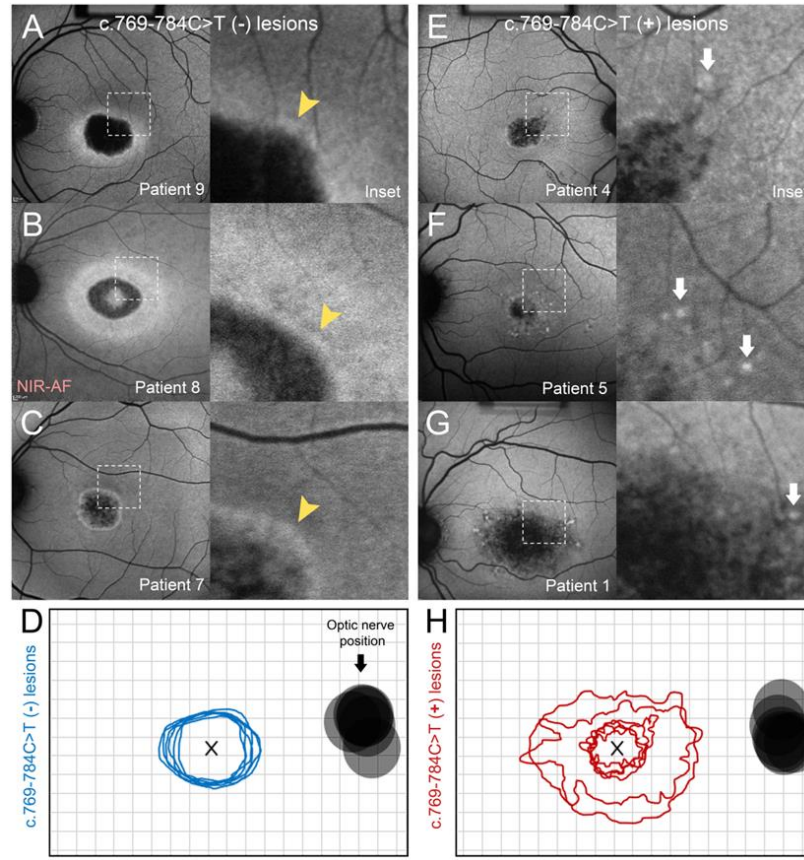
Figure 10. Clinical phenotype of patients homozygous for the p.(Gly1961Glu) allele.[k]

In compound heterozygotes, the c.769-784C > T variant is ~3.6 times more often on the complex

allele with p.(Gly1961Glu) when *in trans* from a non-deleterious variant. It is *in trans* from

21.4% of those chromosomes (12/55) vs. 5.9% of deleterious alleles (5/85; (Fisher's exact test,

P = 0.006) (**Figure 9**), adding to the evidence that the complex allele is more penetrant than the

---

[k] Comparison of phenotypes of patients homozygous for p.(Gly1961Glu), both without (A, B, C, D) and with (E, F, G, H) the c.769-784C > T variant. (A–C) Autofluorescence images of patients homozygous for p.(Gly1961Glu) exhibit bull's eye maculopathy lesions with characteristic smooth, uniform borders (insets, yellow arrowheads). (D) Contours of all lesions associated with homozygous p.(Gly1961Glu) without the c.769-784C > T variant. (E–G) Autofluorescence images of patients homozygous for p.(Gly1961Glu) and also harboring the c.769-784C > T variant, indicated as c.769-784C > T (+), exhibit lesion-centric flecks (white arrows) and a complex pattern of diffuse retinal pigment epithelium atrophy with no distinct border (insets). (H) (D) Contours of all lesions associated with homozygous p.Gly1961Glu without the c.769-784C > T variant. All contours are represented as right eyes, centered at the fovea (x). The position of the optic nerve for each eye is provided. Near-infrared autofluorescence, NIR-AF.

p.(Gly1961Glu) alone (**Figure 8**). Additional data were revealed after exploring the alleles in *trans* from p.(Gly1961Glu) in compound heterozygotes. The p.(Pro1380Leu) variant is frequently identified in *trans* from the complex c.[769-784C > T;5882G > A] allele. Our entire cohort of 645 bi-allelic STGD1 cases includes 17 compound heterozygotes (2.6%) of p.(Pro1380Leu) and p.(Gly1961Glu). Of these, five cases (29.4%) harbor the c.[769-784C > T;5882G > A] complex allele (**Supplemental Table 2**). Previously, the p.(Pro1380Leu) variant had been considered a 'severe' mutation.[46] Frequent pairing with the complex c.[769-784C > T;5882G > A] allele suggests that it may not be a 'severe' mutation, but a 'moderate' one instead. This was also confirmed by our most recent functional analyses, where the 'moderate' status for this variant was confirmed.[114] The explanation for the frequent occurrence of these compound heterozygotes is also likely expected, since both, the p.(Pro1380Leu) and the p.(Gly1961Glu) variants are frequent among the Ashkenazi Jewish population with MAFs 0.0023 and 0.023, respectively.

Fundoscopic examinations of these compound heterozygous patients revealed a spectrum of features consistent with previously described phenotypes of ABCA4 disease. The age of onset varied extensively (mean = 20.8 years, median = 17.0 years, range = 9–48 years, SD = 11.2 years). Measured BCVA was also variable (20/20–20/400), however ffERG testing was consistently normal in all patients (**Supplemental Table 2**). When separating patients who are compound heterozygous for p.(Gly1961Glu) and p.(Pro1380Leu) according to the presence or absence of c.769-784C > T, those without the intronic variant (n = 11) exhibited macular-confined atrophy with very few flecks proximally distributed around the central lesion (**Figure 11A–11C**). In contrast, those with the additional c.769-784C > T variant (n = 6) appeared to have a disproportionately more severe phenotype characterized by a large non-uniform coalescing

regions of chorioretinal atrophy (**Figure 11D–11F**) and an aggressive pattern of confluent flecks across the macula and well into the mid-periphery. Analysis of retinal layer thickness also revealed significant differences between both groups. Each compound heterozygous patient showed thinning of photoreceptor-attributable layers in the macula relative to thicknesses found in healthy individuals (below 95% confidence interval); however, those with the c.769-784C > T showed more thinning of both photoreceptor/RPE (**Figure 11G**) layers, indicating a profound accelertation in disease progression. Consistent with these findings were significant differences in functional loss. Compound heterozygous patients with c.769-784C > T presented with poorer BCVA (P = 0.003, **Figure 11H**) and more attenuated 30 Hz flicker responses (P = 0.007, **Figure 11I**). The exact modifying effect of the previously described[55] p.(Thr1253Met) allele, seen in 9/150 patients with the p.(Gly1961Glu), remains to be established. The disease presentation in two patients, compound heterozygous for the p.([Thr1253Met;Gly1961Glu]) and p.(Pro1380Leu) variants, was milder. They exhibited significantly delayed age of onset (P < 0.0001) compared to all other p.(Gly1961Glu) patients, foveal sparing and generally milder disease features. While this may suggest that the p.(Thr1253Met) allele is 'protective', the small number of patients with this genotype in this study does not allow us to draw definitive conclusions.
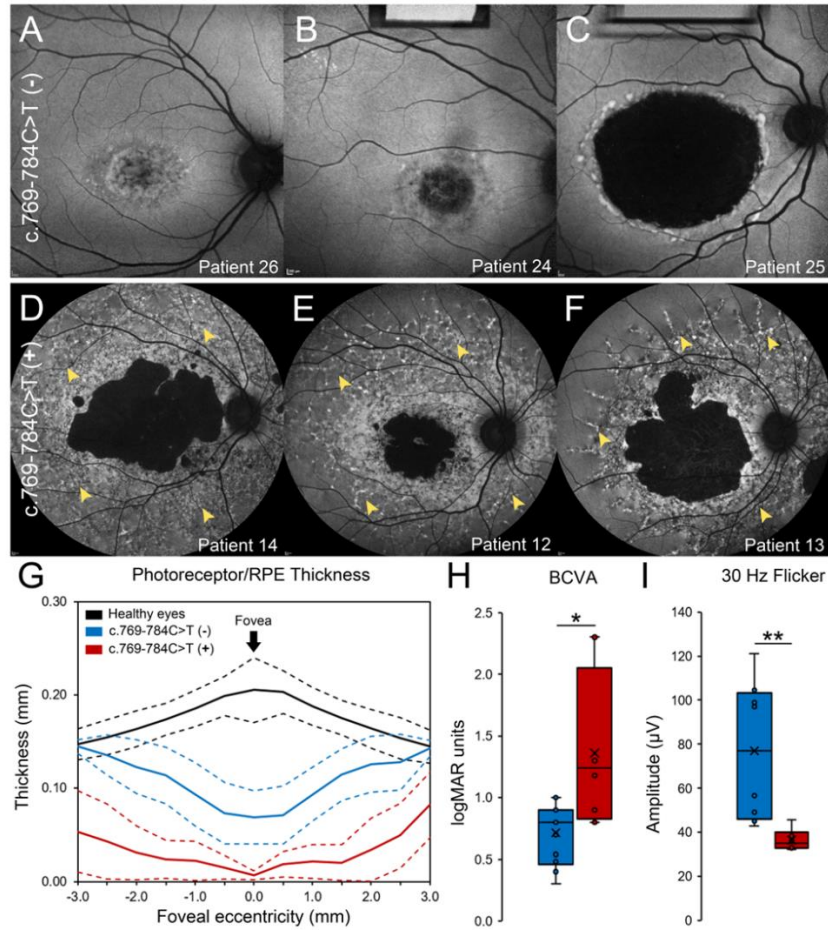
Figure 11*. Clinical phenotype of patients compound heterozygous for p.(Gly1961Glu) and p.(Pro1380Leu).*[1]

---

[1] (A–C) Autofluorescence images of compound heterozygous patients exhibit macular-confined lesions of diffuse RPE and extended chorioretinal atrophy. (D–F) Autofluorescence images of compound heterozygous patients with c.769-784C > T, indicated as c.769-784C > T (+), exhibit large, coalescing lesions of chorioretinal degeneration and confluent accumulations of pisciform flecks extending out to the posterior pole (yellow arrowheads). (G) Analysis of photoreceptor/RPE layer thickness in the macula of healthy eyes (black), eyes of patients compound heterozygous for p.(Gly1961Glu) and p.(Pro1380Leu) (blue), and eyes of patients compound heterozygous for p.(Gly1961Glu) and p.(Pro1380Leu) harboring the c.769-784C > T allele (red). Mean thicknesses (solid lines) and upper and lower 95% confidence intervals (dotted lines) 3 mm from the fovea in the nasal and temporal directions are provided for each group. (H) Statistical comparison of best-corrected visual acuity (BCVA) in compound heterozygous eyes with (red) and without (blue) c.769-784C > T (*, $P < 0.05$). (I) Statistical comparison of 30 Hz flicker amplitudes in compound heterozygous eyes with (red) and without (blue) c.769-784C > T (**, $P < 0.005$). Statistical differences were determined by non-parametric Mann–Whitney U tests.

3.3 Discussion

Our findings shed significant light on the controversy regarding the penetrance of the most frequent disease-causing *ABCA4* variant, p.(Gly1961Glu). Earlier studies had suggested that the variant is not disease-causing due to the high MAFs in some geographical regions of the world, particularly the Horn of Africa, where the MAF of the p.(Gly1961Glu) is as high as 10% in the general population.[55,57] Therefore, if the variant was fully penetrant, the disease prevalence would be at least 1:100, which is 50–100 times more than what is observed.[57] In support of this notion, we identified two homozygous for p.(Gly1961Glu) cases, who are parents of STGD1 patients and yet express no disease features (**Supplemental Figure 1**). At the same time, we studied two homozygous affected individuals in whom we did not find any other *ABCA4* variants in *cis* in the entire *ABCA4* locus since we identified definitive cis modifiers, i.e. patients who harbor additional deep intronic and coding variants on the same allele, in 8/10 affected homozygous patients. The deep intronic variant identified in six of these cases, c.769-784C > T, has been previously shown to have a splicing effect on ~30% of *ABCA4* transcripts derived from this allele, and as such, can be recognized as a modifier variant for the p.(Gly1961Glu) allele. Furthermore, a significantly larger proportion of the c.[769-784C > T;5882G > A] complex allele is found in compound heterozygous patients harboring non-deleterious variants in *trans* (21%) compared to known or expected complete LOF alleles (PVS1) (5%; FET, P = 0.006). In other words, a highly penetrant allele in trans from p.(Gly1961Glu) often offsets the requirement of c.769-784C > T, or any other *cis*-modifier, for disease penetrance. We also demonstrate that the likely more penetrant complex allele c.[769-784C > T;5882G > A] enhances disease expression in cases where the variant in trans is not considered deleterious, e.g. p.(Pro1380Leu).

Although it is likely that the penetrance of the p.(Gly1961Glu) allele requires additional disease-causing or -modifying variants on at least one of the two chromosomes in homozygosity, a few documented homozygous affected cases remain (**Supplemental Table 1**) where no other apparent disease-causing variant in cis can be found after sequencing of the entire *ABCA4* locus. It is possible that other variants such as large CNVs or *cis*-modifiers, similar to c.769-784C > T, exist, especially in specific subpopulations of European and non-European descent. Our data indicate that modifying variants are present, due to a likely founder effect, in specific subpopulations. For example, the c.[769-784C > T;5882G > A] and the c.[3758C > T;5882G > A] alleles are found in patients of Ashkenazi Jewish descent. We have identified many other, much rarer complex alleles (**Supplemental Table 1**, which are found in single cases in some populations prohibiting statistical analyses, but that are also likely specific to subpopulations. The absence of large-scale *ABCA4* screening data in such populations does not allow making definitive conclusions in this regard. For example, in our study of ABCA4 disease in South Asians,[56] we did not encounter the c.[769-784C > T;5882G > A] complex allele in any of the 16 patients with p.(Gly1961Glu), including one who was homozygous for p.(Gly1961Glu). A subsequent analysis of the locus (deep intronic) sequences for the available 11 alleles did not find any frequent modifiers, although the allele number available for analyses is (very) small. Several deep intronic variants were on the same chromosome with the p.(Gly1961Glu) in the homozygous case. Of these, only the c.4635-575 T > C variant, which is absent from gnomAD and was present on 3/11 alleles, could be considered a candidate for a modifier, although no significant splicing effects were predicted for this variant. Another variant, c.3329-240 T > C, was detected in two compound heterozygous cases, which is 14x higher than the highest gnomAD MAF, and it is predicted to create an exon splicing enhancer motif according to ESE

Finder (SC35). Therefore, this variant could also be considered a putative modifier pending the analysis in larger cohorts and functional studies *in vitro*.[35,115-117] Since rare variants on the p.(Gly1961Glu) allele have been identified in coding sequences before[55] and also in this study, similar alleles are likely to exist in deep introns. In summary, our data suggest that there are no other frequent modifiers on the p.(Gly1961Glu) allele in non-Finnish European population. Since the possible remaining *cis*-modifying variants in *ABCA4* coding and non-coding sequences are very rare, unequivocally proving the modifying effect of the remaining *cis*-modifiers from both coding and, especially, non-coding sequences would be complicated.

Finally, it is also possible that in (very) rare cases homozygous p.(Gly1961Glu) alleles without any cis-modifiers can still result in disease expression under environmental pressure, such as an advanced age, excess of vitamin A, extensive UV exposure, etc. We had suggested previously that some of the late onset ABCA4 disease patients can be mistakenly classified as AMD.[46] This is true for cases with the hypomorphic alleles, such as the p.(Gly1961Glu) and p.(Asn1868Ile) variants. From the analysis of 924 AMD cases, we identified one homozygous p.(Gly1961Glu) patient and two bi-allelic patients with the p.(Asn1868Ile) variant in *trans*, where the p.(Gly1961Glu) presented without any cis-modifiers, including the c.769-784C > T variant. Therefore, one can hypothesize that the p.(Gly1961Glu) variant without cis-modifiers, in homozygosity or in trans from other hypomorphic alleles can, in very rare cases, cause disease expression at a very advanced age.

The p.(Glu1961Glu) variant in the *ABCA4* gene is still the most enigmatic mutation. It is, by far, the most prevalent disease-causing allele, which results in a specific phenotype regardless of the allele in trans. STGD1 usually presents as a juvenile-onset macular dystrophy associated with rapid central visual impairment, progressive bilateral atrophy of the foveal retinal pigment

epithelium and the frequent appearance of yellowish flecks, defined as lipofuscin deposits, around the macula and/or in the central and near-peripheral areas of the retina.[118] While most STGD1 patients exhibit elevated autofluorescence,[22,23,103] patients with the p.(Gly1961Glu) allele show only marginal or often not significant elevation compared to normal eyes. Instead, these cases consistently present with a BEM phenotype,[119] which is defined as a circularly confined region of atrophy beginning in the central macula. In both compound heterozygous and homozygous state, the p.(Gly1961Glu) mutation results in BEM characterized by rapid ablation of foveal cone photoreceptors,[42] but otherwise slow progression and slow lipofuscin accumulation.[22,55] Homozygous p.(Gly1961Glu) patients with c.769-784C > T exhibited a more complex pattern of degeneration characterized by a heterogeneous border and the presence of flecks. Conversely, the lesions of homozygous p.(Gly1961Glu) patients without c.769-784C > T were distinctly more uniform and presented with no apparent flecks indicating a marked difference in disease progression (**Figure 9**).[120-122] The effect of c.769-784C > T was more evident in a stronger mutational background, such as in compound heterozygous cases for p.(Gly1961Glu) and p.(Pro1380Leu), where patients carrying the complex allele exhibited more advanced chorioretinal atrophy, fleck accumulation and visual deterioration (**Figure 10**).

In patients, the c.[769-784C > T;5882G > A] complex allele occurs on only 15% of all p.(Gly1961Glu) chromosomes; therefore, it is not necessary for disease expression in most compound heterozygous cases. While the allele in *trans* in those cases is preferably deleterious, or at least 'severe', many patients, especially of late disease onset, have moderate or even mild variants in *trans*. It is estimated that the c.769-784C > T variant affects splicing of ~30% of mRNA derived from this allele; therefore, it causes moderate reduction of available functional ABCA4 protein. Indeed, we also show a marked difference in clinical severity in patients who

harbor this variant with p.(Gly1961Glu), especially in compound heterozygosity (**Figure 10**). The exact functional consequences of p.(Gly1961Glu) are not known. Expression levels of p.(Gly1961Glu) mutant are comparable to wt ABCA4 in transfected HEK293T cells, and this variant is retained in the membrane of vesicles as is the wt protein.[114] However, this mutant cannot bind N-Ret-PE and shows precipitous loss of ATPase activity;[112,114] i.e. the variant is devoid of most any functional activity, which likely result from adverse effects of detergent solubilization of this mutant protein.[114]

Limitations of our study include relatively small numbers of homozygous p.(Gly1961Glu) cases and patient enrollment in two tertiary retinal disease clinics at Columbia University (New York, NY) and The Chicago Lighthouse (Chicago, IL). Therefore, patient ethnicities are representative of two large US metropolitan areas and not of all populations around the world.

In summary, our study explains the paucity of observed homozygous p.(Gly1961Glu) cases in STGD1. We show a statistical enrichment of additional exonic and deep intronic variants, especially c.769-784C > T, on the p.(Gly1961Glu) allele of most homozygous and some compound heterozygous patients and, accordingly, a difference in the disease phenotype of patients with each respective genotype. Taken together, these data indicate that the singular effect of p.(Gly1961Glu) in homozygosity may not exceed the threshold for pathogenicity without the addition of a cis-acting modifier such as c.769-784C > T. In compound heterozygosity, the penetrance of a single p.(Gly1961Glu) allele may be dependent on, or influenced by, the relative strength of the opposite allele.

3.4 Materials & Methods

3.4.1 Patient cohort and clinical evaluation

All study subjects were consented before participating in the study under the protocol #AAAI9906 approved by the Institutional Review Board at Columbia University. The study-related procedures adhered to tenets established in the Declaration of Helsinki. Complete ophthalmic examinations were provided by a retinal specialist, including slit-lamp and dilated fundus examinations. Clinical assessments were made from clinical examination notes, retinal imaging data and research questionnaires. Spectral domain-optical coherence tomography (OCT) scans and corresponding infrared reflectance fundus images were acquired using a Spectralis HRA + OCT (or HRA + OCT) (Heidelberg Engineering, Heidelberg, Germany). Fundus autofluorescence (AF) images were obtained using a confocal scanning-laser ophthalmoscope (Heidelberg Retina Angiograph 2, Heidelberg Engineering, Dossenheim, Germany). Fundus autofluorescence (AF) images were acquired by illuminating the fundus with an argon laser source (488 nm and 787 nm excitation).[19,123] Full-field electroretinograms (ffERG) were recorded using the Diagnosys Espion Electrophysiology System (Diagnosys LLC, Littleton, MA, USA). Prior to acquisition, pupils were maximally dilated and measured before testing using tropicamide (1%) and phenylephrine hydrochloride (2.5%); and the corneas were anesthetized with proparacaine 0.5%. Silver impregnated fiber electrodes (DTL; Diagnosys LLC, Littleton, MA) and Burian-Allen contact lens were used with a ground electrode on the forehead. Normative amplitude and implicit time ranges for each stimulus were provided by the Diagnosys system software. All procedures were performed using extended testing protocols outlined by the International Society for Clinical Electrophysiology of Vision (ISCEV) standard.[108] ffERG classifications were assigned according to electrophysiological attributes described by Lois et

al.[109] Group 1 is characterized by no detectable loss in scotopic or photopic function; Group 2 is characterized by photopic loss, but normal scotopic function; and Group 3 exhibits deterioration of both scotopic and photopic function.

3.4.2 Retinal imaging analysis

Retinal layer thickness was assessed in single 9-mm high-resolution SD-OCT scans through the fovea. Total receptor+ (TREC+) was defined as the distance between the Bruch's membrane/choroid interface and the inner nuclear layer (INL)/outer plexiform layer (OPL) boundary.[107] The boundaries of TREC+ were identified using the automated segmentation program in the HEYEX software (Heidelberg Engineering, Heidelberg, Germany) with manual correction when necessary. Mean thickness, including upper and lower 95% confidence intervals, for all compound heterozygous Gly1961Glu and Pro1380Leu patients were measured at 0.5 mm interval positions in the temporal and nasal directions from the fovea (13 positions total) (**Figure 6**). TREC+ thickness was compared to an age-matched (age 30–60 years) cohort of healthy controls. Thickness plots (mean of grader measurements) were generated in R (https://www.r-project.org) through RStudio (https://www.rstudio.com). Measurements were performed by two independent graders (WL and PYS). Intraclass correlation coefficients (ICC), calculated using the irrICC package (https://cran.r-project.org), showed excellent intergrader agreement for patients (ICC > 0.9667) and controls (ICC > 0.9996). Lesion contour analysis was performed on 30° fundus autofluorescence images collected with the Spectralis HRA + OCT. Lesion contour analysis was performed on 30° fundus autofluorescence images collected with the Spectralis HRA + OCT. All images were normalized according to the internal scaling factor of the image and the border of each lesion was traced using the magic wand tool in the FIJI software (National Institute of Mental Health, Bethesda, MD, USA). Left eyes were flipped

43

along the vertical meridian so that both eyes are represented together (i.e. the nasal retina is left of center for both eyes) aligned at the fovea.

3.4.3 Sequencing analysis

Sequencing of the *ABCA4* gene was performed using 454 Titanium method (Roche) or Illumina TruSeq Custom Amplicon protocol.[34,79] Sequencing of the entire *ABCA4* genomic locus in genetically confirmed STGD1 patients harboring p.(Gly1961Glu) variants was performed using single molecule molecular inversion probes (smMIPs) library preparation and the Illumina NextSeq500 sequencing platform.[35] The variants' segregation with the disease in available families was analyzed by Sanger sequencing.

All variants and their allele frequencies were compared to the gnomAD data set (http://gnomad.broadinstitute.org; accessed November 2020). Variants with similar frequencies to the c.5882G > A variant in the study cohort were determined to be on the same major c.5882G > A haplotype. Remaining variants were assessed for their phase, allele frequency and possible pathogenicity. The possible effect of non-coding *ABCA4* variants on splicing was analyzed using five different algorithms (SpliceSiteFinder, MaxEntScan, NNSPLICE, GeneSplicer and Human Splicing Finder) via Alamut software (http://www.interactive-biosoftware.com) and SpliceAI.[99] New, less frequent haplotypes harboring the c.5882G > A variant and modifying *cis*-variants were identified.

Chapter 4 Characterization of Hyper-Pigmentary Chorioretinopathy Phenotype of *ABCA4*

4.1 Introduction

Pathogenic variation in *ABCA4[1]* is the underlying cause of a group of inherited retinal degenerative disorders ranging from severe early-onset cone-rod dystrophy to adult-onset maculopathy.[2] While each of these conditions are clinically distinct in many ways, all are, etiologically, macular diseases—that is, disease onset and degeneration begins in the macula and progresses outwards toward the peripheral retina. The symptomatic manifestation for patients with maculopathy begins with a central blind spot that enlarges as the disease progresses over time (**Figure 12**). Interestingly, pathogenic variants in *ABCA4* have also been described in patients purported to have retinitis pigmentosa,[124-139] a condition that is the etiological inverse of maculopathies with respect to cellular degeneration (**Figure 12**). Retinitis pigmentosa (RP) is an umbrella term denoting a group of autosomal recessive, dominant and X-linked disorders caused by more than 60 different genes (https://web.sph.uth.edu/RetNet/).[140,141] Most RP-associated encode proteins that play an essential role in the structure and function of rod photoreceptors in the retina.[140,141] Rods are specialized neurons responsible for vision in low light while cones are responsible for vision in bright light and color distinction. The anatomical distribution of rods and cones in the human retina are such that rods are more numerous in the periphery whereas cones more numerous towards the center in the macula. Dysfunction of RP-associated genes results in the degeneration of rods followed by cones in later stages of the diseases.[140,141] As a result, all patients with RP experience the characteristic symptom of nyctalopia and, contrary to maculopathies, gradual "constriction" of the visual field over time (**Figure 12**).
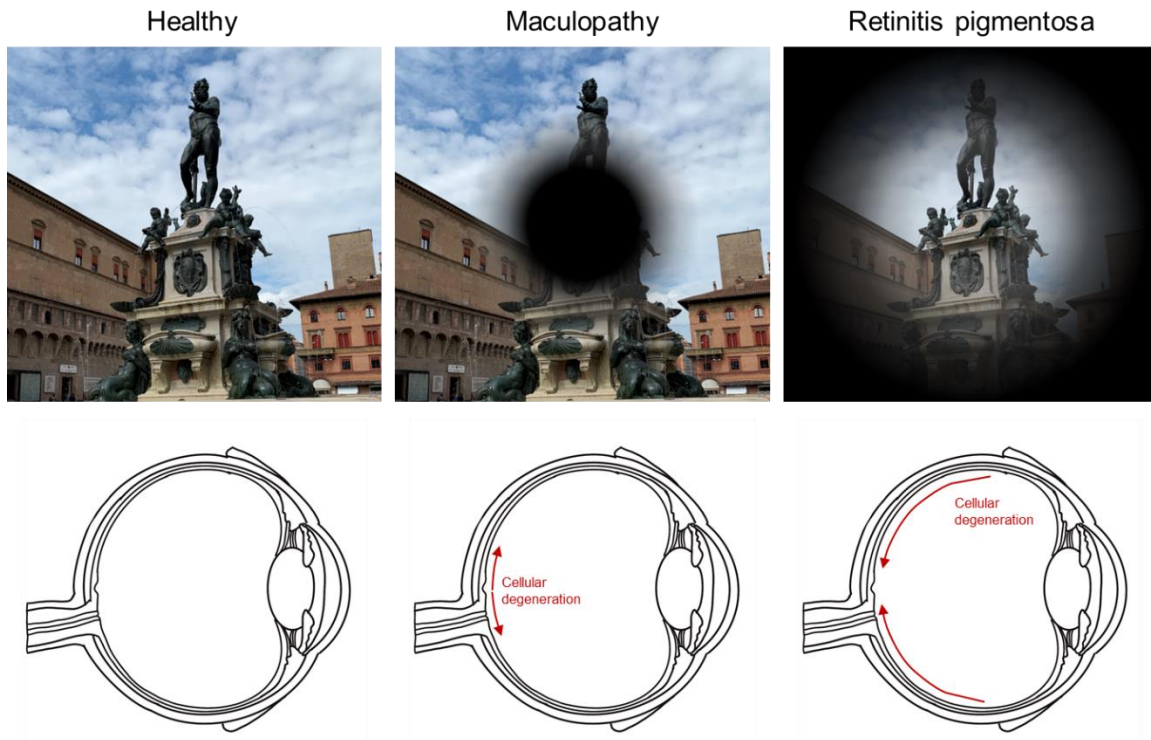
Figure 12. Consequent vision loss and the direction of degeneration in maculopathies compared to "typical" retinitis pigmentosa.

From a clinical perspective, the inclusion of RP among the other phenotypic entities of *ABCA4* may seem unusual; however, there are several genes known to cause both RP and maculopathy (or related cone-rod dystrophy) disorders including *PROM1*,[59,142-144] *RPGR*,[145,146] *PRPH2*,[147-149] *RP1L1*,[150,151] *BEST1*,[152,153] *AIPL1*,[154] *GUCY2D*,[155,156] *CRX*,[157,158] among others. Furthermore, ABCA4 is expressed in both cone and rod photoreceptors in the retinae of humans and other vertebrates.[3-5] A meta-analysis of the RP19 literature however reveals, at best, a questionable connection between *ABCA4* and "typical" RP. In many reports, for instance, the degeneration in patients involves the macula. In the earliest thorough description of RP19,[125] several of the reported patients simultaneous had degeneration in the macula. Several other subsequent reports similarly characterized RP19 in such a manner using paradoxical terminology such as "RP with macular involvement"[133,136,138] or "retinitis pigmentosa inversa".[139] Patients with typical RP

46

have also been included in such studies although, tellingly, such cases are often genetically unsolved or monoallelic (heterozygous) for pathogenic variants in *ABCA4*.[137,159,160]

Despite this precarious evidence, RP continues to be regarded as a phenotypic entity of *ABCA4* in both genetic and clinical databases OMIM (https://www.omim.org/entry/601691), ClinVar (https://www.ncbi.nlm.nih.gov/clinvar), Orphanet (https://www.orpha.net/), etc. Accurate curation of these databases and the corresponding literature is imperative as they are essential resources for not only clinicians, but also other professionals in the field including clinical geneticists interpreting genetic testing results and basic scientists developing therapies for retinal diseases. The aim of this chapter is to comprehensively characterize the clinical phenotype and underlying genetic disposition of patients with RP19 to conclusively establish an empirical consensus on the relationship between RP and *ABCA4*.

4.2 Results

4.2.1 Disease prevalence, fundus characteristics and oculomotor abnormalities

A cohort of 23 individuals (from 19 families) out of our database of genetically-confirmed ABCA4 disease was identified (5.5%, 23/415) as having diagnostic features of RP19 (1) bone-spicule pigment deposition, (2) arteriolar attenuation and (3) waxy disc pallor (**Figure 13A-13F**). Clinical, demographic, and genetic characteristics are summarized in **Supplemental Table 3**. Visual deterioration was reported in within the first decade of life in ~3/4 of patients (mean = 8.5 years, range 6-13 years, IQR = 2) and all were differentially diagnosed with RP at some point in their clinical histories. In addition, nearly half also had ocular motor issues (**Supplemental Table 3**). Nine patients had acquired nystagmus (37.5%) of which 7 were horizontal (P1, P3, P5, P7, P8, P10 and P16) and 1 was rotary (P17). Four patients had intermittent (P3, P9 and P17) and persistent (P4) gaze deviations, specifically right exotropia. Strabismus was noted in

P16. Six patients (P3, P9, P10, P15, P17 and P18) had posterior subcapsular cataracts all of where detected between 20-30 years of age. Ocular histories in several other patients included ocular migraines and discomfort, photopsias and light-aversion (photophopia).
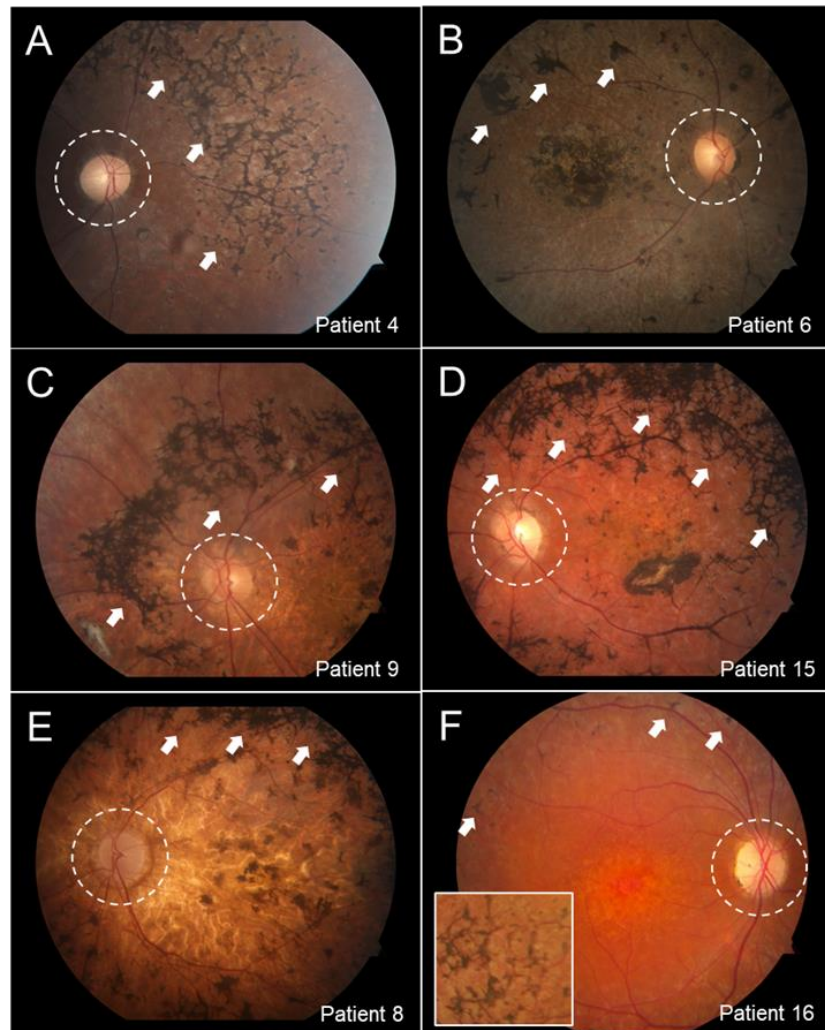
Figure 13. *Funduscopic documentation of retinitis pigmentosa-related features in patients with ABCA4-associated retinitis pigmentosa 19 (RP19, #601718).*

4.2.2 Non-recordable and maculopathy-specific full-field electroretinogram responses

Full-field electroretinogram testing (ffERG) reliably measures the characteristic pattern of early rod degeneration and is therefore a useful tool for distinguishing RP from other macular or cone-

driven retinal diseases such as *ABCA4*-associated maculopathy. Cone and rod ffERG were

recorded in 16 out of the 24 patients, in all of whom revealed generalized retinal dysfunction

(**Figure 14**). Notably, none of the recordings in 16 patients exhibited a rod-cone ERG pattern

consistent with typical retinitis pigmentosa. Both rod-associated scotopic and cone-associated

photopic responses were non-recordable in 3 patients (P3, P6 and P8). Of the remaining 12

patients, photopic responses were extinguished in all except 2 patients (P14 and P19). Under

photopic conditions, light-adapted 3.0 cd.s.m^{-2} (LA 3.0) single flash B-wave amplitudes in P14

were slightly below normal: 51.0 µV OD and 45.0 µV OS. LA 3.0 B-waves amplitudes in P19

were normal in both eyes (64.1 µV OD and 108.7 µV OS). The LA 3.0 cd.s.m^{-2} 30 Hz flicker

(peak-to-trough) amplitudes were likewise reduced in P19 (38 µV OD and 25 µV OS) and

normal in P19 (67.3 µV OD and 69.8 µV OS). Both LA 3.0 single flash and 30 Hz flicker

responses were significantly delayed in P14 and P19. Following dark adaptation, rod-specific

DA 0.01 and maximal (white flash) responses were normal in only one patient (P19).

Recordable but attenuated DA 0.01 cd.s.m^{-2} B-waves in P4, P5, P10, P11, P12, P13, P14, P15,

P16 and P17 ranged from 29.0 µV to 100.0 µV in both eyes. DA 3.0 cd.s.m^{-2} maximal

(combined rod and cone) responses varied from 35.0 to 209 µV. No patients had electronegative

B-to-A wave ratios.
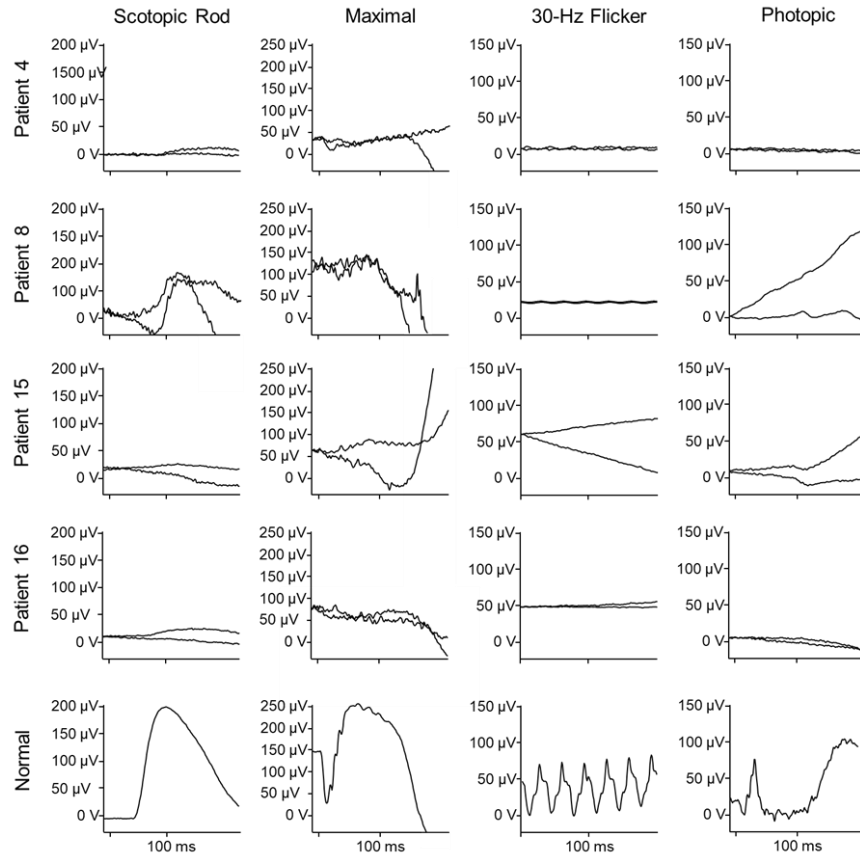
Figure 14. *Full-field eletroretinogram recordings in patients with ABCA4-associated RP19.*[m]

4.2.3 Degeneration of the macula and pathognomonic ABCA4 disease features on ultra-widefield imaging

The onset of degeneration beginning in the peripheral retina typically spares central vision in patients with RP until very late in disease;[141] however, all 24 patients had profound central vision loss attributable to extensive atrophy in the macula. Best-corrected visual acuity (BCVA) in 14 patients was hand motion (HM) to counting-fingers (CF) and not correctable beyond 20/200 in all others (**Supplemental Table 3**). Collectively, the visual prognosis in these patients was

---

[m] Representative ffERG traces from healthy, unaffected eye are shown in bottom row for comparison.

significantly worse than all other ABCA4 disease patients combined (n=391) (P<0.0001, log-rank test) (**Figure 15**).
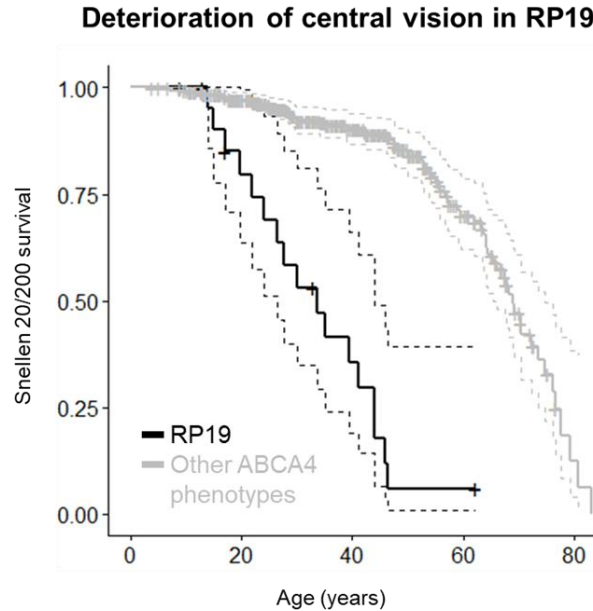
Figure 15. *Kaplan-Meier curve showing central vision loss in RP19 compared to other ABCA4 disease patients.*

To assess the spatial distribution of degeneration in peripheral retina, ultra-widefield (UWF) pseudo-color (red and green reflectance) and medium wavelength (532-nm excitation) autofluorescence images were acquired. Outside the macula (**Figure 16A-16D, dotted circle**), disease changes were localized primarily to the posterior pole—although in some cases extended as far as the equatorial regions (P1, P8, P10 and P18). Nevertheless, UWF imaging was consistent with a macular-driven or cone-rod dystrophy in all cases.
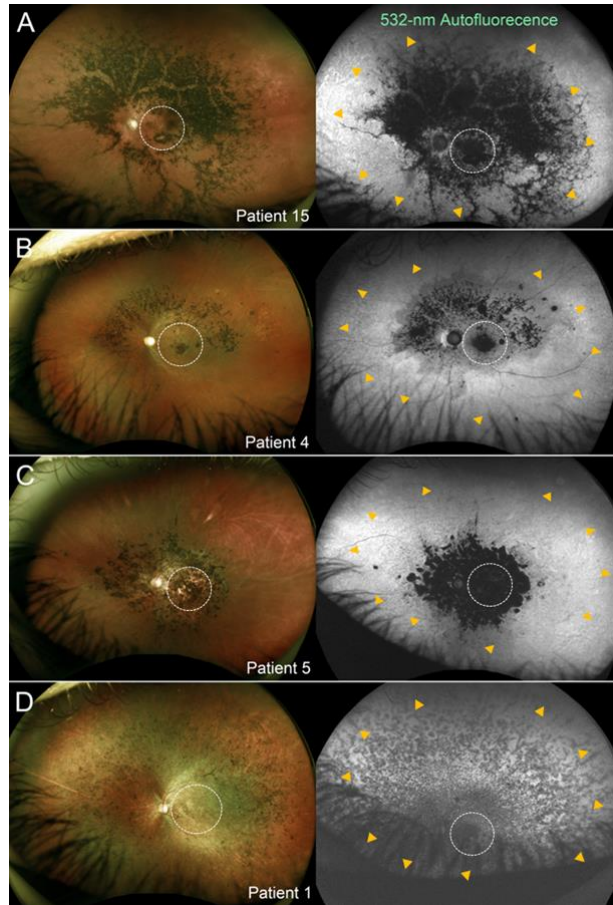
Figure 16. *Macular-based degeneration in RP19.*

Consistent with the underlying *ABCA4* variation in these patients, the fundus in less progressed

cases exhibited a "beaten bronze" appearance indicative of lipofuscin accumulation.[161,162] To

better visualize this and other ABCA4-associated features, short-wavelength autofluorescence

(SW-AF) imaging was acquired in all patients. Confluent lipofuscin-laden fleck deposits were

apparent in non-atrophic regions (particularly in the periphery) (**Figure 17A and 17B, insets**).

Residual peripapillary sparing was detectable in most (19/24, 79%) (**Figure 17A and 17B**) but

not all cases (**Figure 17C and 17D**).
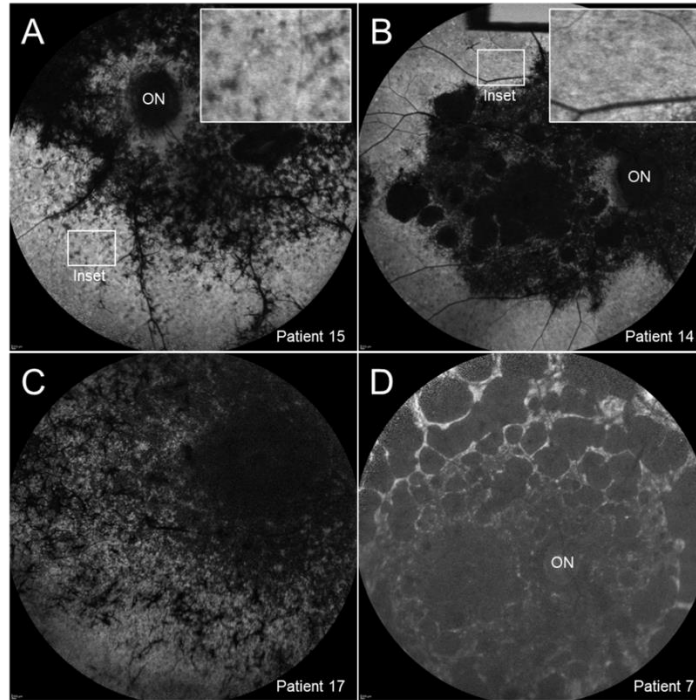
Figure 17. Detection of lipofuscin-laden flecks and peripapillary sparing in RP19 patients.

4.2.4 Enrichment of homozygosity and familial consanguinity

Direct sequencing of the *ABCA4* locus identified two pathogenic variants in *ABCA4* in all cases (**Supplemental Table 3 and Table 1**). Although a portion of the alleles were null or protein-truncating variants, there was a higher-than-expected fraction of missense variants represented given the severity of this phenotype. Excluding familial alleles, genotypes consisting of two alleles with missense variants were found in 10 out of 19 (52.6%) unrelated cases. This biallelic missense percentage exceeds that of other comparatively severe ABCA4 disease phenotypes such as rapid-onset chorioretinopathy (1/15 (6.7% biallelic missense).[110] In addition, 13/19 (68.4%) of the unrelated genotypes in this cohort are homozygous. This percentage is significantly higher (P<0.0001, two-tailed FET) compared to the original cohort of 415 screened in which the percentage of homozygous genotypes was 5.4% (21 homozygous allele genotypes out of 391 unrelated genotypes).

Table 1. *Summary of predicted pathogenicity and classification criteria of all ABCA4 variants identified in patients with RP19*

cDNA (NM_000350.2)	Protein	Allele count	VEP annotation	dbSNP	Allele frequency	M-CAP	REVEL	Eigen	CADD	SpliceAI
c.93G>A	p.(Trp31*)	2	Stop gained	rs1191816747	3.98E-06	-	-	0.81	37	-
c.223T>G	p.(Cys75Gly)	2	Missense	-	Absent	0.666	0.909	0.85	25.5	-
c.571-1G>T	p.(?)	2*	Splice acceptor	rs61748533	Absent	-	-	1.13	25.7	Acceptor loss at -1bp (Δscore = 99%)
c.1804C>T	p.(Arg602Trp)	2	Missense	rs61749409	4.38E-05	0.498	0.934	0.59	35	-
c.2160+1G>C	p.(?)	2*	Splice donor	rs61749427	3.98E-06	-	-	1.23	26.7	Donor loss at +1bp (Δscore = 100%)
c.2894A>G	p.(Asn965Ser)	2	Missense	rs201471607	1.35E-04	0.551	0.779	0.61	25.9	-
c.2947A>G	p.(Thr983Ala)	1	Missense	rs1368508052	3.98E-06	0.691	0.874	0.48	24.1	-
c.3210_3211dup	(p.Ser1071CysfsTer14)	1	Frameshift	rs387906385	1.99E-05	-	-	-	-	-
c.3259G>A	p.(Glu1087Lys)	1	Missense	rs61751398	1.19E-05	0.738	-	1.13	32	-
c.3988G>T	p.(Glu1330*)	6*	Stop gained	-	Absent	-	-	0.08	36	-
c.4139C>T	p.(Pro1380Leu)	1	Missense	rs61750130	2.34E-04	0.391	0.867	0.7	28.1	-
c.4667G>A	p.(Arg1556Lys)	2	Missense	rs1385119665	Absent	0.931	0.852	0.81	33	-
c.4849-1G>A	p.(?)	1	Splice acceptor	rs61750156	2.39E-05	-	-	1.21	34	Acceptor loss at -1bp (Δscore = 85%); Donor loss at -1170 (Δscore = 65%)
c.4919G>A	p.(Arg1640Gln)	1	Missense	rs61751403	3.18E-05	0.343	0.771	0.82	33	-
c.5018+2T>C	p.(?)	2	Splice donor	rs61750562	Absent	-	-	1.15	25.5	Donor loss at +2bp (Δscore = 93%); Donor gain at -99bp (Δscore = 29%)
c.5186T>C	p.(Leu1729Pro)	1	Missense	rs61750567	4.18E-06	0.346	0.79	0.58	28.4	-
c.5318C>T	p.(Ala1773Val)	4	Missense	rs760549861	7.56E-05	0.123	0.79	0.83	32	-
c.6088C>T	p.(Arg2030*)	2	Stop gained	rs61751383	2.78E-05	-	-	0.54	45	-
c.6112C>T	p.(Arg2038Trp)	2	Missense	rs61750643	1.59E-05	0.331	0.941	0.75	31	-
c.6658C>T	p.(Gln2220*)	1	Stop gained	rs61753046	5.17E-05	-	-	1.06	48	-
c.[66G>A;859-9T>C]	p.(?)	1	-	-	-	-	-	-	-	-
c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	5*	-	-	-	-	-	-	-	-
c.[4594G>A;5603A>T]	p.([Asp1532Asn;Asn1868Ile])	2	-	-	-	-	-	-	-	-

In-depth family histories collected from each patient also revealed familial consanguinity in 12/13 of these unrelated homozygous patient/genotypes (**Table 1, Figure 18**). The consanguineous families of two patients also included additional individuals with apparent retinal dystrophies. The older son of P1 (who married her paternal second cousin) is affected with typical retinitis pigmentosa (**Figure 18B, Figure 20C**). The family history of P18 included a paternal uncle with unspecific retinal dystrophy, however, this individual was not examined (**Figure 18D**).
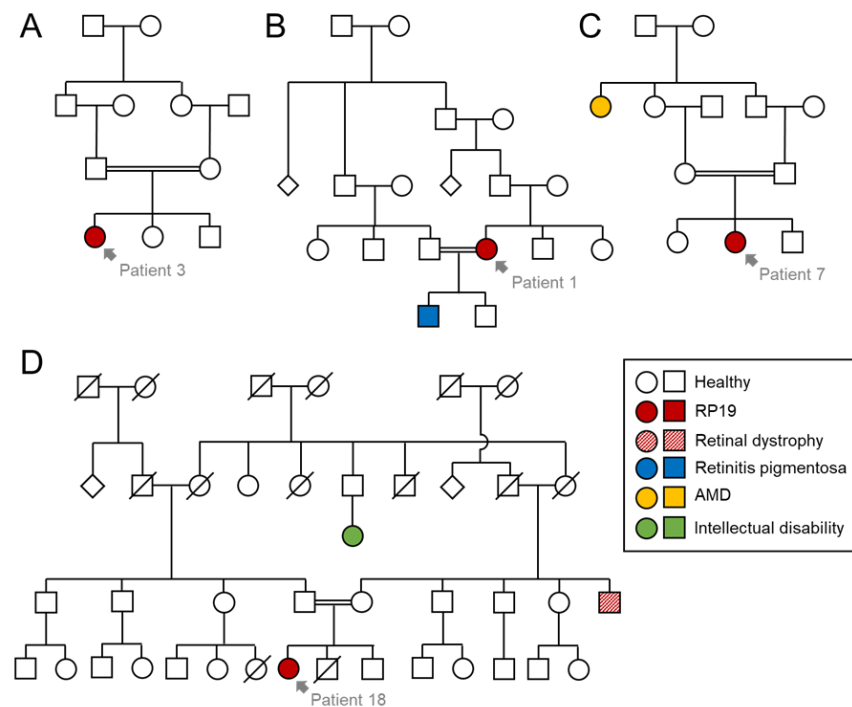
Figure 18. *Consanguineous pedigrees with RP19.*

The prevalence of consanguinity in this group and several other underlying observations led us to question whether this unusual disease phenotype may be modified by or the result of additional pathogenic variation outside of *ABCA4*. It is apparent that the core etiology of RP19 is not

typical RP; however, most of the auxiliary features expressed by this group—bone-spicule pigment deposition, nystagmus, early-onset cataracts, among other-- are nevertheless considerable phenotypic deviations from ABCA4 disease. Moreover, the predominant composition of missense alleles in this cohort is incongruous with the severe nature of this phenotype relative to other ABCA4 disease phenotypes.

4.2.5 Whole-exome sequencing analysis

To identify possible variation outside of the *ABCA4* locus, we performed whole exome sequencing (WES) on 12 patients and relatives (when available). Given the consanguinity in this cohort, we first tried to identify a shared locus (or loci) between multiple patients by identifying the presence of consecutive homozygous regions or "runs of homozygosity" (ROH). As expected, consanguineous cases had, on average, larger ROH (255.9 Mb) than non-consanguineous cases (62.5 Mb). There were no shared ROH regions between all 12 patients nor among only the consanguineous cases at the default minimal percentage of homozygous variants (MPHV) threshold of 88%. The only shared ROH region among multiple patients was a 1.05 Mb region at chromosome 15 at a MPHV threshold of 50% in 5 patients (P1, P2, P7, P9 and P10) (**Figure 19A**). This peak was no longer detectable after raising the MPHV threshold back to 88% (**Figure 19B**). This shared 1.05 Mb ROH encompassed 3 genes, *RBFOX1*, *METTL22* and *ABAT*, neither of which are known to be associated with retinal disease. There were also no obviously pathogenic variants in these 3 genes in the other 9 patients.
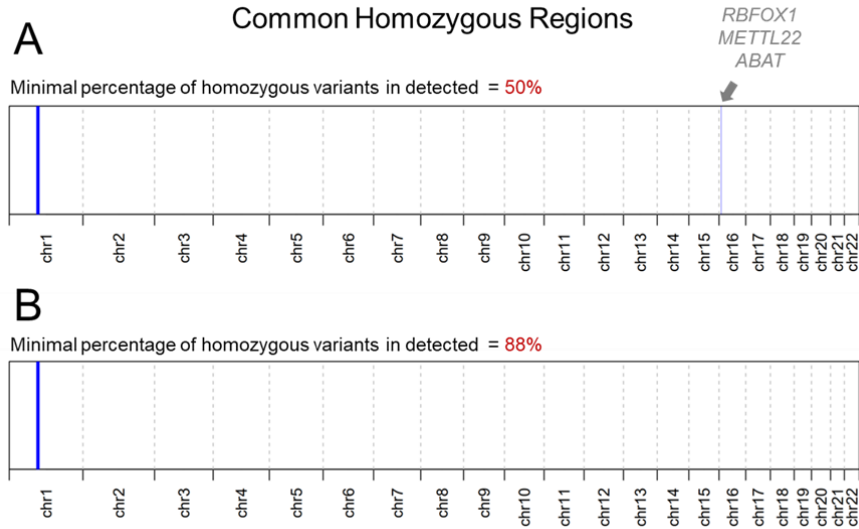
Figure 19. Runs of homozygosity (ROH) analysis 5 patients with RP19.

Given the absence of a possible shared locus between multiple patients, we analyzed cases separately for pathogenic variants in individual candidate genes. We generated an annotated list of variants for each patient for further analysis based on the following initial criteria: (1) minor allele frequency (MAF) ≤ 0.5% in gnomAD (general population; last accessed January 2023), (2) detected in coding, canonical splice site (+/- 2 nucleotides from the exon boundary) regions (3) missense, nonsense or frameshift coding effect and (4) found in retinal disease (RetNet, https://web.sph.uth.edu). The average number of variants found in each patient ranged from 7 to 43 (mean = 16.6). Putatively pathogenic variants in a candidate gene (other than *ABCA4*) were only found in one patient, P9, who was homozygous for the missense variant, c.671C>T (p.Pro224Leu) (MAF = 0.0000437) (CADD-PRED = 28.6) in the *ROM1* gene, which encodes the rod outer segment membrane protein. *ROM1* has been previously reported to cause digenic RP[163] and, in fact, this specific variant has been reported to be causal in an individual with ADRP.[164] We also have two cases (one published (Ma et al.)[165]) in which the homozygous

frameshift variant c.712delC (p.Leu238Cysfs*78) causes a late-onset *ABCA4*-like maculopathy phenotype. We re-analyzed the list variants for each patient increasing the MAF filter to ≤2%, covering non-coding variants up to +/- 10 and extending the ontology criteria to include genes known to be expressed in retinal tissue but found no additional candidate genes.

Given the unique family history of P1, (**Figure 18B**) we also performed and analyzed WES in her affected son (P1.1). We did not find any RP-related genes that were maternally inherited however there were several paternally inherited variants including a heterozygous canonical splice site variant, c.119-2A>C, in *NR2E3* (**Supplemental Table 4, Figure 20**). The functional effect of this specific variant (Exon 2 skipping) has been described in patient with enhanced S-cone syndrome (ESCS, MIM# 268100),[166] the phenotype primarily associated with *NR2E3*; however, families with both autosomal recessive and dominant RP have also been described.[167,168] This variant is unlikely to be causal since the RP phenotype in the son is inconsistent with autosomal dominant RP and the father from whom the variant segregated is unaffected.
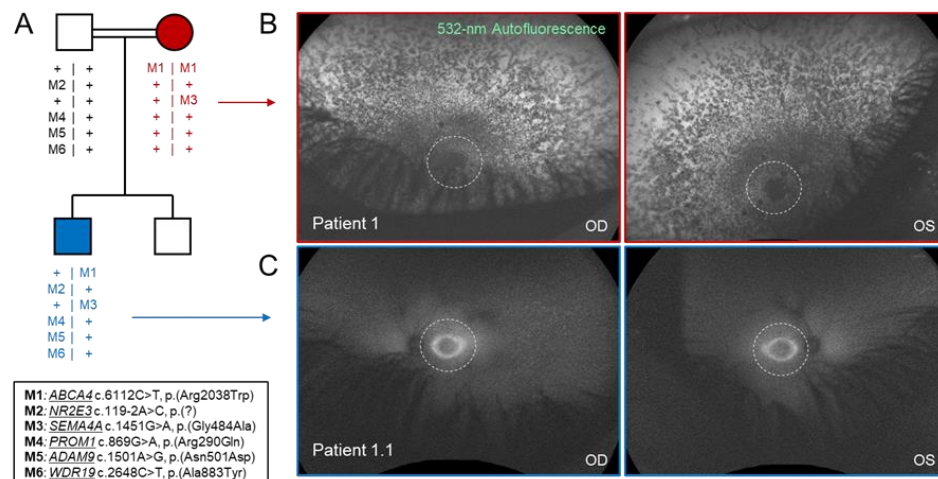
Figure 20. Haplotypes segregating in a family segregating RP19 (P1) and "typical" RP.

4.3 Discussion

This study had two main goals. First, we conclusively determined that RP19 is not in fact typical RP but rather a severe maculopathy with secondary RP- and LCA-like features. Some of these include ocular motor issues like nystagmus and early-onset posterior subcapsular cataract (most common type associated with RP).[169] Like other cases of LCA and RP, this nystagmus is likely not congenital but acquired given the pattern seen in our patients. While there is no known connection between ABCA4 and extra-ocular issues, the underlying cause in this case may be psychophysical, wherein the early precipitous loss in central vision results in issues with fixation. Like other cases, this may be treatable using established low vision therapy. There are no known connections between *ABCA4* variants and the formation of early onset PSC. Likewise, the etiology of PSC in RP is also uncertain although elevated aqueous flare.[170] Other risk factors include diabetes mellitus, hypertension, high myopia, asthma, history of steroid intake, and intraocular inflammation however none are certain.[171,172] Moving forward we propose a revising the RP19 nomenclature for this phenotype to "hyper-pigmentary chorioretinopathy" (HPC).

The presence of the RP "triad" of features in ABCA4 disease is not entirely unusual. Both thinning of the retinal vascular arcades and optic disc pallor are commonly observed in very advanced stages of degeneration in the fundus. Pigmentary deposition is also observed (albeit infrequently) in ABCA4 disease. This specific pigmentary deposit however localizes exclusively with active lesions of atrophy and are less morphologically distinct (i.e., not bone-spicule shaped in appearance).[173,174] The bone-spicule pigment type that is pathognomonic to RP and seen in HPC is exceedingly rare in ABCA4 disease. The etiology of bone-spicule pigment (BSP) formation is largely unknown however the current understanding is that it originates from deteriorating RPE and migrates to the surface of the fundus.[175,176] Interestingly, despite it being

a diagnostic hallmark, BSP deposition is not a ubiquitous feature in all RP. Large-scale meta-analyses have found that BSP are found more common among RP associated with photoreceptor-specific genes. This has led to the hypothesis that BSP migration from RPE is contingent upon the rapid death of adjacent photoreceptors, which would otherwise impede its anterior movement toward the surface of the retina. Extending this theory to ABCA4 disease opposes the view that, at least in the HPC sub-phenotype, RPE death occurs subsequently to photoreceptor death. Further studies will be needed to confirm this mechanism, particularly the precise genetic etiology of HPC.

While all HPC patients in this study have two confirmed pathogenic *ABCA4* alleles, three specific observations support hypothesis that this rare and unusual phenotype may be modified by additional variation beyond the *ABCA4* locus. First are the presence of auxiliary features which are not characteristic of ABCA4 disease. Second, is the discordantly high proportion of biallelic missense genotypes in this cohort (52.6%). The severity of missense alleles can be comparable to null alleles, notably those that occur in *trans* to hypomorphic allele genotypes proportion in HPC is profoundly higher than other severe ABCA4 diseases phenotypes (cone-rod dystrophies, advanced choriocapillaris degeneration and rapid-onset chorioretinopathy (ROC)).[110] The stronger evidence for the consequential presence of additional variation is the enrichment of consanguineous individuals. The inherent risk of expressing recessive diseases among consanguineous families is a known phenomenon and indeed, families segregating multiple affected relatives and individual with co-morbidities have been widely described.[177,178] It is important to note, however, that our study does not provide direct evidence for this phenomenon in HPC. We can just make the preliminary conclusion that if such variation does exist, it is likely to vary between individuals. The scope of our analyses can also be extended to

non-coding variation and structural variants. Pathogenic deep intronic and structural variants are rare but still make up a sizeable portion of disease in the population.[179] There is growing evidence supporting the role of mutational burden on phenotypic diversity. The small number of individuals with HPC and diverse ethnic makeup of our cohort would make association testing challenging.

In summary, we have comprehensively characterized the rare ABCA4 disease phenotype, erroneously categorized as retinitis pigmentosa 19 (MIM# 601718). This disorder has diagnostic features of typical RP such as bone-spicule pigment deposition, early-onset cataracts and acquired nystagmus, however, it is etiologically a cone-rod disorder with considerable degeneration which begins in the macula within the first decade of life. We propose revising the retinitis pigmentosa 19 nomenclature to "hyper-pigmentary chorioretinopathy" to avoid confusion in the field.

4.4 Materials & Methods

4.4.1 Study subjects and ethics approval

Written and signed informed consent to participate in this study were obtained from all individuals described in this study. All procedures were performed according the institutional review board-approved protocol #AAAI9906 at Columbia University Medical Center. The study adhered to all tenets set out in the Declaration of Helsinki. All patients enrolled in the study underwent a complete ophthalmic examination, which included slit-lamp and dilated-fundus examinations, best-corrected visual acuity (Snellen) measurement and assessment of clinical histories.

4.4.2 Electroretinogram testing

Full-field electroretinograms (ffERG) were recorded using the Diagnosys Espion (E3) Electrophysiology System (Diagnosys LLC, Littleton, MA, USA) in accordance with the recommended protocol of the International Society for Clinical Electrophysiology of Vision (ISCEV).[180] Pupils were maximally dilated to a minimum of 7-8 mm in diameter using topical 1% tropicamide and 2.5% phenylephrine prior to testing. Dawson-Trick-Litzkow (DTL) electrodes (silver impregnated fiber) over the cornea with ground electrodes attached to the forehead were used in all patients. Normative amplitude and implicit time ranges for each stimulus were provided by the Diagnosys system software.

4.4.3 Retinal imaging

Imaging across all the modalities was conducted after maximum pupil dilation (>7 mm) with 1% tropicamide and 2.5% phenylephrine hydrochloride. Short-wavelength (488 nm) and near-infrared autofluorescence (787 nm) images as well as spectral domain OCT (SD-OCT) scan images were acquired using the Spectralis HRA+OCT scanning laser ophthalmoscope (Heidelberg Engineering). Ultra-widefield (optomap) color and 532-nm autofluorescence images were acquired using the Optos 200 Tx scanning laser ophthalmoscope (Optos PLC).

4.4.4 Molecular Analysis

Genomic DNA was extracted and purified from peripheral blood lymphocytes from each patient, and whole-exome sequencing was performed by Psomagen using SureSelect Human All Exon V8 (Agilent). The sequence reads were aligned with the Human Genome version 19 (hg19) reference genome using Burrows-Wheeler Aligner and processed using Genome Analysis Toolkit according to the best practice recommendations. After variant calling, we narrowed our analyses to variants of genes previously associated with retinal disease

(https://web.sph.uth.edu/RetNet) at a minor allele frequency (MAF) of <0.2 based on the gnomAD database. We then performed functional annotation of 221 resulting variants with ANNOtate VARiation (ANNOVAR)[84] using pathogenicity scores from the dbnsfp 4.2a data set.[181,182] The phase of *ABCA4* variants identified in each patient was either confirmed using parental screening imputed based on their cooccurrence patterns in the general population (https://gnomad.broadinstitute.org/variant-cooccurrence).

Chapter 5 A genotype-phenotype correlation matrix for ABCA4 disease based on long-term prognostic outcomes

5.1 Introduction

Pathogenic variants in the ATP-binding cassette, sub-family A, member 4 (*ABCA4*), gene are the underlying molecular cause of a large and complex group of autosomal-recessive retinal degenerative disorders characterized by progressive loss of central vision.[1] The most well-known phenotype is the eponymous Stargardt disease (STGD1, MIM #248200);[38,183] however, advances in genetic screening capabilities, aided by high-resolution diagnostic imaging technology, have broadened the phenotypic profile of ABCA4 disease to an expansive clinical spectrum encompassing severe adolescent-onset to mild late-onset retinal disorders.[183] This phenotypic heterogeneity is matched by an equally extensive array of pathogenic variations across the approximately 140 kb–spanning *ABCA4* locus (1p22.1). To date, more than 2,300 disease-causing variants have been identified in patients.[30] Consistent with the model that clinical phenotypes are dependent on the residual activity of ABCA4 protein,[125,184] variants resulting in null alleles, such as stop-gain, frameshift, canonical splice site, and large copy number variants, have been documented in the most severe phenotypes, such as cone-rod dystrophy, rapid-onset chorioretinopathy (ROC), and even generalized choriocapillaris dystrophies with retinitis pigmentosa–like features.[43,44,125]

More recently, the complex genetic architecture of milder ABCA4 disease manifestations has been uncovered. The most frequent pathogenic allele, c.5882G>A [p.(Gly1961Glu)], is associated with a slowly progressing disease trajectory in patients who often present with transient phenotypes, such as bull's eye maculopathy and occult macular dystrophy.[41,42,48] Despite being highly prevalent in patients, the disease penetrance of this allele has been disputed

as its frequency in the general population is also relatively high; minor allele frequency [MAF] is ~0.5% in Europeans and in certain ethnic groups is much higher.[56,57] We recently resolved this controversy, for the most part (**Chapter 3**), by showing that the contribution of an additional deep intronic variant, c.769-784C>T,[33,106] present in *cis*, is required for clinical penetrance, particularly in homozygotes. Alleles causing late-onset ABCA4 disease, such as c.5603A>T [p.(Asn1868Ile] and c.4253+43G>A, occur at even higher frequencies in the population of European descent (up to 7% MAF) and, unlike p.(Gly1961Glu) and other disease alleles, are only clinically penetrant under the condition that the allele in *trans* is sufficiently deleterious.[46,47,185]

Steady progress in defining genotype-phenotype correlations has been made, and the addition of such knowledge to the medical repertoire has inarguably elevated the clinical care of patients. Studies to date have often relied on cross-sectional cohorts of a patient population that include all age groups. As a result, the correlated "phenotype" studied is often a stage-specific feature, e.g., bull's eye maculopathy or occult macular dystrophy. Such information, while no doubt useful at the diagnostic stage, is not informative of an individual patient's long-term prognosis. To address this issue, we constructed a genotype-phenotype correlation matrix based on the most temporally advanced phenotypes of 112 patients aged 50 years or older who have 2 confirmed pathogenic variants in *ABCA4* coupled with comprehensive clinical characterization. We also reclassified many frequent disease-causing alleles, thereby further clarifying the impact of *ABCA4* variants on clinical outcome. Our findings provide structure to the complex genotype-phenotype correlation landscape of ABCA4 disease and establish a quantitative approach for predicting the prognosis of individual patients by clinicians and genetic counselors and for assessing the

severity of pathogenic variants. The prognostic matrix will also aid in selecting specific patient groups for clinical trials, depending on the specific therapeutic application.

Text and figures from this chapter have been recently published (Lee et al. *Journal of Clinical Investigation Insight*. 2022 Jan 25;7(2):e156154).

5.2 Results

5.2.1 Four clinically defined prognostic outcomes of ABCA4 disease

Demographic, clinical, and genetic characteristics of all 112 patients in the study are summarized in **Supplemental Table 5**; supplemental material available online with this article; https://doi.org/10.1172/jci.insight.156154DS1 Clinical data from the most recent visit for each patient were used in the study. Each patient was categorized into 1 of 4 prognosis categories based on the observable spatial progression of *ABCA4*-associated disease features in the fundus (**Figure 21A**) by age 50 years or older. Patients categorized as having prognosis 1 (n = 28) had the mildest disease outcome (in the cohort), manifesting early retinal pigment epithelium atrophy within the central macula without any apparent pisciform flecks. Patients with prognosis 2 (n = 31) were at a more progressed stage of chorioretinal atrophy across the macula and developed nascent flecks that appeared outside the vascular arcades (**Figure 21A, yellow arrowheads**). All patients with prognosis 3 (n = 20) had multifocal regions of chorioretinal atrophy, which, in some cases, extended beyond the macula and exhibited a pattern of highly confluent flecks in non-atrophic regions. Patients with prognosis 4 (n = 33) progressed to the stage characterized by the large atrophic, coalescing lesions across the entire posterior pole.
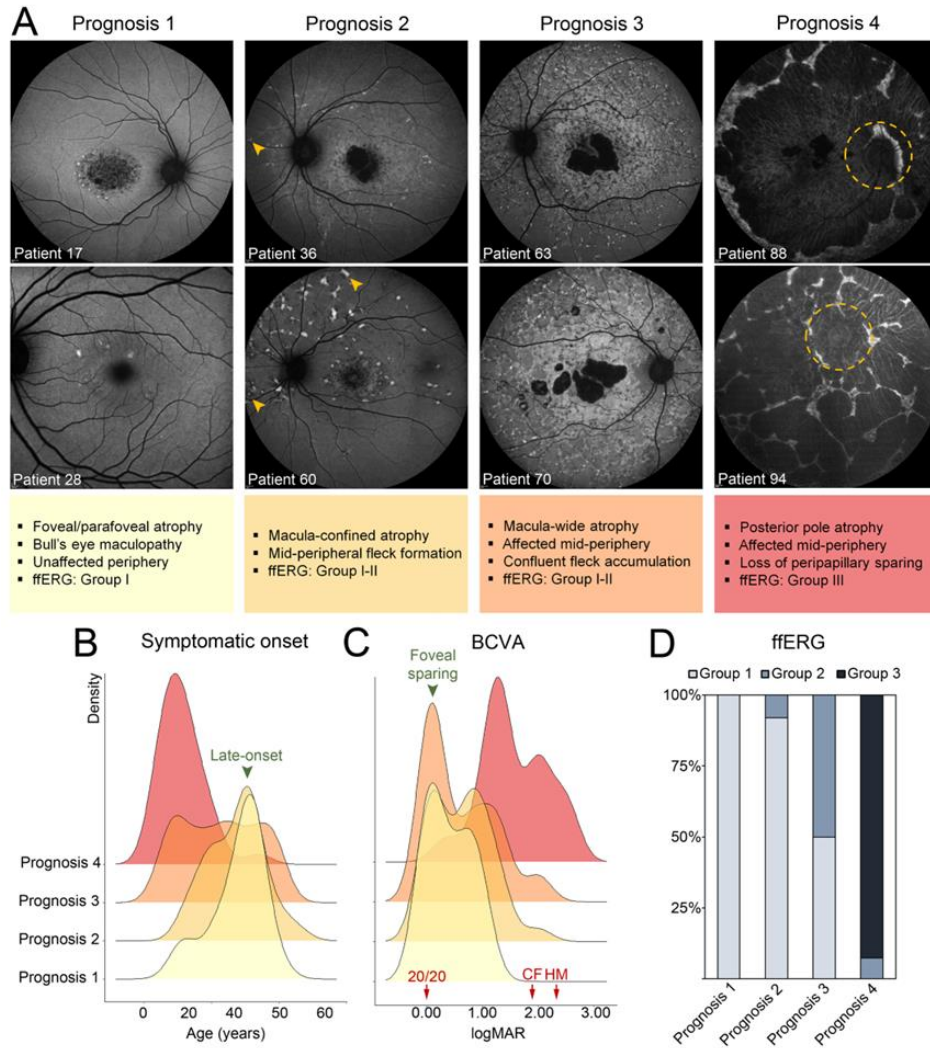
Figure 21. *Figure 22. Clinical characteristics of 4 prognostic outcomes observed in 112 patients (≥50 years of age) with ABCA4 disease.[n]*

[n] Each prognosis was defined according to observable spatial progression of disease features detected at the most recent visit in each patient. (A) Representative autofluorescence images and clinical descriptions of patients in each prognosis classification. Extramacular development of flecks in prognosis 2 is indicated by yellow arrowheads. The position of the optic nerve is encircled by the dotted yellow line. The fields of view are 55° in all autofluorescence images except P28 (30°). (B) Ridgeline plots of the distribution of ages at which visual symptoms of all patients were first reported for each prognosis category (bandwidth = 5.83). (C) Density plots of the best-corrected visual acuity (BCVA) of the least-impaired eye of all patients (bandwidth = 0.246). BCVAs were presented in logMAR units with corresponding Snellen equivalents (20/20, counting fingers [CF] and hand motion [HM], red arrows) provided. (D) Proportion of ffERG groupings according to the classification by Lois et al. (37) for each prognosis category. ffERG, full-field electroretinogram; group 1, normal responses; group 2, attenuation of cone responses; group 3, attenuation of cone and rod responses; logMAR, logarithm of the minimum angle of resolution.

There were no significant differences in the mean age of patients between prognosis groups (**Appendix C**). The mean age of symptomatic onset was earliest among patients with prognosis 4 (17.1 years) compared with the milder prognostic groups, which had peak distributions at 41.7 years (prognosis 1) and 40.9 years (prognosis 2) due to the large number of late-onset disease cases in these latter groups (**Figure 21B**, **Appendix C**). Best corrected visual acuity (BCVA) from the most recent visits was also poorest among patients with prognosis 4, of whom approximately 40% were assessed by counting fingers or by an assessment for even worse visual acuity (P < 0.00001; **Figure 21C**, **Appendix C**). Comparatively, BCVA distributions were multimodal among patients with prognoses 1 to 3, most of whom had 20/200 (logMAR 1.00) or worse and 20/20 in cases with foveal sparing. Full-field electroretinogram responses were largely unremarkable in prognoses 1 and 2 (**Figure 21D**). Significant defects were found in prognosis 3 (50% Lois Group 2) and prognosis 4 (93% Lois Group 3, see Methods). There were no significant differences in mean age of patients across prognosis groups (P = 0.254) (**Appendix C**).

5.2.2 Classification of p.(Gly1961Glu), p.(Asn1868Ile), and a potentially new class of rare hypomorphic alleles

Genotypes consisting of the major disease-causing allele, p.(Gly1961Glu), and the frequent hypomorphic allele, p.(Asn1868Ile), were the most prominent variants among the mild phenotypes, together accounting for 56% of patients with prognosis 1 and prognosis 2 (**Figure 22A**). Despite the advanced age of this cohort, 3 patients (patients 12, 18, and 20) presented with early-stage bull's eye maculopathy. As we have previously shown, p.(Asn1868Ile) is highly associated with foveal sparing, which is a major contributing factor to the delayed symptom onset age in most patients (**Figure 22B**).[46]
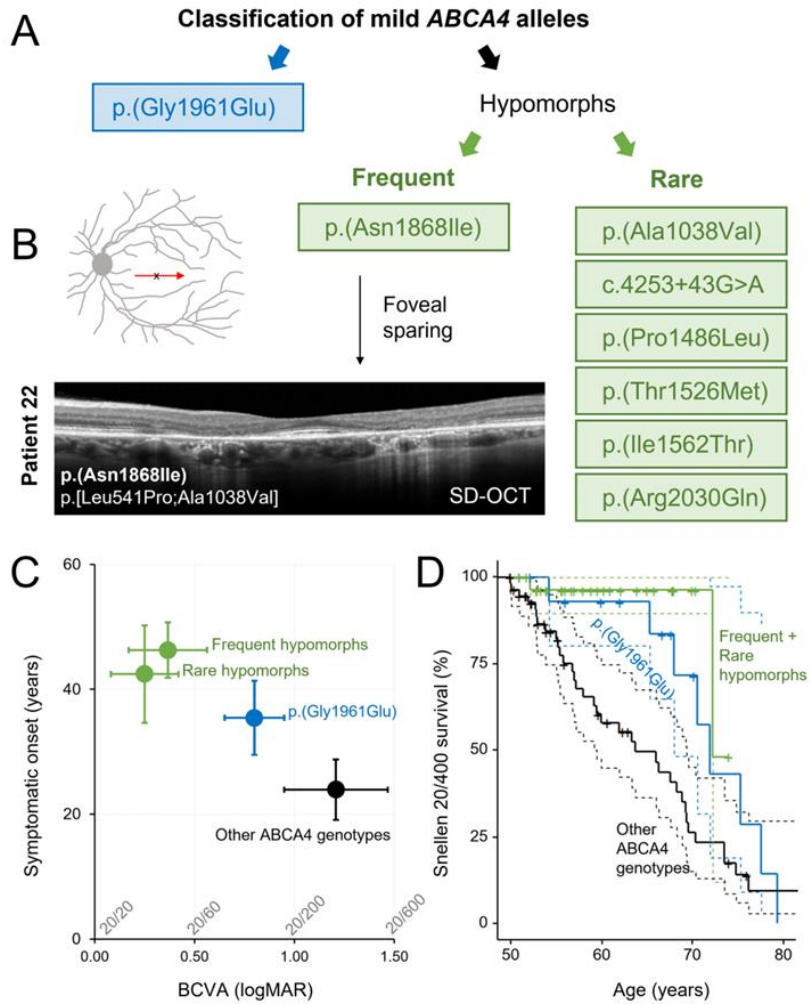
Figure 22. *Classification and phenotypic characterization of mild ABCA4 alleles.[o]*

[o] (A) Mild ABCA4 alleles identified in patients with mild prognoses included p.(Gly1961Glu) and 2 hypomorphic allele subgroups: frequent hypomorphs, which consisted of p.(Asn1868Ile), and rare hypomorphs, which consisted of p.(A1038V), c.4253+43G>A, p.(Pro1486Leu), p.(Thr1526Met), p.(Ile1562Thr), and p.(Arg2030Gln). (B) Horizontal (6 mm) spectral-domain optical coherence tomography (SD-OCT) scan showing structural preservation of the fovea in patient 22, an allele-specific subphenotype common among p.(Asn1868Ile) genotypes. (C) Scatter plot of average age of onset (years) versus average best corrected visual acuity (BCVA) of the least deteriorated eye in patients within all patients/genotypes with p.(Gly1961Glu) (blue), frequent hypomorph p.(Asn1868Ile) (green), rare hypomorph (green), and all other allele combinations (black) in the study cohort. Horizontal and vertical bars represent ±95% confidence intervals. BCVAs are provided as logMAR units with corresponding Snellen equivalents listed above the axis. (D) Survival analysis showing the probability of the least affected eye retaining better than Snellen 20/400 in patients with p.(Gly1961Glu) (blue curve) or with rare and frequent hypomorphic alleles (green curve) and all other patients (black curves). Color-matched dotted lines represent 95% confidence intervals for each individual curve.

Among the remaining patients in the mild prognosis categories, we identified another group of patients with 6 recurring alleles, p.(Ala1038Val), c.4253+43G>A (18), p.(Pro1486Leu), p.(Thr1526Met), p.(Ile1562Thr), and p.(Arg2030Gln), that have features in common with p.(Asn1868Ile), most notably, delayed symptom onset due to foveal sparing (**Supplemental Table 5** and **Supplemental Figure 2**). Disease features in the fundus of these cases were confined to a definable area around the vascular arcades in a reticular appearance studded along the peripheral boundary with elongated "tails" projecting eccentrically in a radial pattern (**Supplemental Figure 2**; and **Supplemental Figure 3**).

Generalized dysfunction of the cone and rod systems was not detected on ffERG testing (**Supplemental Figure 2**). Although each of these variants is exceedingly rare in the general population (0.005 > MAF > 0.00005), unaffected homozygotes have been reported for p.(Ala1038Val), p.(Ile1562Thr), and c.4253+43G>A, resulting in some cases in conflicting interpretations of pathogenicity. Furthermore, as has been observed with p.(Asn1868Ile), the alleles in trans in these genotypes are mostly loss-of-function alleles, including an 8.4 kb deletion identified in patient 58 (**Supplemental Figure 3**). Considering these differences, we separated these mild *ABCA4* alleles into 3 classes: p.(Gly1961Glu), frequent hypomorph, and rare hypomorph.

5.2.3 Classification of PVS1 and severe non-PVS1 alleles

The distribution of PVS1 (i.e., null or loss-of-function alleles) (**Table 2**) was skewed toward the most severe clinical phenotypes though at a lower-than-expected proportion. Genotypes with a PVS1 allele comprised about 33.3% of prognosis 4 cases while the remaining approximately 66.7% consisted mostly of missense variants and, in part, functionally validated deep intronic and synonymous variants. The majority of these missense alleles have been observed to be the

causal allele in *trans* from p.(Asn1868Ile) (**Supplemental Table 5**).[46,47] Using our current data

set, we further classified 5 additional alleles, p.(Thr1019Met), p.(Ala1598Asp),

p.([Asp1532Asn;Asn1868Ile]), p.([Gly863Ala;Asn1868Ile]), and c.5714+5G>A, as severe based

on their recurrence in compound heterozygous and/or homozygous patients with prognosis 3 or

prognosis 4. To distinguish these severe non-PVS1 alleles from moderate/milder alleles, we

grouped them into a separate "severe" subclass (**Table 2**).

Table 2. Classification criteria and list of all pathogenic ABCA4 alleles in the study cohort.[p]

Allele class	Classification criteria	Alleles
p.(Gly1961Glu)	(Allele-specific)	**p.(Gly1961Glu)**
Frequent hypomorph	(Allele-specific)	**p.(Asn1868Ile)**
Rare hypomorph	Mild prognostic association and hypomorphic allele features (late-onset symptoms, foveal sparing)	**p.(Ala1038Val)**, **c.4253+43G>A**, **p.(Pro1486Leu)**, **p.(Thr1526Met)**, **p.(Ile1562Thr)**, **p.(Arg2030Gln)**
Moderate	No intrinsic indication of severity (variant effect on protein); determined by functional studies to be moderate; undetermined clinical association	p.(Arg24His), p.(Ile214Asn), p.(Leu257Arg), c.859-9T>C, p.(Trp339Gly), p.(Lys346Thr), p.(Arg508Cys), p.(Arg602Gln), p.(Gln623Arg), p.(Arg653His), p.(Phe655Cys), p.(Val675Ile), p.(Leu844Arg), p.(Thr972Asn), **p.(Val989Ala)**, p.(Ser1096Leu), **p.(Arg1108Cys)**, p.(Arg1108His), p.(Thr1112Asn), p.(Thr1117Ile), p.(Gln1332_Cys1339dup), p.(Arg1640Gln), p.(Ser1696Asn), p.(Leu1784Arg), p.(Met1882Ile), p.(Val1896Asp), p.(Arg2040Gln), p.(Arg2077Gly), **p.(Arg2106Cys)**, **p.(Arg2107His)**, p.(Cys2150Tyr), p.[Cys1490Tyr;Asn1868Ile], **p.([Leu257Aspfs*3,Gly1961Glu])**, **p.([Thr1253Met;Gly1961Glu])**, c.4539+2001G>A, c.[4539+2028C>T;302+68C>T]
Severe	Observed in *trans* to hypomorphic allele in patients	p.(Arg18Trp), **p.(Cys54Tyr)**, p.(Gly88Arg), p.(Arg602Trp), p.(Cys641Arg), p.(Arg653Cys), c.2382+179G>A , p.(Ala848Asp), **p.(Asn965Ser)**, p.(Gly1203Arg), **p.(Pro1380Leu)**,‡ p.Phe1417del, p.(Tyr1557Cys), p.(Arg1705Trp), p.(Phe1714Ser), **p.(Leu2027Phe)**, p.(Arg2077Trp), p.(Glu2096Lys), **p.[Leu541Pro;Ala1038Val]**, p.[Trp1408Arg;Arg1640Trp]
	Validated in functional studies in HEK293T and patient-derived cell lines	**c.768G>T**,† c.1100-6T>A, **c.3050+5G>A**, c.4253+5G>T, p.(Gln1513Arg), c.4773+3A>G, c.5196+1056A>G, **c.5461-10T>C**, c.6342G>A†
	Associated with Prognosis 3 and Prognosis 4	**p.(Thr1019Met)**,‡ **p.(Ala1598Asp)**,‡ **p.[Gly863Ala,Gly863del;Asn1868Ile]**, **p.[Asp1532Asn;Asn1868Ile]**,‡ **c.5714+5G>A**
PVS1[q]	Null or loss-of-function variants (nonsense, frameshifts, canonical +/- 1 or 2 splice sites, large multi-exonic deletions)	p.(Asn14Lysfs*38), p.(Ser84Thrfs*16), p.(Asp202Thrfs*39), c.570+331_768+4523del, p.(Leu296Cysfs*4), p.(Val521Serfs*47), **p.(Trp663*)**, **p.(Arg681*)**, c.2160+1G>C, c.2587+2T>C, p.(Ser1071Cysfs*14), p.(Trp1408*), p.(Cys1502*), c.4539+1G>T, c.4540-2A>G, p.(Leu1534Trpfs*1), c.5018+2T>C, p.(Ile1687Phefs*15), p.(Val1706*), p.(Val1764Trpfs*14), **p.(Arg2030*)**, p.(Ala2044Valfs*25), p.(Arg2149*), p.(Gln2220*), c.6148-698_6670del

[p] Variants in bold were found in 2 or more patients in the study. †Synonymous variants validated by Braun et al. (2013) and Sangermano et al. (2018) to have the following effects: c.768G>T (p.(Leu257Valfs*17)) and c.6342G>A (p.([Val2114_Ser2129delfs*5,=])). ‡Homozygous in individual with Prognosis 3 or Prognosis 4. Classification criteria for all individual alleles are provided in **Supplemental Table 6**.

[q] The PVS1 classification was used in accordance with the ACMG/AMP Standards and Guidelines for the Interpretation of Sequence Variants.

5.2.4 Classification of moderate variants

After classifying 65% of alleles in the study cohort as either hypomorphic, mild or severe, a remaining group of 36 unique variants (35%, 56 total alleles) did not meet any of the aforementioned classification criteria. These alleles were uniformly distributed across prognosis categories as compared with the other classified allele groups, which skewed accordingly toward mild or severe prognoses (**Figure 23A**). The coding effect in 93% of these alleles was missense (**Figure 23B**). The 3 non-missense variants in this group were an exonic in-frame duplication, a deep intronic 15-nucleotide deletion, and the c.859-9T>C variant, for which prior midi-gene studies in HEK293T cells have shown a moderate effect as the variant results in 75% of wild-type *ABCA4* RNA.[33] Considering the nonspecific genetic attributes of these alleles and their collectively uniform distribution across prognosis categories, we classified them into a moderate group.
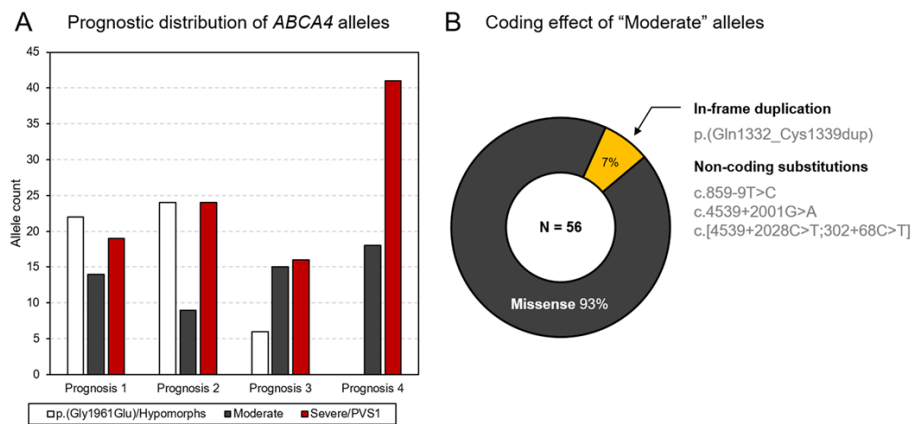
Figure 23. *Clinical and genetic characteristics of moderate ABCA4 alleles.*[r]

---

[r] (A) Distribution of p.(Gly1961Glu) and hypomorphs (white bars) and moderate (gray bars) and severe/PVS1 (red bars) alleles across prognosis categories. (B) Coding effect of alleles designated as moderate in patients with ABCA4 disease.

5.2.5 Construction of a genotype-phenotype correlation matrix

We generated probability matrices representing correlations among the 4 clinical prognosis categories (prognoses 1–4) and all possible genotypic combinations for the following allele classes: p.(Gly1961Glu), frequent hypomorph, rare hypomorph, moderate, severe, and PVS1 (**Table 2** and **Figure 24**).
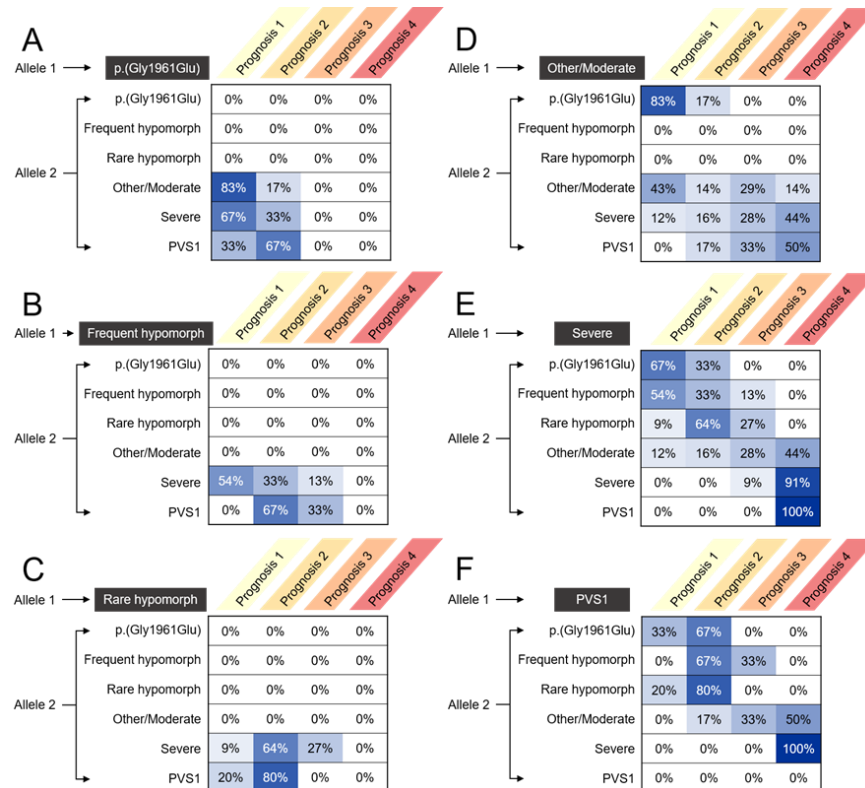
Figure 24. *Prognostic probabilities (%) of all possible combinations for each allele class.*[s]

Genotypes consisting of a p.(Gly1961Glu), frequent hypomorphic, or rare hypomorphic allele had the mildest prognostic outcomes, with most cases having either prognosis 1 or prognosis 2 (**Figure 24A–24C**). Genotypes of these 3 allele classes were also the least heterogeneous in

---

[s] (A) p.(Gly1961Glu), (B) frequent hypomorph, (C) rare hypomorph, (D) moderate, (E) severe and (F) PVS1. Percentages represent the observed fraction of patients across each prognosis category for a given allele 1 and allele 2 combination.

terms of prognostic distribution (P = 0.1164, 2-sided Fisher's exact test [FET], see **Appendix C**) compared with both moderate and severe/PVS1 genotypes (P < 0.001, 2-sided FET, see **Supplemental Data 4**). This is due, at least in part, to the absence of homozygotes and cases with other mild allele combinations. The apparent non-penetrance, coupled with the consistent clinical phenotype, suggests that the p.(Gly1961Glu) allele exhibits a form of "clinical dominance" whereby the allele in *trans*, while necessary for disease expression, has minimal to no effect on the phenotypic variability. The frequent and rare hypomorphic alleles require a "severe", mostly deleterious, allele in *trans* and therefore represent haploinsufficiency cases, where the disease expresses only after certain functional deficiency threshold is reached.

Conversely, all prognosis categories were represented in moderate, severe, and PVS1 allele combinations, and the additive severity of the allele in *trans* strongly correlated with the prognostic severity for these allele combinations. For instance, moderate allele genotypes with another moderate allele in *trans* gave a 43% probability of having prognosis 1 whereas having a p.(Gly1961Glu) allele in *trans* increased the prognosis 1 probability to 83%, and having a severe or PVS1 allele in *trans* reduced the prognosis 1 probability to 0% to 12% and increased the probability of prognoses 3 and 4 to 44% to 55% (**Figure 24D**). Similar trends appeared for both severe and PVS1 genotypes. Prognosis correlations between PVS1 and severe allele genotypes were also remarkably similar, suggesting very little clinical distinction between the 2 allele classes (**Figure 24E and 24F**). To simplify these observations for clinical applicability, we excluded allele combinations that were not present in patients for any prognosis, thereby collapsing each allele matrix into only those representing the genotypes of all patients across the study cohort (**Figure 25**).

Allele 1	Allele 2	Prognosis 1	Prognosis 2	Prognosis 3	Prognosis 4
p.(Gly1961Glu)	Moderate	83%	17%	0%	0%
	Severe	67%	33%	0%	0%
	PVS1	33%	67%	0%	0%
Frequent hypomorph	Severe	54%	33%	13%	0%
	PVS1	0%	67%	33%	0%
Rare hypomorph	PVS1	20%	80%	0%	0%
	Severe	9%	64%	27%	0%
Moderate	Moderate	43%	14%	29%	14%
	Severe	12%	16%	28%	44%
	PVS1	0%	17%	33%	50%
Severe	Severe	0%	0%	9%	91%
	PVS1	0%	0%	0%	100%
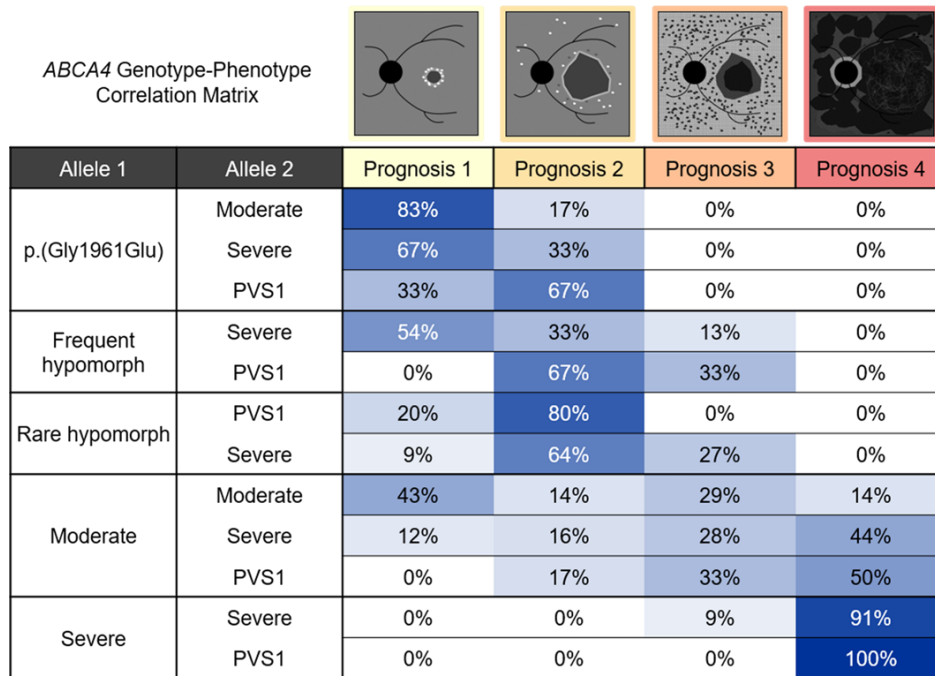
Figure 25. Genotype-phenotype correlation matrix based on the long-term prognostic outcomes of 112 genetically confirmed patients with ABCA4 disease.[t]

5.3 Discussion

Advances in genomic medicine in recent decades have allowed genetic testing in the clinic to be a routine option for patients with monogenic diseases. While this has undoubtedly improved the standard of care for patients, the utility of a genetic result rarely extends beyond diagnostic confirmation. The underleveraging of variant level insight in the clinic is attributable to the lack of concrete genotype-phenotype correlations, which are difficult to assess for several reasons. First, Mendelian disorders like ABCA4 disease are both rare and profoundly heterogeneous. Prior studies have noted strong trends with specific alleles;[119,121,186-189] however, most cohorts are typically insufficient in size and scope to make conclusions that are applicable to clinical care. Moreover, cross-sectional study cohorts themselves are demographically heterogeneous,

---

[t] Percentages represent the observed fraction of patients across each prognosis category for a given allele 1 and allele 2 combination. For the list of unclassified variants, see Table 1 or Supplemental Table 2.

particularly in terms of age, adding further limitations, such as unknown disease trajectory and clinical outcome of younger patients.

The large clinical and genetic repository we have built over 20+ years has allowed us to overcome most of these issues. Using the well-characterized clinical data of an age-restricted (≥50 years of age) cohort of 112 patients, we were able to precisely dissect apart the complex genotype-phenotype correlation landscape of ABCA4 disease in a quantitative manner that can be immediately used to assess and predict the long-term prognosis of patients following genetic testing. The correlation matrix can be improved upon by the addition of more cases in follow-up studies to increase statistical power and accommodate other *ABCA4* variants not described in this study. These data also provide precise insight into magnitude differences in disease severity between different alleles, which should be considered in the selection of patients for clinical trials.

These data can also be used to clinically classify the pathogenicity of different *ABCA4* alleles. Analyzing patients with the mildest prognoses, for instance, identified a class of rare hypomorphic variants that exhibit clinical overlap with p.(Asn1868Ile) cases, including slowly progressing disease and persistent sparing of the fovea. Results of prior functional and clinical studies of these variants were also consistent with mild characterization. For instance, transgenic expression of human p.(Ala1038Val) in both *Xenopus laevis* tadpole retinae and HEK293T cells revealed no observable defects in subcellular localization.[137,190] The latter study also showed that p.(Ala1038Val) mutant structure closely resembles the wild-type ABCA4 structure using single-particle analysis (cryo-electron microscopy).[190] The clinical phenotypes of all well-characterized patients harboring p.(Ala1038Val),[190] p.(Arg2030Gln),[191,192] p.(Pro1486Leu),[159]

p.(Thre1526Met),[159,191,192] and p.(Ile1562Thr)[193,194] alleles in the literature are also consistent with milder disease in general and with specific hypomorphic features.

These and other mild ABCA4 alleles, including p.(Gly1961Glu) and the frequent hypomorph p.(Asn1868Ile), also exhibit some collective characteristics that are inconsistent with most autosomal-recessive diseases. Under an additive pathogenicity model that has been proposed for *ABCA4*,[195] patient phenotypes are expected to vary according to the combined effects of both *ABCA4* alleles, and indeed, the phenotypic outcomes of moderate and severe/PVS1 alleles vary widely depending on the allele in trans (**Figure 25D–25F**). Mild alleles, however, appear to be "clinically dominant" in that all genotypes are invariably mild in overall severity (long-term prognosis), and additionally, each respective allele has unique and consistent sub-phenotypic features (e.g., optical gap in cases with the [p.(Gly1961Glu)] variant) irrespective of the type of allele in trans. This phenomenon may be partially explained by the non-penetrance of mild genotypes resulting in more "homogeneous" genotype combinations in observed cases. The underlying mechanisms resulting in sub-phenotypes, while of diagnostic value for solving cases without genetic confirmation, remain unknown.

Our analysis also reclassified several non-PVS1 alleles as clinically severe, such as the c.5714+5G>A substitution in intron 40, which was previously reported to have a "moderate" effect as it results in approximately 39% correctly spliced mRNA in HEK293T cells.[33] Consistent with other clinical studies,[125,191,196,197] we also found the variant to be exclusively associated with severe phenotypes (compound heterozygous in 2 patients with prognosis 3 and 3 patients with prognosis 4), which led us to conclude that the allele is at least clinically severe in patients. Several other missense alleles were also classified as severe, including p.(Ala1598Asp), p.(Thr1019Met), and p.([Asp1532Asn;Asn1868Ile]), based on their recurrence in patients

(including homozygotes) with prognoses 3 and 4. These, and a large group of other missense alleles, comprised an unexpectedly large proportion of genotypes leading to the most severe prognostic outcomes. This observation should caution against the common interpretation that most missense variants, at least in the *ABCA4* gene, are less severe than PVS1 variants.

This study has several limitations. While the patient cohort is large considering the rarity of this disease, not all possible *ABCA4* genotypes are represented. Notably, some biallelic PVS1 genotypes, which are known to underlie the most severe ABCA4 disease phenotypes, such as retinitis pigmentosa-like, ROC, and cone-rod dystrophy,[43,44,110,125,198] were not included. Prognostic assessment in these cases, however, is usually unambiguous as visual deterioration begins early in life and the disease progresses rapidly. The 4 prognostic classifications defined in the study may also not fully represent the breadth of clinical outcomes in ABCA4 disease. Further studies based on our study design in larger, preferably multiethnic, cohorts of more comprehensively characterized patients would help address many of the current limitations and critically advance precision medicine for ABCA4 disease. In summary, we constructed a genotype-phenotype matrix based on the long-term prognostic outcomes of 112 genetically confirmed patients with ABCA4 disease. Two major disease-causing variants of *ABCA4*, p.(Gly1961Glu) and p.(Asn1868Ile), accounted for more than half of the genotypes (patients) with the mildest prognoses. We also identified a potentially new class of rare hypomorphic variants among cases with mild prognoses, with similar phenotype to p.(Asn1868Ile) cases. Cases with the p.(Gly1961Glu) allele exhibit "clinical dominance" in their consistent clinical features irrespective of the allele in trans. We identified a large group of missense variants that are associated with the more severe prognoses and clinically reclassified others, including c.5714+5G>A, that were previously suggested to be non-severe. The genotype-phenotype

correlation matrix provides prognostic probabilities based on underlying *ABCA4* genotype and can be used as a tool to assess disease severity in patients and as a framework for design of and selection of patients for clinical trials.

5.4 Materials & Methods

5.4.1 Study participants and clinical characterization

Patients diagnosed with Stargardt or ABCA4-associated disease were recruited from the Department of Ophthalmology at Columbia University Irving Medical Center. In total, 112 unrelated patients harboring 2 pathogenic variants in *ABCA4* and ≥50 years of age were included in the study. The lower age limit threshold of 50 years was chosen to ensure that all major genotype groups were accommodated in the analysis, particularly patients with the common hypomorphic allele, p.(Asn1868Ile), whose median age of symptom onset is approximately 35 years (IQR 28–48 years). Each patient underwent a complete ophthalmic examination by a retinal physician, which included slit-lamp and dilated fundus examination, BCVA (Snellen), color fundus photography, fundus autofluorescence (AF, 488 mn, 532 nm, and 787 nm), SD-OCT scanning, and ffERG testing. Conversion of CF and HM to logMAR units was calculated in accordance with Schulze-Bonsel et al. (35). In short, CF was replaced with the calculated decimal acuity of 0.014, which corresponds to approximately Snellen 20/1500 or logMAR 1.875; HM was replaced with the decimal acuity of 0.005, which corresponds to approximately Snellen 20/4000 or logMAR 2.300.

Imaging across all modalities was conducted following pupil dilation (>7 mm) with tropicamide (1%) and phenylephrine hydrochloride (2.5%). Fundus AF (488 nm) images and 9 mm horizontal foveal SD-OCT scans were acquired with the Spectralis HRA+OCT (Heidelberg Engineering). Ultra-widefield AF images were acquired with an Optos 200 Tx (Optos PLC).

ffERGs were recorded on silver-impregnated fiber electrodes (DTL, Diagnosys LLC) on the Espion Visual Electrophysiology System (Diagnosys LLC) in accordance with International Society for Clinical Electrophysiology of Vision (ISCEV) standards.[180] ffERG classifications were assigned according to electrophysiological attributes described by Lois et al.[109] Group 1 is characterized by no detectable loss in scotopic or photopic function; group 2 is characterized by photopic loss but normal scotopic function; and group 3 exhibits deterioration of both scotopic and photopic function.

Prognosis classifications (prognoses 1, 2, 3, or 4) were determined by 2 independent graders using 55° AF (488 nm) images of each eye for all study patients. In patients with interocular discordance, the prognosis classification was assigned according to the more advanced eye. Discordant evaluations between graders were adjudicated by an additional grader. Notes from the corresponding clinical exam, which included direct and indirect ophthalmoscopy details, were reviewed to confirm the final prognosis group assignment in each patient. All 3 graders were blinded to the ABCA4 genotype of each patient at the time of prognosis classification.

5.4.2 Molecular analyses

Screening of the *ABCA4* gene was performed by next-generation sequencing as previously described.[34,79] All detected possibly disease-associated variants were confirmed by Sanger sequencing and analyzed with Alamut software (Interactive Biosoftware). Segregation of the new variants with the disease was analyzed in families if family members were available. Functional annotation of variants was determined using computational software including ANNOVAR using pathogenicity scores of M-CAP, REVEL, Eigen, and CADD-PHRED (v1.6). As a general guideline, pathogenic consequences are predicted for variants with scores over 0.025 for M-CAP, 0.5 for REVEL, 0.5 for Eigen, and 20 for CADD. The allele frequencies of all

variants were compared to those in the Genome Aggregation Database (gnomAD, accessed October 2021).

5.4.3 Statistical analysis

A detailed summary of all statistical calculations is provided in **Appendix C**. Comparison of mean characteristics between prognosis categories was determined by a 1-way ANOVA test with post hoc Tukey's honestly significant differences and Kruskal-Wallis tests. Significance was set at α level less than 0.05. Density plots were generated using the ggridges package in R version 4.0.4. FETs for count data (2×3 contingency table) were used to compare the distributions of mild, moderate, and severe allele combinations across prognosis categories.

Study approval. All study procedures were defined under protocol AAAI9906 approved by the Institutional Review Board at Columbia University Irving Medical Center. The study adhered to tenets set out in the Declaration of Helsinki.

Chapter 6 Genotype is a Significant Factor in Determining the Measurability of Atrophic Lesions in ABCA4 disease

6.1 Introduction

Stargardt disease is a profoundly heterogeneous disorder due to the large and complex landscape of pathogenic variants in the *ABCA4* gene. Although there are no FDA-approved treatments thus far, several therapeutic options have been and are currently in clinical trials.[199,200] As is the case with any scientific experiment, clinical trials need to be well-designed and properly controlled in order to reliably assess treatment efficacy. While the double-blind placebo approach has been relatively successful for many approved treatments, it is difficult to estimate how many failed trials have been due to confounding factors.

There are several unique challenges in designing an effective clinical trial for ABCA4 disease. The first is the selection of a primary outcome measure. Unlike many systemic disorders, visualizing disease manifestation requires the aid of tools such as an ophthalmoscope to visualize the fundus. Similarly, the documentation of disease progression requires sophisticated imaging modalities. Symptomology in ABCA4 disease, while singular, is entirely subjective.

The currently approved outcome measure for ABCA4 disease is growth rate of chorioretinal lesions.[199,200] Several studies have previously characterized the different types of atrophic changes for ABCA4 disease;[44,201-207] however, the type relevant to clinical trials have been called "definitely decreased autofluorescence" (DDAF).[206,207] While this nomenclature may be problematic given our understanding of autofluorescence and ABCA4 disease,[208] we will proceed with this term for simplicity. DDAF is an atrophic stage characterized by an apparent loss of both photoreceptors and RPE based on SD-OCT imaging.[206,207] Growth rates of DDAF

are highly variable and some studies have assessed several contributing clinical factors.[121,187] None however have thoroughly investigated the underlying contribution of genotype.

Given the combined challenges posed by genetic heterogeneity, this study evaluates how genotype effects various aspects of a clinical trial that uses DDAF growth rate from the selection of participants, measurement considerations and putative differences in progression rate.

6.2 Results

6.2.1 Individuals with "measurable" DDAF lesions comprise a small fraction of ABCA4 disease

The Starting Cohort consisted of 415 patients with at least 2 confirmed disease-causing variants in *ABCA4*. We then screened for individuals that had "quantifiable" DDAF lesions through 3 consecutive filters: Filter 1 selects for patients with DDAF lesions in one or both eyes, Filter 2 selects for patients (from the output of Filter 1) with DDAF lesions that are entirely visible within a standard 30° SW-AF image (**Figure 26A**) (lesions extending outside this magnification are excluded, **Figure 26B**) and Filter 3 selects for patients (from the output of Filter 2) with DDAF lesions that are well-delineated along its entire perimeter (**Figure 26C**). DDAF lesions with obscure borders are excluded (**Figure 26D**).
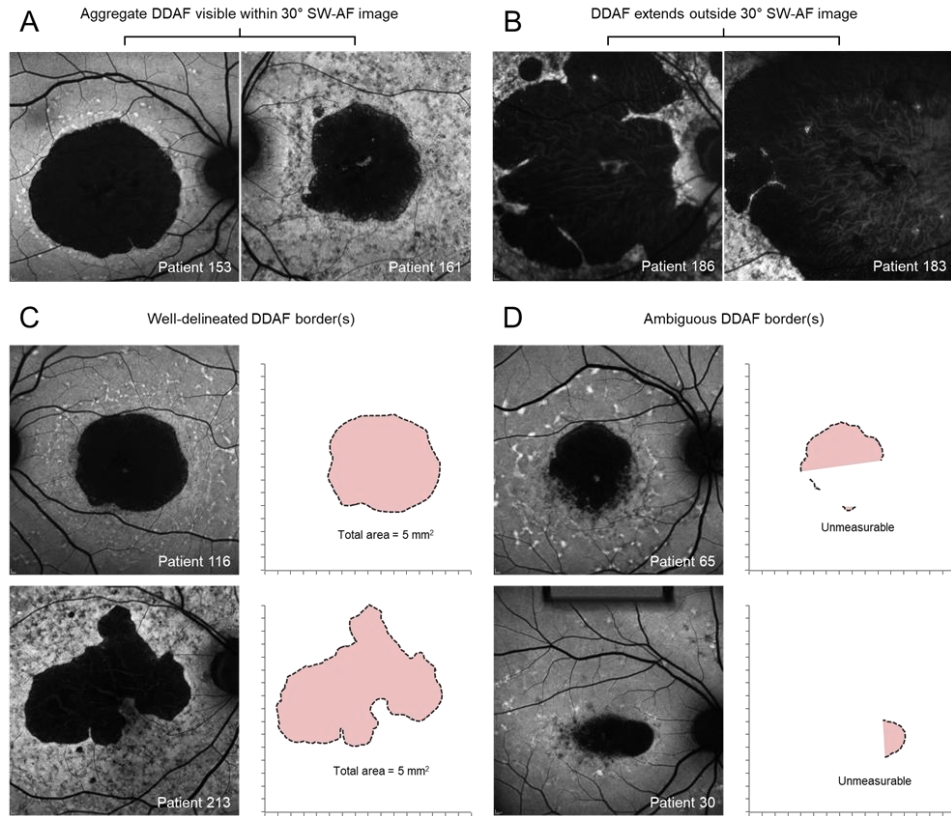
Figure 26. *Selection criteria or Filter 2 and Filter 3.*

Out of 415 patients, 216 (52.0%) had a DDAF lesion in at least one eye at baseline (Filter 1). As expected, those with DDAF were significantly older (mean age = 48.4 years, SD = 16.2) than those without DDAF (mean age = 28.6 years, range 10-83 years, SD = 13.3) ($P<0.0001$, Mann-Whitney U test) (**Table 3**).

Table 3. *DDAF vs. Other lesion types.*

Lesion type	Patients (n)	Female/Male ratio	Age at visit (years)					
			Mean	**Median**	**Min**	**Max**	**SD**	**IQR**
DDAF	216 (52%)	110/106 (1.04)	48.4	48.0	10.0	83.0	16.2	23.0
Other lesion types[‡]	199 (48%)	114/85 (1.34)	28.6	26.0	5.0	72.0	13.3	16.9
Cohort total	415	224/191 (1.17)	38.8	37.0	5.0	83.0	17.9	28.0

Of these 216, 122 patients had DDAF lesions visible within a standard 30° SW-AF image (Filter 2) and 79 (out of 122) well-delineated DDAF lesions comprising the final Target Cohort (mean age = 51.9 years, range 24-80 years, SD = 12.0) (P<0.0001) (**Table 4**).

Table 4. *Filter 1 and Filter 2.*

DDAF lesion size	DDAF lesion border	Patients (n)	Age at visit (years)					
			Mean	**Median**	**Min**	**Max**	**SD**	**IQR**
>30° SW-AF image	-	94 (43.5%)	51.6	52.5	17.0	83.0	17.1	31.0
<30° SW-AF image	Ambiguous	43 (19.9%)	34.8	32.0	10.0	72.0	14.2	19.0
	Well-delineated	79‡ (36.6%)	51.9	52.0	24.0	80.0	12.0	13.0
	DDAF total	216						

Follow up imaging was available for 46 (out of 79) patients/eyes that comprise the Progression Cohort. The mean time interval between the baseline and follow up visit in this group was 4.0 years (SD = 2.8 years) and the DDAF growth rate ranged from 0.113 mm^2/year to 3.18 mm^2/year (mean = 1.06 mm^2/year, SD = 0.75) (**Figure 27**). A square root transformation was applied to the baseline and follow up sizes to obtain the linear progression rates: range, 0.034 mm/year to 0.487 mm/year (mean = 0.197 mm/year, SD = 0.102).
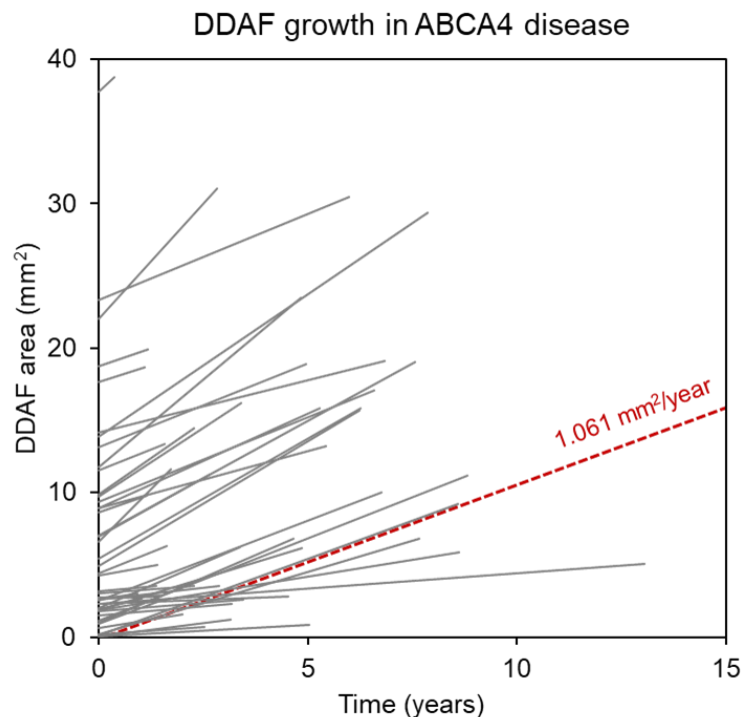
Figure 27. *Measured and estimated DDAF growth rates in ABCA4 disease.*

6.2.2 The genotypic profile of the target cohort skews toward milder alleles

ABCA4 genotype categories were defined according to our previous study (see *Materials & Methods*). The genotype fractions in the Starting Cohort were: p.(Gly1961Glu) (20.5%), p.(Asn1868Ile) (10.4%), Rare hypomorphs (7.7%), Other *ABCA4* alleles (39.5%), 1 PVS1 allele (17.8%) and 2 PVS1 alleles (4.1%). These percentages, however, did not remain consistent across the screening (filtering) process for patients with reliably quantifiable DDAF lesions. Specifically, there was an "inversion" in the mild-to-severe genotype ratio between the Starting and Target cohorts—that is, the milder genotypes, p.(Gly1961Glu), p.(Asn1868Ile) and Rare hypomorphs collectively accounted for ~1/3 of the Starting Cohort and ~2/3 for the Target Cohort whereas the remaining, comparatively more severe genotypes comprised ~2/3 and ~1/3 of the Starting and Target cohorts, respectively (**Figure 28A**). It should also be noted that

p.(Gly1961Glu) became the most frequent genotype in the Target Cohort, alone accounting for

~25% (22 patients) while PVS1 genotypes were almost entirely eliminated (2 patients remain)

(**Figure 28B**).
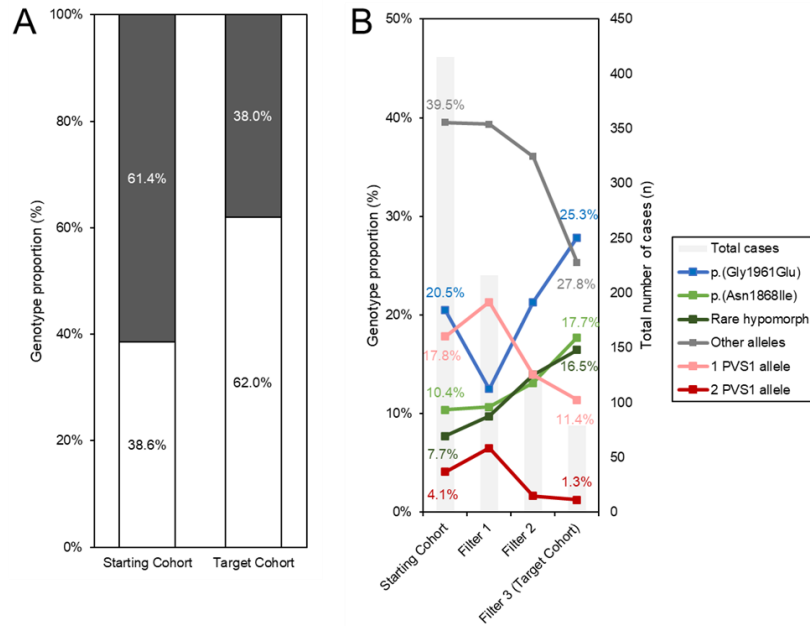
Figure 28. Proportional shifts in the filtering process between the Starting and Target cohorts.

6.2.3 DDAF growth rates are significantly different across genotypes

Several specific trends help to explain the genetic profile of the Target Cohort. Filter 1 screens

for individuals with DDAF lesions and thus selects against mild allele genotypes like

p.(Gly1961Glu) genotypes in which the prevalence of DDAF is the lowest (31.8%, $P < 0.0001$)

and conversely, for severe genotypes such as 2 PVS1 alleles (82.4%, $P < 0.05$) (**Supplemental

Table 7**). The large proportion of mild allele genotypes in the final Target Cohort however, is

determined by the last two filters which select for DDAF lesions that are smaller (i.e., those that

are entirely visible within a 30° SW-AF image) (Filter 2) and with well-defined borders that can

be reliably segmented for quantification (Filter 3). Both of these qualities are significantly

positively associated with mild allele genotypes (OR>1) and negatively associated with other and severe allele genotypes (OR<1) (**Table 8**) (**Figure 29**).
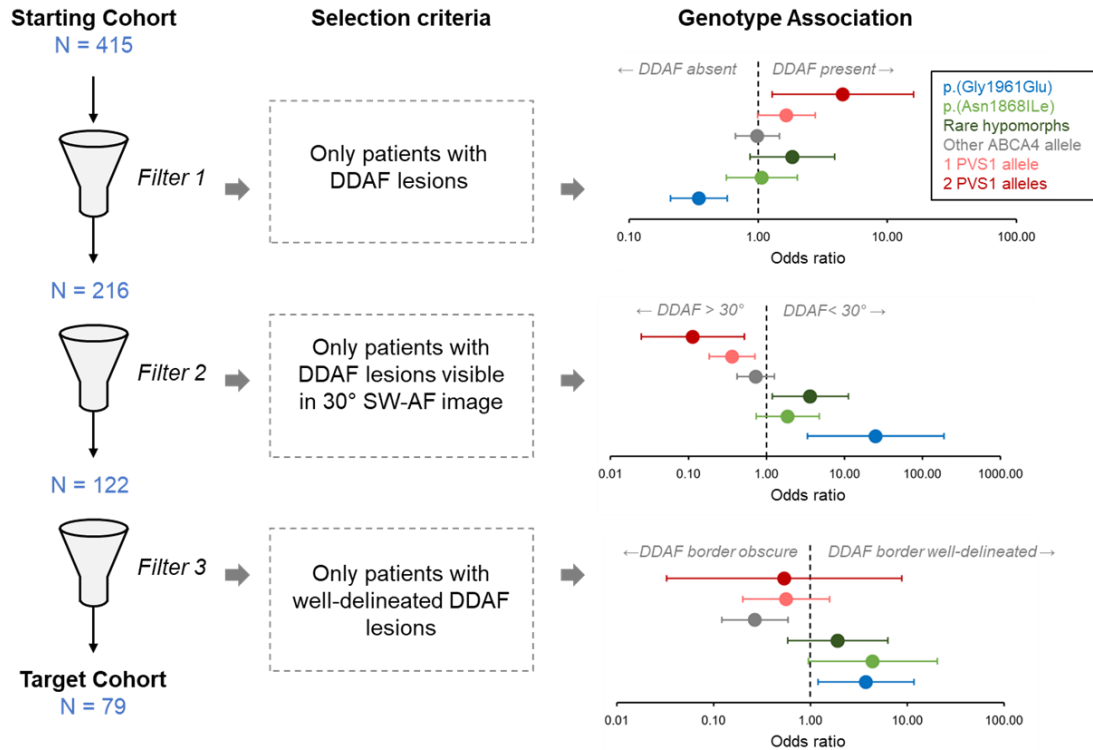
Figure 29. *Genotype associations associated with selection filters.*

After stratifying the linear DDAF growth rates in the Progression Cohort, we found significant differences among genotypes (P < 0.00001, one-way ANOVA; P = 0.00442, Kruskal-Wallis Test) (**Figure 30**) (**Appendix C**). Pairwise differences were significant between p.(Gly1961Glu) (0.096 mm/year) versus Rare hypomorphs (0.226 mm/year) (P < 0.01) and Other *ABCA4* alleles (0.226 mm/year) (P = 0. 00001), and between p.(Asn1868Ile) (0.187 mm/year) and Other *ABCA4* alleles (0.226 mm/year) (P = 0.00255) (**Appendix C**).
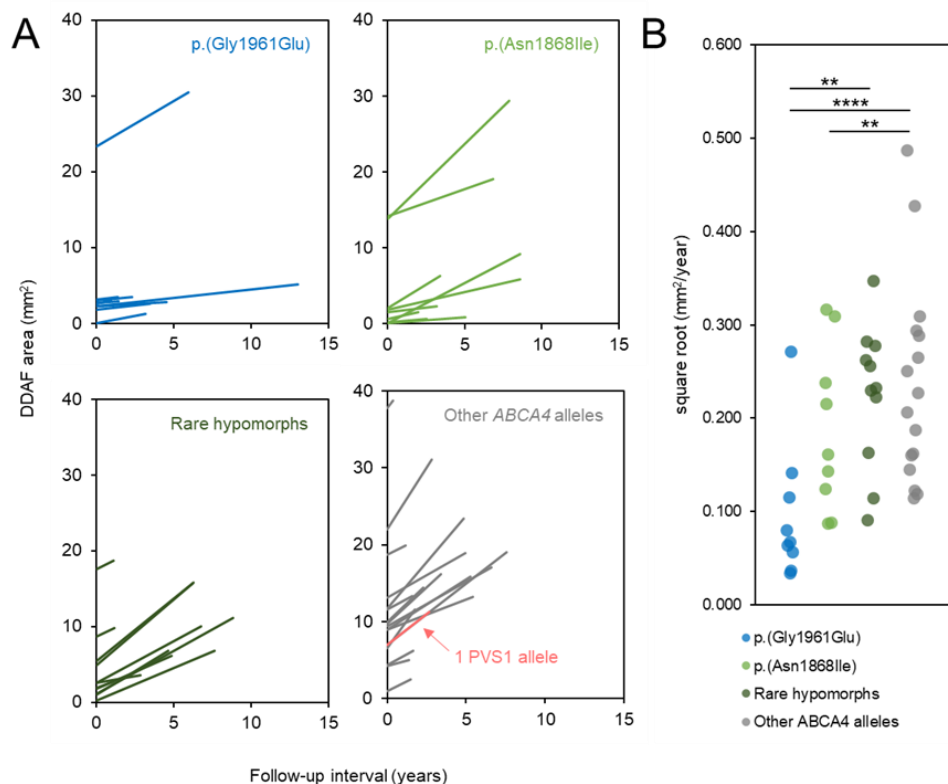
Figure 30. Measured and estimated DDAF growth rates between genotypes in ABCA4 disease.

6.2.4 PVS1 alleles are associated with a distinct and heterogeneous, immeasurable form of DDAF

We found several other significant genotype associations related to the morphometric complexity of individual DDAF lesions. While the majority of DDAF lesions were generally unifocal, enclosed regions of homogenously dark autofluorescence (**Figure 31A**), a distinct form of DDAF characterized by poorly delineated amalgamations of smaller coalescing lesions was found in a subgroup of at least 30 individuals (**Figure 31B**).
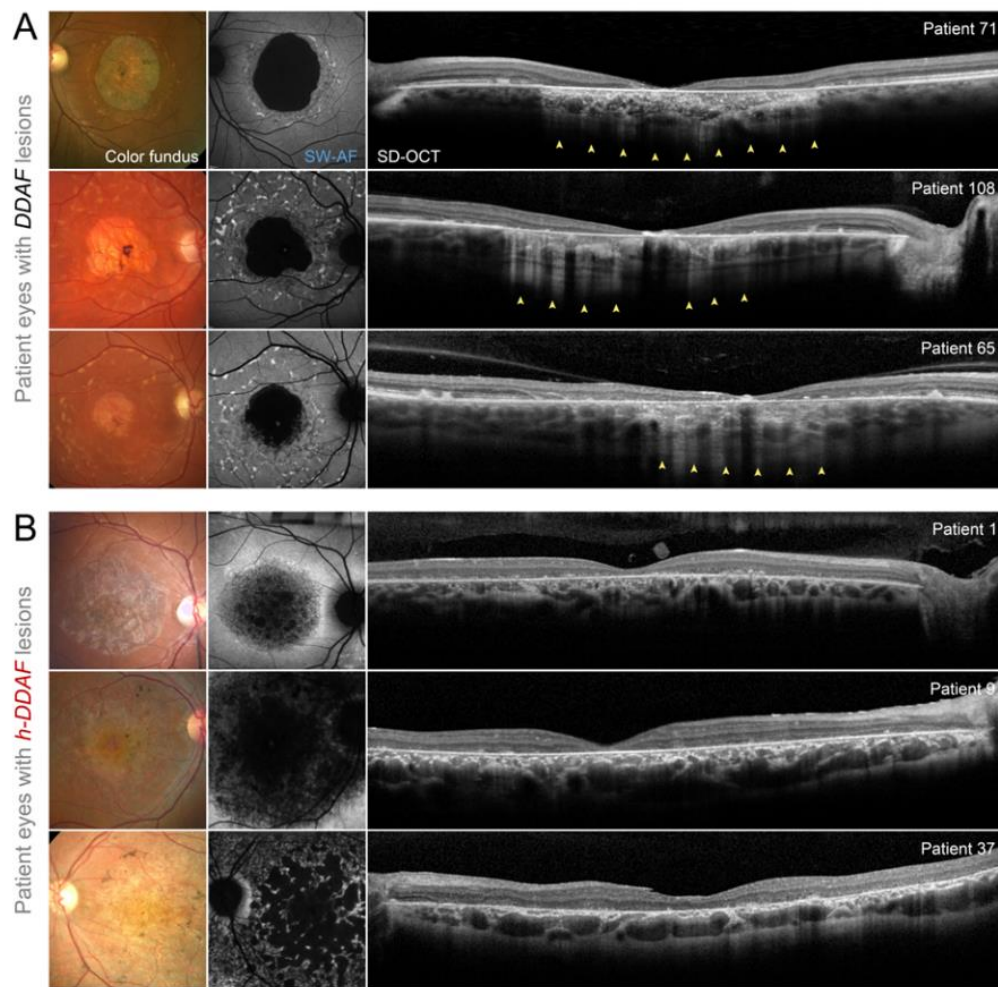
Figure 31. Structural distinction between DDAF and h-DDAF in ABCA4 disease.[u]

This DDAF subtype, which we designate as heterogeneous-DDAF (h-DDAF), was found among the earliest age groups (mean age = 27.5 years, range 10-49 years), including all individuals below age 20 years. Genetically, h-DDAF lesions were almost exclusively associated with severe genotypes, particularly those consisting of 2 PVS1 alleles (76.5%, P < 0.001, 2x3 FET) (**Table 5**).

---

[u] Yellow arrowheads indicate regions hyper-transmission of the OCT signal due to complete loss of outer retinal layers (neuroretina and RPE).

Table 5. *DDAF and h-DDAF prevalence across ABCA4 genotypes.*

ABCA4 genotypes	DDAF	h-DDAF	Other lesion types	Total
p.(Gly1961Glu)	27 (31.8%)	0 (0%)	58 (68.2%)	85
p.(Asn1868Ile)	23 (53.5%)	0 (0%)	20 (46.5%)	43
Rare hypomorphs	21 (65.6%)	0 (0%)	11 (34.4%)	32
Other alleles	77 (47%)	8 (4.9%)	79 (48.2%)	164
1 PVS1 allele	37 (50%)	9 (12.2%)	28 (37.8%)	74
2 PVS1 allele	1 (5.9%)	13 (76.5%)	3 (17.6%)	17
			Cohort total	415

6.2.5 DDAF lesions associated with milder alleles are morphometrically less complex

Among lesions with more well-defined shapes, we observed that those found in eyes with mild allele genotypes were more uniformly round in appearance whereas lesions in eyes with more severe genotypes exhibited more complex shapes (**Figure 32A**).
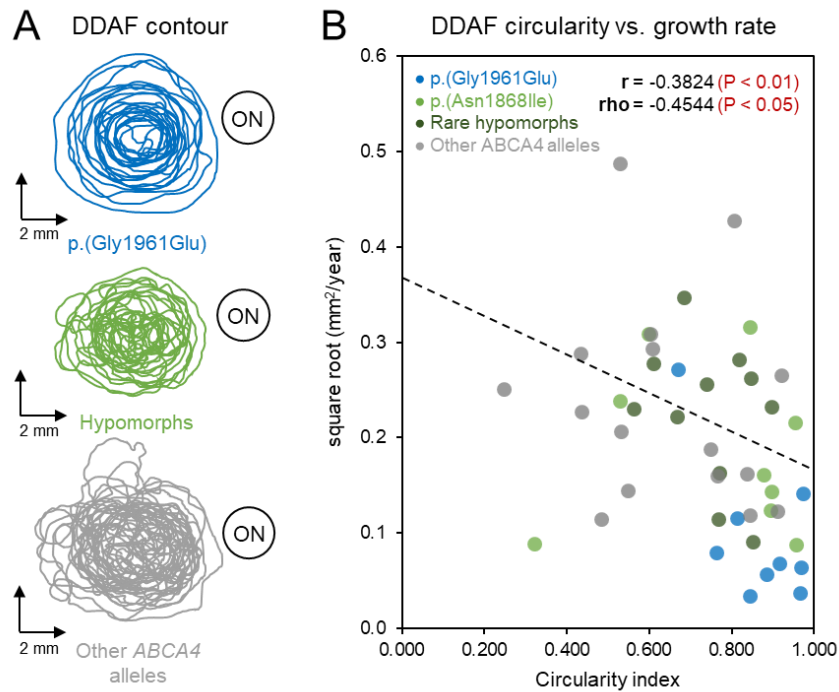
Figure 32. *Morphometric analysis of DDAF circularity and correlation to growth rates.*

91

To assess this quantitatively, we determined the circularity index (CI, $4\pi*area/perimeter^2$) ($0 \leq$ CI ≤ 1, where 1 = perfect circle) for each lesion at the baseline visit in the target cohort (n=79). CI ranged from 0.248 (Patient 350) to 0.973 (Patient 267); however, on average, DDAF lesions in eyes with p.(Gly1961Glu) were the most uniformly circular (CI = 0.866) and those with other *ABCA4* alleles were the least uniformly circular (CI = 0.640) (**Table 6**).

Table 6. *Circularity index of DDAF lesions in ABCA4 disease.*

ABCA4 genotypes	Patients/eyes (n)	Mean age (years)	Baseline perimeter (mm) [SD]	Baseline area (mm2) [SD]	Mean circularity index [SD]
p.(Gly1961Glu)	22	51.2	10.86 [5.39]	10.44 [9.64]	0.873 [0.105]
p.(Asn1868Ile)	14	48.7	7.36 [4.93]	5.02 [6.12]	0.765 [0.195]
Rare hypomorphs	13	52.2	7.80 [3.99]	4.61 [4.81]	0.746 [0.109]
Other alleles	28	49.1	13.07 [5.88]	10.00 [7.88]	0.683 [0.180]
1 PVS1 allele	1	40.3*	17.28*	6.96*	0.293*
2 PVS1 allele	1	24.9*	18.17*	19.22*	0.732*

Given the apparent relationship between genotype, lesion shape (CI) and DDAF growth rate (**Tables 6 and 7**), we calculated significant negative correlation coefficients between CI and DDAF growth rates (Pearson's r = -0.39, P = 0.007655) (Spearman's rho = -0.44, P = 0.00195) indicating that more uniformly circular lesions progress more slowly relative to those of more complex shapes (**Figure 32B**). No significant correlation was found between CI and baseline DDAF lesion size (P = 0.530).

Table 7. *DDAF growth rates across ABCA4 genotypes.*

ABCA4 genotypes	Patients/ eyes (n)	Mean age (years)	Mean baseline DDAF size (mm^2)	Mean surface progression (mm^2/year)	Mean linear progression (mm/year)	Mean circularity index
p.(Gly1961Glu)	9	45.5	4.576	0.355	0.096	0.866
p.(Asn1868Ile)	9	48.8	3.819	0.73	0.187	0.763
Rare hypomorphs	11	52.6	4.355	1.05	0.226	0.746
Other alleles	16	50.4	11.554	1.62	0.236	0.64
1 PVS1 allele	1	40.3*	6.957*	1.61*	0.269*	0.293*
2 PVS1 allele	0	-	-	-	-	-
All	46	49.4	6.854	1.061	0.197	0.726

6.2.6 Outer retinal tubulation are associated with DDAF lesions associated with rare and frequent hypomorphic allele genotypes

Out of 415 patients in the Starting cohort, outer retinal tubulation (ORT) were observed in 18 (4.3%) patients. Most of these patients were generally older with a mean age of 59.4 years (range 31-71 years), reported recent onset of visual symptoms and had relatively preserved vision. On individual SD-OCT scans, ORT varied in size from approximately 100 µm to 500 µm in diameter and were typically observed traversing the outer nuclear layer (ONL) above the RPE. Displacement of the overlying outer plexiform and inner nuclear layers occurred with larger ORT. Morphologically, ORT cross-sections were uniformly round or ovoid with hyper-reflective material within the internal cavities (**Figure 33A and 33D, insets a-f**). ORT typically occurred at the borders of DDAF; however, many were found in areas further away from the lesion borders (**Figure 33B, 33C, 33E and 33F**). The transverse length ORT across the fundus also varied and complex branching was observed in several patients (**Figure 33B and 33E**).
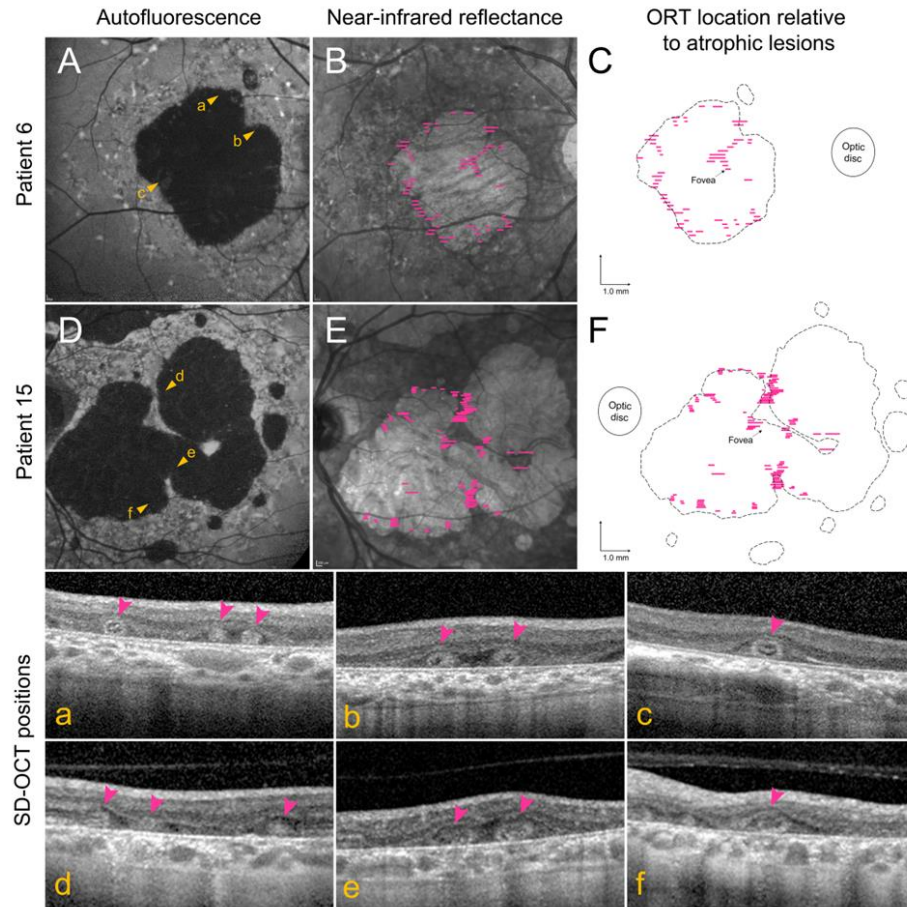
Figure 33. *Spatial distribution and morphology of outer retinal tubulation (ORT) in Stargardt (ABCA4) disease.*[v]

Three hypomorphic *ABCA4* alleles, p.(Asn1868Ile), p.(Ile11562Thr) and p.(Arg2030Gln),

accounted for 10/16 (62.5%) genotypes. Each of the three genotypes was found to be

significantly enriched in the cohort, individually (P<0.05) and combined as a group (P<0.0001)

(**Appendix C**). There was a notable absence of the common p.(Gly1961Glu) allele. The

---

[v] (A, D) Fundus autofluorescence images of right and left eyes of patients 6 and 15, respectively. The positions of several representative ORT are indicated (yellow arrowheads) in each image with corresponding spectral domain-optical coherence tomography scans below (insets a-f). (B, E) Spatial distribution and configuration of the ORT cavitary network mapped onto infrared reflectance images of the same eye in patients 6 and 15. Individual pink lines represent the horizontal diameter and corresponding position of observed on serial SD-OCT scans. (C, F) Location of ORT relative to the borders of atrophic lesions (dotted lines) observed on autofluorescence imaging.

remaining genotypes consisted of both severe and non-hypomorphic variants in patients who presented with ORT at a significantly younger age (44.5 years, P<0.001) (**Appendix C**).

6.3 Discussion

Assessing the growth rate of atrophic (or DDAF) lesions has been the only primary outcome measure for multiple clinical trials of macular diseases including age-related macular degeneration (AMD) and Stargardt disease as approved by the FDA (clinicaltrials.gov).[199,200] For most cases included in these two disorders, however, the incidence, onset and morphology of atrophic lesions can vary significantly between individuals,[127,209-211] thereby affecting the types of patients available for clinical trials. For instance, geographic atrophy is only found in a subset of dry AMD[212] and can present in different patterns.[211,213-217] Furthermore, visible differences have been noted between atrophy in ABCA4 disease and AMD suggesting etiological differences between both entities.[211]

The choice of using atrophy in ABCA4 disease trials appears logical at first, given that it is an easily distinguishable manifestation of tissue loss in the retina and believed to be prevalent among patients.[200] In the cross section of patients included in this study, >50% of patients presented with some kind of atrophy of the fundus. Other large cross-sectional studies have reported similar observations.[121,207] However, after several practical considerations, the remaining pool of directly applicable individuals was found to be much smaller than expected (~20%, 84/418) and with very specific genotypic and demographic profiles.

Based on the results of this study, clinical trials using DDAF growth as an endpoint according to the autofluorescence images standard 30° image field will inevitably assemble a cohort in which more than half of the patients will have genotypes consisting of one of the milder *ABCA4* alleles. The underlying reasons for this are two-fold. The first, and the most significant, is the technical

limitation of image size. While patients with more severe *ABCA4* alleles develop atrophy sooner, the faster progression in these cases results a large proportion of cases with lesions that are too large to measure in a standard fundus image. Wider field imaging modalities are available, including the 55° lens for the Spectralis+OCT module; however, studies have shown that measure accuracy decreases with larger images.[103] Secondly, DDAF lesions in patients with mild *ABCA4* alleles are inherently more reliably measurable. Considering both criteria, an over-representation of mild allele cases is therefore unsurprising.

Demographically, most patients with "measurable" DDAF were concentrated in the 41-60 year age group with none below 24 years of age. Atrophy in younger individuals were observed but the numbers were very small and the atrophic lesions were neither well-demarcated nor organized enough for objective assessment. Most of these individuals belong to the previously described rapid-onset chorioretinopathy (ROC) phenotype group[110] characterized by severe *ABCA4* alleles and onset of symptoms in the first decade followed by a precipitous loss of vision. The "fast-progressing" nature of this phenotype may suggest that these patients appear to be ideal candidates for clinical trials, which desire to demonstrate the most robust results in the shortest period of time; however, an endpoint other than DDAF will be needed.

Choosing an ideal endpoint for ABCA4 disease is indeed challenging. While there are many existing imaging and functional modalities, there is likely none that will encompass the entire spectrum of heterogeneous disease such as those attributable to *ABCA4*. Based on the findings of this study, the DDAF endpoint may also be hindered by time, given the known rates of growth ($\sim$0.6-2.0 mm^2/year),[121,201,204,218] and the fact that most patients included will be on the milder end of the severity spectrum. Depending on the type of treatment being evaluated, it is also unclear what the short-term therapeutic effects will be on DDAF. For instance, it is not clear at

what threshold of bisretinoid accumulation commits the cell to degeneration resulting in DDAF. Visual cycle modulation or A2E sequestration may fix an upstream problem but may not alleviate deleterious effects already imparted on the cell. Therefore, even with long-term efficacy, existing lesions may continue to grow over the evaluation period of a trial. For such situations, more specialized methods such as quantitative autofluorescence (qAF)[22,103,219] may be a better indicator, or should at least be considered in conjunction with other acceptable endpoints.

In summary, the current study demonstrates that while atrophy, specifically DDAF, is a relatively common feature of ABCA4 disease, there are far fewer cases available for clinical trial selection than expected, when considering technical considerations such as measurability and imaging capacity. It is important to note that further exclusion criteria such as systemic health and geography, among others, will likely further reduce this pool of available candidates. Based on the criteria of this study, most patients will also likely have mild allele genotypes which could potentially extend the timeframe necessary to show efficacy. Depending on the type of treatment being evaluated, the quantification of DDAF may not be the ideal endpoint for ABCA4 disease and other, more encompassing, modalities should be considered.

6.4 Materials & Methods

6.4.1 Study subjects and ethics approval

Written and signed informed consent to participate in this study were obtained from all individuals described in this study. All procedures were performed with approval by the institutional review board protocol #AAAI9906 at Columbia University Irving Medical Center. The study adhered to all tenets set out in the Declaration of Helsinki. All patients enrolled in the study underwent a complete ophthalmic examination, which included slit-lamp and dilated-

fundus examinations, best-corrected visual acuity (Snellen) measurement and assessment of clinical histories.

6.4.2 Molecular analysis

Direct sequencing of the *ABCA4* gene was performed as previously described.[34,79] All detected possibly disease-associated variants were confirmed by Sanger sequencing and analyzed for pathogenicity based on familial segregation, allele frequency and *in silico* predictions. Phase was either confirmed by screening of familial samples or imputed based on observed co-occurrence in the general population (https://gnomad.broadinstitute.org/variant-cooccurrence) if family members were available. Functional annotation of variants was determined using computational software including ANNOVAR using pathogenicity scores of M-CAP, REVEL, Eigen, and CADD-PHRED (v1.6). As a general guideline, pathogenic consequences are predicted for variants with scores over 0.025 for M-CAP, 0.5 for REVEL, 0.5 for Eigen, and 20 for CADD-PHRED. The allele frequencies of all variants were compared to those in the Genome Aggregation Database (accessed October 2021).

6.4.3 Retinal image acquisition and analysis

All genetically confirmed patients with a well-captured 30° field short-wavelength autofluorescence (SW-AF, 488-nm excitation) images and fovea-centered spectral domain optical coherence tomography (SD-OCT) from the Spectralis HRA + OCT scanning laser ophthalmoscope (Heidelberg Engineering) were included in the study (n=415).

Definitely decreased autofluorescence (DDAF) lesions were defined based on Strauss et al.[207] and confirmed by the presence of "choroidal hyper-transmission" on corresponding SD-OCT scans from the same visit (Filter 1). The distinction between DDAF lesions with well-defined or obscure borders (Filter 3) was made by two independent graders (WL and PYS). Disagreements

between the two primary graders were adjudicated by a third independent grader (SHT). All

graders were masked to patient genotypes each of the selection steps (cohort assembly, Filter 1,

Filter 2 and Filter 3). The area (in mm^2) and perimeter (mm) of DDAF lesions were also

measured by two independent graders (WL and PYS) using the RegionFinder function (HEYEX

software, Heidelberg Engineering). DDAF growth rate was calculated as the difference between

follow up and baseline DDAF size (mm^2). To account for the effect of the baseline lesion size

on the growth rate, the linear growth rate was determined by square root transformation of the

baseline and follow up visits. In this analysis, graders were masked to both genotype and time-

point (visit 1 vs. visit 2). The study eye was determined at random. Intraclass correlation

coefficients (ICC), calculated using the irrICC package (https:// cran.r-project.org), showed

excellent intergrader agreement for all area (ICC > 0.8999) and perimeter (ICC > 0.9008)

measurements.

Lesion contour analysis was performed on 30° SW-AF images of both eyes of all patients with

"measurable" lesions (n=79, 158 eyes) using a custom plugin in the FIJI software (National

Institute of Mental Health, Bethesda, MD, USA). All images were normalized according to the

internal scaling factor of the image (given in pixels/μm). Left eyes were flipped along the

vertical meridian so that both eyes are represented together (i.e. the nasal retina is left of center

for both eyes) aligned at the fovea. Circularity index was determined as follows:

$$Circularity\ index = \frac{4\pi(DDAF\ area)}{DDAF\ perimeter^2}$$

Where the circularity index (CI) (also known as isoperimetric quotient) is the ratio between 4π

times measured area of the DDAF lesion and DDAF perimeter squared. CI values range from 0

to 1, where 1 is a perfectly uniform circle.

The presence of outer retinal tubulation (ORT) was assessed in volumetric SD-OCT raster scans in both eyes of each patient. Transverse mapping of ORT onto corresponding positions in fundus images were performed manually using the spatial correlation tool in the HEYEX software (Heidelberg Engineering, Heidelberg). Lesion perimeters were outlined and merged with ORT positions using the pixel thresholding function and alignment function, respectively, in FIJI (National Institute of Mental Health, Bethesda, MD, USA). Image proportions were maintained according to the intrinsic scaling factor (pixels/μm) of each fundus autofluorescence image.

6.4.4 Statistical analysis

A detailed summary of all statistical calculations is provided in the **Appendix C**. Comparison of means were determined by both 1-way ANOVA test with post hoc Tukey's honestly significant differences and Kruskal-Wallis tests for non-normal distributions. Significance level was set at α level less than 0.05. Correlation between CI and square root transformed DDAF area were assessed using Pearson's correlation coefficient (r) and Spearman's rank correlation coefficient (ρ, rho).

Chapter 7 *CERT1* mutations perturb human development by disrupting sphingolipid homeostasis

7.1 Introduction

Sphingolipids play numerous essential roles in membrane structure, signal transduction, and brain development and function [220-222]. The central nervous system is particularly affected by disturbances in sphingolipid production or clearance: defective production can cause hereditary sensory neuropathy, spastic paraplegia, or infantile epilepsy syndrome, whereas toxic accumulation of sphingolipids underlies a number of devastating inborn errors of metabolism such as Gaucher disease, Farber disease, Niemann-Pick type A, Krabbe, Tay-Sachs, and Sandhoff diseases [223,224]. Sphingolipid metabolic fluxes therefore must be tightly regulated through homeostatic circuits [225]. A key checkpoint in sphingolipid biosynthesis occurs at contact sites between the endoplasmic reticulum (ER) and the *trans* Golgi membrane, where the Ceramide Transporter (CERT) transfers ceramide (Cer) from the ER to the *trans* Golgi for its conversion to sphingomyelin (SM) [226]. When sufficient rates of SM production are reached, CERT is phosphorylated and undergoes a conformational change that renders it inactive [227].

Given the central position of CERT in sphingolipid metabolism, its malfunction should be detrimental to human health, and particularly to neural function. Thus far, one case report and several case-control screening studies have described associations between variants in *CERT1* (ceramide transporter 1), the gene that encodes CERT, and neurological abnormalities [228-236]. Nonetheless, there has not been a systematic assessment of the mutational landscape of *CERT1* in humans, and whether or how *CERT1* mutations cause neurological disease remains to be proven.

In this chapter, we characterized 31 unrelated individuals with 22 distinct missense variants in *CERT1*, 18 of which have not been previously reported. These patients have a syndromic presentation characterized by infantile hypotonia; mild dysmorphologies (affecting the face, hands or feet); variable degrees of intellectual disability, motor and speech delays; increased pain tolerance; and seizures.

101

Text and figures from this chapter have been recently published (Gehin, Lone, Lee and Capolupo et al. *J Clin Invest. 2023 Mar 28:e165019. doi: 10.1172/JCI165019.*

7.2 Results

7.2.1 Characterization of human CERT1-associated phenotypes

Most of the *CERT1* variants reported so far have been associated with intellectual disability,[237] but the clinical phenotype has been comprehensively characterized for only one of these individuals, who bears the p.(Ser135Pro) variant.[233] We therefore sought additional affected individuals harboring *CERT1* variation from multiple international disease consortia and databases. We identified 50 individuals who carry a potentially pathogenic missense variant in *CERT1* (**Supplemental Table 8 and Supplemental Table 9**). Notably, we did not identify any cases with protein truncating variants or large CNV. We obtained thorough clinical histories for 31 patients (**Supplemental Table 8, Appendix D**): 27 reported here in present study cohort, three who were initially included in genetic screening consortia for intellectual disability,[229,231,238,239] and the patient who had been previously characterized.[233] Familial segregation confirmed that *CERT1* variants occurred *de novo* in 93% (25/27) of cases. Biparental samples were not available for S1, S19, S20, or S24. The p.(Val326Phe) variant in S21 was inherited from her reportedly unaffected father, whereas the p.(Ala449Val) variant in S26 was inherited from her mother, who was diagnosed with intellectual disability.

The detailed clinical information we were able to gather revealed cognitive, motor, and speech delays of variable degrees of severity. Of the subjects for whom we had neonatal information, only four were of average weight at birth; most were born slightly to significantly underweight. Similarly, only four of the subjects did not present some form of apparent developmental delay

by the end of the first year of life (4/26, 15%), with the latest onset being at age four. Fifteen (out of 24) had neonatal feeding difficulties, often accompanied by hypotonia or failure to thrive. These were likely early manifestations of what would later become frank motor delays, affecting 26 out of 29 subjects (**Figure 34A**). Intellectual disability ranged from mild to profound, as per the criteria of the Diagnostic and Statistical Manual of Mental Disorders (5th edition) (DSM-V) (**Figure 34B**); the latter subjects are nonverbal, lack age-appropriate daily living skills and require safety supervision. Neurobehavioral abnormalities frequently led to a diagnosis of autism spectrum disorder (ASD) (19/27, 70%); some patients displayed stereotypic hand movements (14/18), self-injurious behavior (9/19), high pain tolerance (9/18), disrupted sleep patterns (9/21), attention deficit-hyperactivity disorder (ADHD) (10/19), or aggression (6/20). Multiple seizure semiologies were reported (16/29). Neuroimaging frequently revealed a thin corpus callosum, ventriculomegaly, delayed myelination, and cerebellar atrophy.
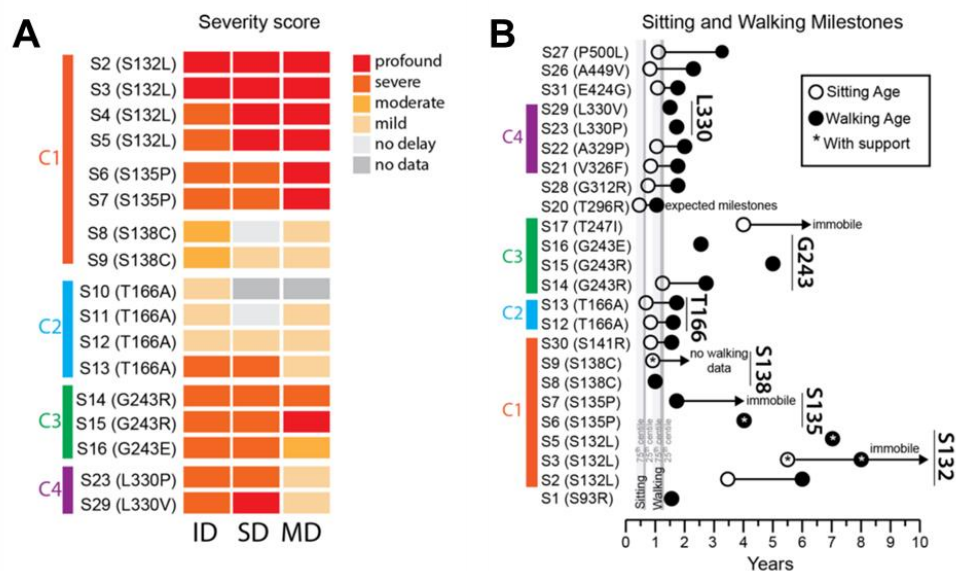
***Figure 34**. Clinical characterization of individuals with pathogenic variation in CERT1.[w]*

---

[w] (A) Range of severity in motor delays compared to the 75th percentile (light grey) and 25th percentile (dark grey), adapted from values published by the Denver Developmental Screening Test II. Open and closed circles indicate

103

Subtle facial dysmorphisms included anteverted nares with depressed or broad nasal bridge, enlarged earlobes, synophrys, micrognathia, dental anomalies (protruding incisors and diastema), and palatine ridges. Anomalies affecting the hands, feet, or digits included 3rd/4th finger syndactyly, club foot, or hallux varus (sandal gaps); the first metatarsal also tended to be short, while the fifth fingers tended to be long (**Figure 35**).

---

delayed sitting and walking ages, respectively; asterisks indicate the subject needs sitting or walking support. Open arrows indicate subjects who are currently immobile or have not yet developed independent walking. Clusters mutations belong to are indicated. (B) Heatmap displaying the degree of intellectual disability (ID), speech delay (SD), and motor delay (MD) of patients bearing frequent CERT1 mutations. See Supplemental Table 2 for scores.
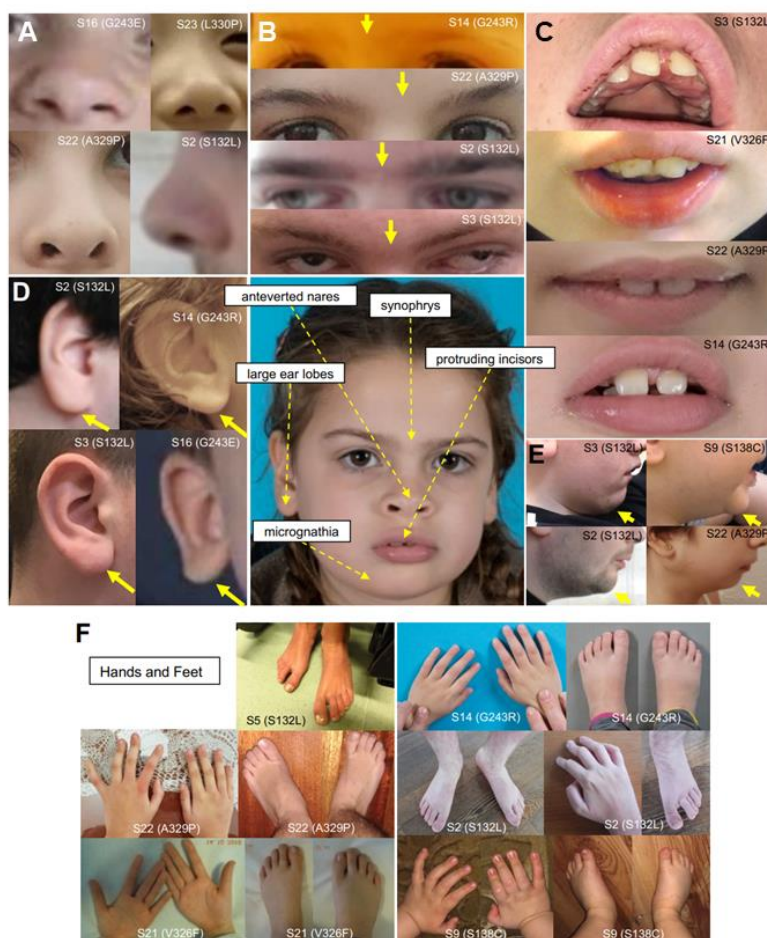
Figure 35. Craniofacial and peripheral dysmorphologies in CerTra subjects.[x]

7.2.2 Genotype-Phenotype Correlations

The majority (27/31, 87%) of variants from our enrolled patients occurred in the region between the PH and START domains (**Figure 36**), as was the case for previously reported variants.[233]

The first four serine residues of the SRR (132-SMVSLVSGASGYSATSTSS-150) are hotspots for mutations (*Cluster 1*): Ser-132 (n=7; four enrolled), Ser-135 (n=4; two enrolled), Ser-138 (n=4; two enrolled), Ser-141 (n=1; one enrolled) (**Figure 36**). An alanine substitution at Thre-

---

[x] Enlargement of craniofacial dysmorphologies in patients showing: (A) anteverted nares, (B) synophrys, (C) protruding incisors, (D) large ear lobes, (E) micrognathia, and (F) hand and foot abnormalities.

166 was found in five subjects in our cohort (four enrolled; *Cluster 2*). Between residues Asp-240 to Gly-254 (*Cluster 3*), missense variants were found at Asp-240 (n=1), Gly-243 (n=4, four enrolled), Thr-247 (n=1, one enrolled), and Thr-251 (n=1, one enrolled). Four variants p.(Val326Phe), p.(Ala329Pro), p.(Leu330Val), and p.(Leu330Pro) are located at the C-terminal end of the CERT FFAT motif (*Cluster 4*).
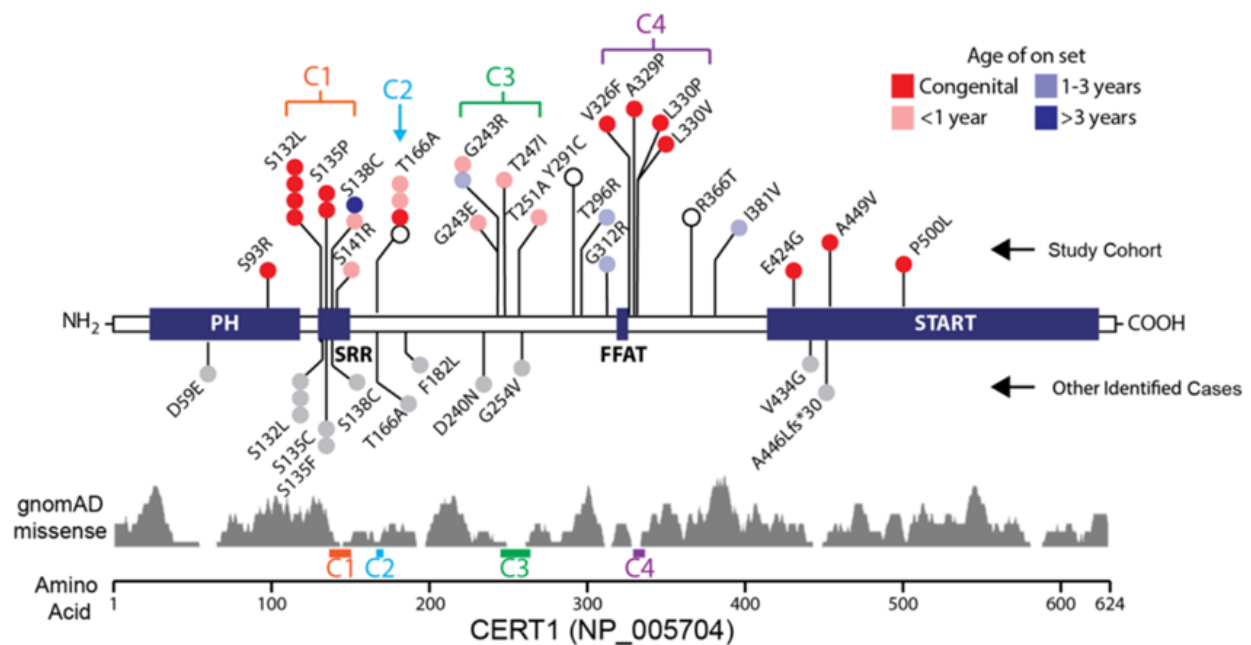
Figure 36. Missense variants in CERT1 lead to a neurodevelopmental syndrome.[y]

---

[y] Schematic representation of functional domains in CERT. The N-terminal Pleckstrin Homology (PH) domain interacts with phosphoinositide phosphatidylinositol-4-phosphate [PtdIns(4)P] on the trans Golgi. The serine rich region (SRR) is the target of protein kinase D (PKD) and casein kinase 1γ2 (CSNK1G2) phosphorylation. The FFAT (two phenylalanines in an acidic tract) motif interacts with integral membrane proteins VAP-A and VAP-B on the ER (82), and a C-terminal (StAR)-related lipid transfer (START) domain extracts ceramide from the ER membrane and delivers it to the trans Golgi. The schematic shows coding variants in CERT1 (NP_005704) in our cohort of 31 subjects above the gene diagram and other individuals identified from clinical databases (Decipher v9. 31, ClinVar and VKGL) below it. Colors indicate the age of onset; grey, no information available. The distribution of gnomAD singleton missense variants for healthy subjects is plotted below.

For the variants that recurred in multiple subjects, we were able to sketch out broad genotype-phenotype correlations. The most severe phenotypes—with congenital or perinatal onset, profound to severe intellectual disability, and the greatest motor delay—resulted from mutations at Ser-132 and Ser-135 (**Figure 34** and **Figure 36**). Subjects bearing mutations at Ser-138, Thr-166, or Gly-243 tended to not have perinatal difficulties and achieved early developmental milestones (S8 did not show difficulties until 4 years of age) but then regressed or slowed in their development. Seizures likely contribute to the developmental delays: neither subject with a p.(Ser138Cys) mutation had seizures and have only moderate ID, but S13, the most severely affected of the p.(Thr166Ala) carriers, has an epileptic encephalopathy that apparently halted her development at 19 months. All variants associated with more severe phenotypes were predicted to be deleterious by CADD-PHRED, REVEL, M-CAP, and Eigen. Approximately half of the variants among the more moderately affected had inconsistent pathogenicity predictions across algorithms. On average, variants among patients had significantly greater CADDv1.6, M-CAP, REVEL and Eigen pathogenicity scores than gnomAD singleton missense variants (P<0.001, Mann-Whitney U-test).

Variants p.(Asp59Glu) (16), p.(Thr166Ala) (20) and (Phe182Leu) (10) were previously identified in ASD cohorts (**Supplemental Table 9**). To determine the extent to which *CERT1* mutations contribute to autism, we conducted a targeted *de novo* analysis interrogating the Simons Foundation Powering Autism Research (SPARK) initiative database, which includes genomic data from nearly 35,000 ASD individuals. No *de novo* protein-altering variants were observed in the Simons exome cohort. Denovo-db (v.1.6.1) confirmed the negative findings for the exome cohort but identified one de novo genome missense mutation (pos=74721285, T>C, c.880A>G, p.(Thr294Ala), exon 6) in one out of 516 trios. This analysis, along with the fact that

none of the subjects in our cohort was diagnosed with ASD as a primary condition, indicates that

CERT1 variants are unlikely to be a significant contributor to autism but instead cause a

recognizable neurodevelopmental syndrome distinct from ASD, which we will refer to as

Ceramide Transporter (CerTra) syndrome.

7.3 Discussion[z]

In this study we report that multiple *de novo* heterozygous variants in *CERT1* cause an autosomal

dominant developmental syndrome we named CerTra, which is characterized by various degrees

of developmental delay, motor delay, cognitive impairment, behavioral abnormalities, and

seizures. Several CerTra mutations reduce or abolish the capability of CERT to be inactivated in

response to excessive sphingomyelin production. This leads to uncontrolled ceramide transfer

out of the ER, with several consequences: i) reduced ceramide at the ER relaxes the homeostatic

inhibition of serine palmitoyltransferase [240,241], resulting in increased *de novo* sphingolipid

synthesis; ii) reduced ceramide at the Golgi, the site for glucosylceramide synthesis, impairs

glycosphingolipid production; and iii) excessive CERT activity likely competes with ceramide

desaturase activity at the ER, leading to the transfer of a significant amount of dhCer to the *trans*

Golgi and production of dhSM. Elevated dihydro-sphingolipid levels cause neuropathology. For

example, biallelic loss-of-function mutations in *DEGS1*, which catalyzes the final conversion of

dhCer to Cer, results in increased dihydro-sphingolipid levels, causing a multisystem

---

neurological disorder of both the central and peripheral nervous system characterized by hypomyelination and leukodystrophy (MIM# 618404) [242,243]. Recessively inherited loss-of-function mutations in alkaline ceramidase 3 (*ACER3*) also result in increased dihydro-sphingolipid formation and a leukodystrophy phenotype (MIM# 617762) [244]. Impaired glycosphingolipid synthesis is a further cause of neuropathology: loss-of-function mutations in *ST3GAL5* or *B4GALNT1*, which encode two key enzymes in ganglio-series glycosphingolipid synthesis (i.e., GM3 and GM2 synthases), cause neurodevelopmental regression (MIM# 609056) and spastic paraplegia (MIM# 609195), respectively [245-250]. Recently, a specific group of mutations in serine palmitoyltransferase subunits were reported to cause childhood-onset amyotrophic lateral sclerosis [251-253], and hereditary spastic paraplegia [254]. All identified mutations disturbed the homeostatic control of *de novo* sphingolipid synthesis, resulting in greatly elevated sphingolipid levels.

Several CerTra mutations cause lipid metabolic derangements similar to those associated with the above listed conditions, nonetheless, the molecular mechanisms of CERT dysregulation appear to be varied. Mutations in the SRR directly impair CERT-inactivating phosphorylation, while mutations in the coil-coil domain (CCD) hinder the conformational change that follows this event. We have not directly addressed the effects of other mutations close to the FFAT motif or in the PH and START domains. These mutations likely affect CERT interaction with VAP proteins [255] and the final events of CERT inactivation that involve intramolecular interactions between the PH and START domains [256]. Four consecutive variants (p.(Arg366Tgr), p.(Ile381Val), p.(Glu424Gly), and p.(Ala449Val)) do not affect CERT localization nor phosphorylation state. When analyzed in greater depth these variants present some doubts about their actual pathogenicity: i) each of them is associated with a single patient and they are not part

of a recognizable cluster of mutations; ii) p.(Ile381Val) has been found in one healthy subject; iii) p.(Ala449Val) has been found in a subject whose parents (both) presented with intellectual disability independently of the mutational state of *CERT1*; iv) p.(Arg366Thr) has been found in a subject where the very little information we have would point towards a very mild phenotype. Further studies will be required to assess the actual role of these variants and, more in general, to untangle the cellular mechanisms by which sphingolipid metabolism disruption leads to CerTra syndrome. Nonetheless, the characterization of this new disease entity and of the most frequent variants associated with it has already revealed unforeseen operating principles of sphingolipid homeostasis, led to the definition of a new structural region in CERT, and delineated a possible use of CERT inhibitors for the treatment of CerTra patients.

7.4 Materials & Methods

7.4.1 Subject identification and clinical characterization

Initial screening and identification of anonymized individuals harbouring *CERT1* variation was conducted through the following public databases and tools: ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), Decipher [257], GeneMatcher [258], and VKGL-NL Rotterdam (https://www.vkgl.nl/nl/) (**Supplemental Table 10**). Each patient underwent a full clinical examination by a neurologist and/or medical geneticist. Clinical data were directly abstracted from medical records provided by the referring clinician(s) which included behavioral assessment and electroencephalograms (EEG). Developmental delay and cognitive ability tests were performed on subjects using the following tests: Gesell Developmental Schedules, Chinese National Health Commission Developmental Evaluation Scale, Wechsler Intelligence Scale for Children, Peabody Picture Vocabulary Test-4 (PPVT-4) and Bayley Scales of Infant Development II. When possible, standardized assessment of impairment in conceptual, social,

and practical domains in each subject was made in accordance with the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) and noted to be mild, moderate, severe or profound. Degrees of intellectual disability (ID) by verbal and nonverbal IQ scores are classified as follows based on verbal and nonverbal IQ scores: mild (IQ 50-55 to ~70), moderate (IQ35-40 to 50-55), severe (IQ 20-25 to 35-40), and profound (IQ <20-25).

To characterize the craniofacial/skeletal dysmorphia, we performed a deep phenotyping analysis for all the patients whose families agreed to share photos. This was done by the team of Eva Bermejo-Sánchez at the Institute of Rare Disease Research (IIER), Madrid (Spain). Dysmorphology analyses were performed blind to the variant detected for each patient. Each parameter was compared to age-, sex-, and ethnicity-matched healthy cases. Briefly, a detailed reading of all the clinical reports was performed to extract a first list of dysmorphic features that was then used to evaluate all the subjects. All the available photographs were assessed based on those features, giving a matrix of 96 rows that were scrutinized for all patients. Thus, the same traits were assessed by the same observer using homogeneous criteria for all the patients. See Supplemental Clinical Appendix for all the features for each patient.

7.4.2 Sequencing and genetic analyses

A table summarizing all genetic testing procedures and results is provided in **Supplementary Table 10**. Before the identification of *CERT1* mutations by trio or singleton exome sequencing, most patients underwent extensive genetic screenings that included karyotyping, aCGH, CAG and CGG repeat expansion and metabolic testing, none of which was significant. All subjects underwent various diagnostic screening procedures including standard-of-care tests for neurodevelopmental diseases such as karyotyping, chromosomal microarray or panel sequencing, and testing for Fragile X syndrome.

CERT1 variants were identified by either research or clinical diagnostic whole exome sequencing. Methodology and analysis pipelines for each are provided in detail in **Supplementary Table 10**. Allelic segregation was verified in all except four subjects (S1, S19, S20 and S24) for whom familial samples were not available.

7.4.3 Pathogenicity analysis

All identified variants were analysed in accordance with ACMG Guidelines for variant interpretation and classification [259]. Minor allele frequencies (MAF) of all *CERT1* variants were obtained from The Genome Aggregation Database (gnomAD) [260], BRowse All Variants Online (BRAVO) (https://bravo.sph.umich.edu) and the 1,000 Genomes Project (https://www.genome.gov/27528684/1000-genomes-project). Functional annotation of variants was carried out with ANNOtate VARiation (ANNOVAR) [261] using pathogenicity scores of M-CAP [262], REVEL [263], Eigen [264] and CADD [264]. As a general guideline, pathogenicity is predicted for variants with scores over 0.025 for M-CAP, 0.5 for REVEL, 0.5 for Eigen and 20 for CADD. Prediction scores of CerTra variants were compared to other missense variants found in singleton cases amongst the general population in gnomAD (i.e., missense variants reported in only one healthy individual in the gnomAD database).

Chapter 8 Expanding the Genetic and Clinical Landscape of *PUM1*-associated diseases

8.1 Introduction

The *PUM1* gene encodes the ubiquitously expressed RNA-binding protein Pumilio homolog 1 (PUM1) which has critical role in post-transcriptional processes fundamental to the development, maintenance and homeostasis of various cell types throughout the body, particularly neurons.[67,70,74] The importance of PUM1 in neurons is punctuated by the finding that *Pum1*-haploinsufficient mice (*Pum1*[+/-]) progressively develop ataxia and subsequent Purkinje cell degeneration over time.[67] The phenotype in knockout mice (*Pum1*[-/-]) is comparative more severe and, interestingly, etiologically distinct. Specifically, *Pum1*[-/-] progeny born at a significantly low Mendelian ratio, are considerably smaller in size, have seizures, and exhibit more pronounced cerebellar degeneration.[67] This phenotypic heterogeneity is recapitulated in patients, where mild heterozygous missense variants such as p.(Thr1035Ser) that reduce PUM1 levels by ~25% result in an adult-onset pure cerebellar ataxia phenotype known as PRCA (PUM1-related cerebellar ataxia), and large genomic deletions reducing its levels to ~40-60% underlie a more severe, early-onset neurodevelopmental disorder known as PADDDAS (PUM1-associated developmental delay and seizures).[67]

Many diseases are known to exhibit clinical heterogeneity however the degree to which PADDAS and PRCA differ due to such modest alterations in gene dosage is unprecedented. Understanding this phenomenon requires investigation at multiple levels including elucidation of the full clinical landscape of possible disease variation and phenotypes in patients. Since the publication of the original cohort of patients, a few additional sporadic cases with PADDAS (n=4),[72,77,78] PRCA (n=2),[265] West syndrome (n=1)[266] and an epileptic Dravet-like syndrome

$(n=1)$[76] have been described. To accelerate disease characterization in this rare but important disease, we recently spearheaded a multicenter project to collect and study all affected individuals in the world harboring pathogenic variation in *PUM1*. Although the study is ongoing, we have identified over 60 sporadic and familial cases, the analysis of which extensively expands our current knowledge of the disease landscape.

8.2 Results

8.2.1 Study cohort

In total, 62 affected individuals (from 54 families) with putatively pathogenic variants in *PUM1* were included in the study. Demographic, genetic and clinical characteristics are summarized in **Supplemental Table 11**. The mean and median age (years) at the most recent clinical examination were 20 and 9 years, respectively (range 0.1-81 years, IQR = 30). The age distribution was heavily skewed towards individuals below 20 years of age ($\beta_2 = 3.49$). The female-to-male ratio was 28/26 (1.08). In addition to PADDAS and PRCA, several other distinct clinical entities were represented in this cohort including an apparent adolescent-onset form of PRCA (mean onset age = 6.25, SD = 5.6) and isolated developmental delay without ataxia. Both *de novo* and inherited heterozygous variants in *PUM1* were identified as the strongest or only possible causal variant in each patient following a diagnostic genetic screening (*See Methods*) (**Supplemental Table 12** and **Supplemental Table 13**).

8.2.2 Expanded genotypic profile of PADDAS: large deletions, intragenic duplications and splice variants

As originally described,[72] the prominent phenotype causing PADDAS are copy number variants (CNV), specifically large deletions spanning the *PUM1* the locus. The current study includes 6 additional simplex (non-familial) cases (S15, S16, S17, S58, S67 and S68) with PADDAS

114

features harboring *de novo PUM1*-spanning chromosomal deletions (**Figure 37, Supplemental Table 12**).
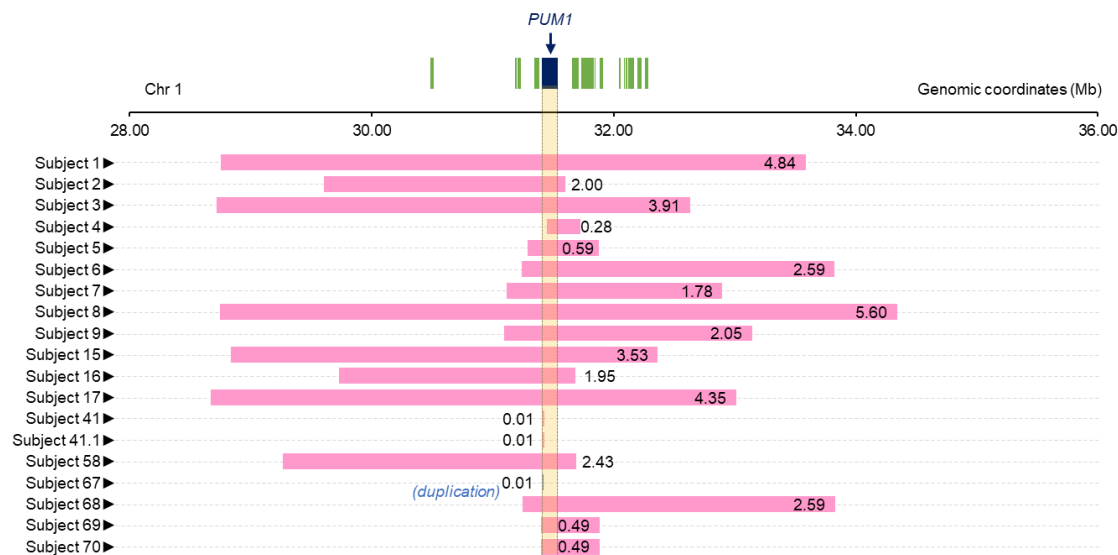
Figure 37. Copy number variants (CNV) spanning PUM1.[aa]

All six subjects were documented to have developmental delay (intellectual disability, speech delay, etc.), cerebellar features (ataxia, gait abnormalities or incoordination), congenital dysmorphia and seizures. The earliest symptoms in each case were epileptic spasms or tonic-clonic seizures first noted around infancy (post-natal to 4 months) apart from Subject 68, who experienced his first generalized tonic seizure at approximately 8 years of age. Four additional non-CNV loss-of-function (LOF) variants in patients with PADDAS were identified a including a 9.04kb intragenic duplication chr1:31409393:31418428:DUP (StrVCTVRE score = 0.836) in Subject 67 (**Figure 37, Supplemental Table 12**), canonical splice site duplication, c.1158+1_1158+2dup (SpliceAI Δscore = 1, donor loss) (Subject 18) (**Figure 37**), stop-gain at

---

[aa] Large genomic deletions (pink) and duplications (blue) including PUM1 (minimal region in yellow shaded area) on chromosome 1p35.2 (shown in red) were identified in 19 patients. S41 and S41.1, and S69 and S70 are related individuals (17 independent alleles).

exon 15, p.(Arg837*) (Subject 38) (**Figure 38**)[77] and a 4 nucleotide deletion variant in exon 21

(c.3267_3270del) resulting in a frameshift resulting in premature downstream termination codon

(p.(His1090Profs*16)) (**Figure 38**).
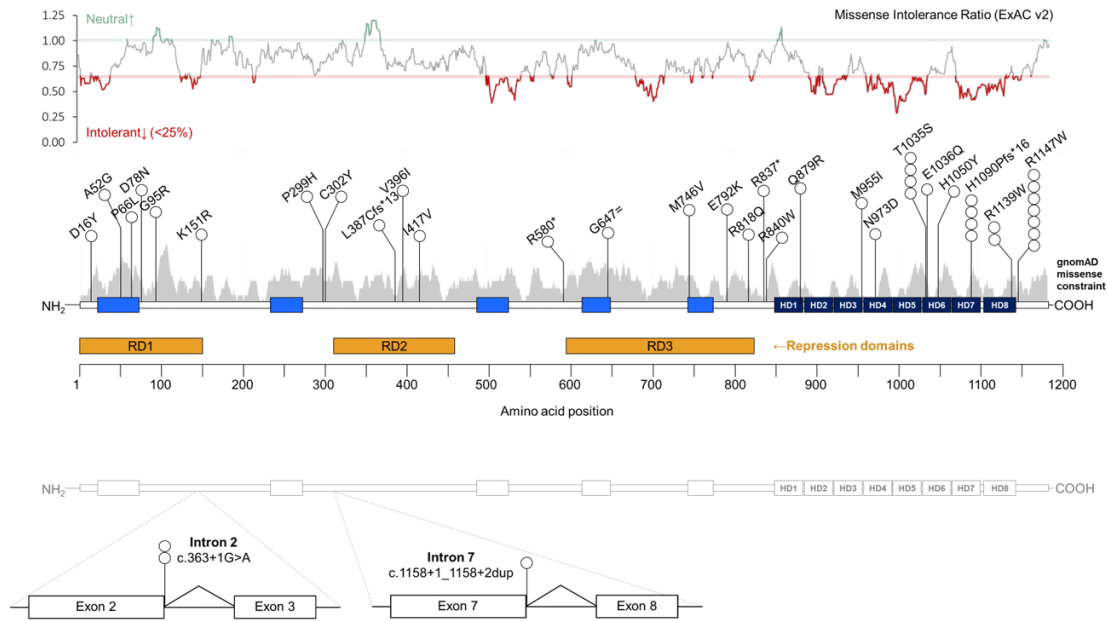
Figure 38. Lollipop plot of coding variants across the PUM1 protein (NP_001018494.1).[bb]

8.2.3 Phenotypic characterization of p.(Arg1147Trp)-associated PADDAS

This study also includes 3 novel cases harboring the p.(Arg1147Trp) missense variant, in

addition to the 4 previously documented sporadic cases,[72,76-78] increasing the current total

number of known *de novo* p.(Arg1147Trp) alleles to n=7. The clinical phenotype of these 7

cases were largely consistent with the PADDAS phenotype similar to those with LOF variation

---

[bb] The eight homology domains (HD) (dark blue) and low complexity regions (light blue) are shown. The distribution of missense variants across PUM1 in gnomAD are shown in gray (valleys indicate constrained regions). Plot of the missense tolerance ratio (MTR) across PUM1 is presented above according to ExAC v2. Green denotes regions of neutrality (MTR ≥ 1) whereas red denotes regions of intolerance (MTR within bottom 25th percentile). The 3 amino-terminal repression domains (RD) of PUM1 (Enwerem et al.) are shown in orange. (C) The exonic positions of the two non-coding variants, c.363+1G>A and c.1158+1_1158+2dup, and their corresponding location on the PUM1 protein.

(large deletions, frameshift, stop-gain, etc.) however, there are several comparatively distinct

features among p.(Arg1147Trp) cases. While growth delay and short stature are documented but

infrequent features in PADDAS, it is consistent and significant feature in all p.(Arg1147Trp)

cases. The height of Subject 37 at age 15, for instance, was measured at 82 cm which falls 3

standard deviations below the population average. Comparatively, the frequency of short stature

is significantly enriched among p.(Arg1147Trp) cases relative to both other PADDAS cases and

the entire cohort (FET, P<0.005). Ocular ptosis and craniofacial dysmorphia including

anteverted nares, hair dysplasia and low-set ears, were also distinct features noted in all

p.(Arg1147Trp) cases (**Figure 39**).

Figure 39. Facial photographs (frontal) showing characteristic features of p.(Arg1147Trp) variant of PUM1.[cc]

8.2.4 Inherited loss-of-function alleles and intra-familial heterogeneity of PADDAS

We further identified 4 unique families segregating apparent LOF alleles across multiple affected

individuals exhibiting discordant PADDAS phenotypes (**Figure 40**). The proband (S19) of

Family A (**Figure 40A**), is a 4-year-old boy with infantile intellectual disability, mild gait

---

[cc] Features shown include bilateral ptosis (unilateral in Subject 72) (white arrows), hair dysplasia, anteverted nares and low-set ears relative to the eye-line. Photograph of S72 was adapted from Figure 1A of Ye et al.

disturbance, short stature, hypotonia, facial dysmorphia and ventricular septal defect who inherited the non-coding c.363+1G>A variant from his moderately affected mother (S20), in whom the variant occurred *de novo*. This c.363+1G>A occurs early (near N-terminal) in the peptide and is predicted to abolish the canonical splice donor of exon 2 (SpliceAI Δscore = 0.99, donor loss; SF2/ASF site lost). This generational pattern of a PADDAS case(s) inheriting a LOF variant from a comparatively less affected parent was observed in the three remaining families: Family B (chr1:31409179:31421689:DEL) (**Figure 40B**), Family C (p.(His1090Profs*16)) (**Figure 40C**), and Family D (chr1:31394262:31883885:DEL) (**Figure 40D**).
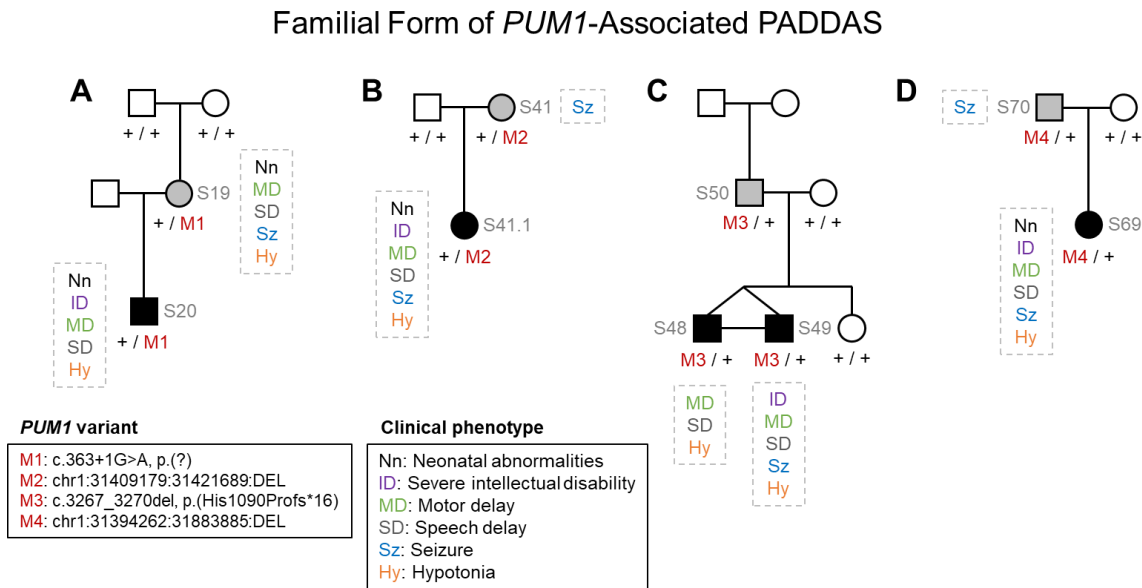
Figure 40. Familial pedigrees showing the segregation of variable Pumilio1-associated developmental disability, ataxia, and seizure (PADDAS) phenotypes and heterozygous loss of function alleles in four families.[dd]

[dd] (A) The affected boy (S19, age 4.5 years) and mother (S20, age 31 years) of Family A segregating the canonical splice site variant, c.363+1G>A (M1). (B) The affected girl (S41, age 6 years) and mother (S41.1, age 34 years) of Family B segregating a 12.51 kb intragenic deletion (M2). (C) The affected monozygotic boys (S48 and S49, age 6 years) and mildly affected father (S50, age 41 years) of Family C segregating the c.3267_3270del, p.(His1090Profs*16) frameshift variant (M3). (D) The affected infant girl (S69, age 1.5 years) and father (S70, age 24 years) of Family D segregating a 0.49 Mb deletion (M4). Allele status (M/+, +/+) is listed under individuals who were genetically screened; + denotes a wild-type allele.

8.2.5 Expanded genotypic profile of PRCA: additional missense and a pathogenic synonymous variant

We identified 6 novel cases (S34, S35, S45, S46, S59 and S62) exhibiting classical PRCA features. Consistent with the affected individuals (Family X) with p.(Thr1035Ser), the average onset of progressive cerebellar ataxia in these new cases began at a mean age of 45.5 years (range = 33-61 years) and magnetic resonance imaging (MRI) of the brain showed characteristic patterns of diffuse moderate cerebellar atrophy, with preservation of the cerebral and brainstem structures, in all cases. The *PUM1* variants were all missense substitutions apart from a synonymous variant, p.(Gly647=), identified in S62. The variant occurs in the middle of Exon 13 where the substitution (c.1941C>T, GGC to GGT) is predicted to activate a strong cryptic splice donor (SpliceAI Δscore = 0.91, donor gain, SpliceSiteFinder-like = 72.8, MaxEntScan = 6.7, NNSPLICE = 0.9, GeneSplicer = 4.6 and Human Splice Finder = 79.9) (**Figure 41A**). While the effect of this variant is predicted to induce skipping of exon 13 (**Figure 41B**, **Supplemental Table 13**) however alternative splicing at this activated upstream donor could also result the partial excision of Exon 13[ee] (**Figure 41B**). The comparable strengths of both donor sites could also result in a mixture of mRNA isoforms.

[ee] According to the sequencing data from this patient, 1 out of 103 reads (~1%) from this allele shows a 147-bp downstream excision consistent with all *in silico* predictions alternative splicing at the activated cryptic donor. A mini-gene experiment performed by Salvatore Botta in the lab shows ~200bp deletion at the beginning (5' end) of exon 13 which shifts the reading frame and results in a premature stop codon in small proportion of cases. This latter finding is conflicts the *in silico* predictions.
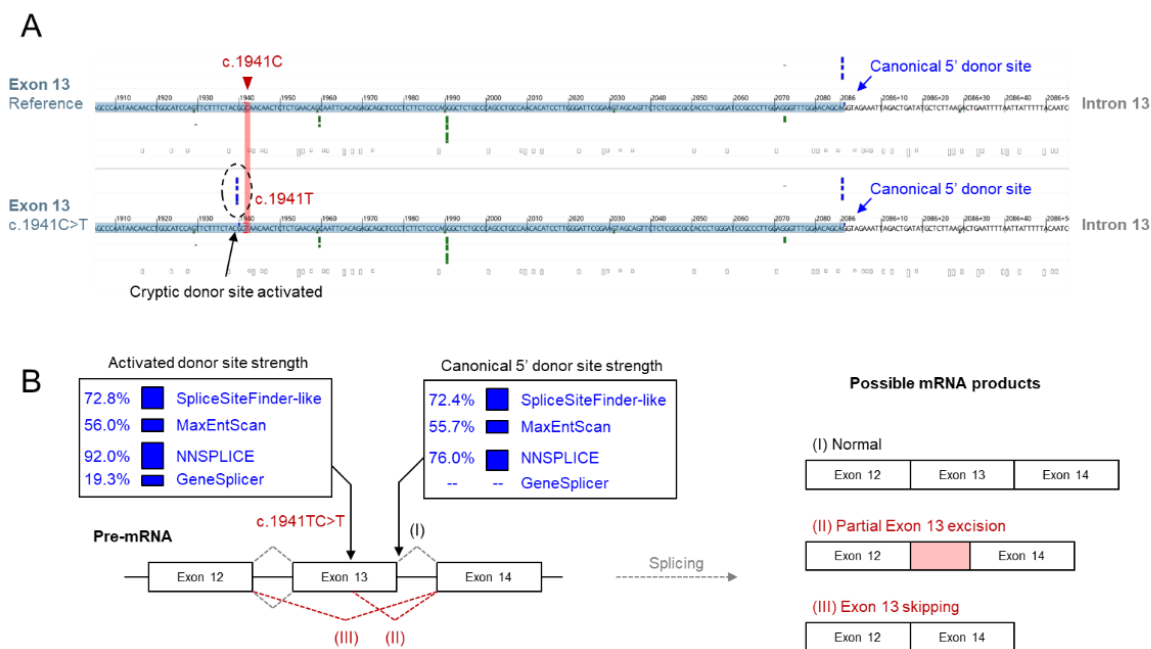
Figure 41. *Predicted effect and splicing outcomes of the synonymous (silent) variant, c.1941C>T, p.(Gly647=) in PUM1.[ff]*

As with p.(Thr1035Ser), all 6 of these variants exhibit incomplete penetrance as evidenced by

the presence of alleles in the general population (gnomAD) or familial segregation. The

p.(Gly647=), p.(His1050Tyr) and p.(Met746Val) alleles were inherited from unaffected parents,

although positive family history was noted for Subject 46.

[ff] According to the sequencing data from this patient, 1 out of 103 reads (~1%) from this allele shows a 147-bp downstream excision consistent with all in silico predictions alternative splicing at the activated cryptic donor. A mini-gene experiment performed by Salvatore Botta in the lab shows ~200bp deletion at the beginning (5' end) of exon 13 which shifts the reading frame and results in a premature stop codon in small proportion of cases. This latter finding is conflicts the *in silico* predictions. (A) Reference (top) and mutant (Subject 62) (bottom) splice window showing exon 13 (light blue) and intron 13. The position of c.1941C>T substitution is highlighted in red. Blue bars at the exon 13-intron 13 boundary indicate the predicted strength of the canonical 5' donor site; blue bars within exon 13 in the mutant splice window (dotted circle) indicate the strength of a cryptic donor site activated 2bp upstream of the c.1941C>T substitution (SpliceAI Δscore = 91%, donor gain at +2bp). (B) Schematic representation of WT and two predicted aberrant splicing outcomes due to the c.1941C>T variant. Insets indicate the comparable strengths of the canonical 5' donor site and activated cryptic donor sites according to the prediction algorithms SpliceSiteFinder-like, MaxEntScan, NNSPLICE and GeneSplicer. Possible mRNA products include (I) normal splicing of all exons, (II) partial 147 bp excision of exon 13 and (III) outright skipping exon 13 (EX-SKIP).

8.2.6 *De novo* hypomorphic variants cause an early-onset form of PRCA

Interestingly, several *de novo* missense variants (p.(Lys151Arg), p.(Gly95Arg) and

p.(Met955Ile)) were found in small group of patients with a distinctly early-onset form of PRCA

(S23, S61 and S74). Early cerebellar signs including gait abnormalities, dysarthria, tremors and

chorea were noted as early as infancy and early adolescence, after which ataxia and cerebellar

atrophy progressively manifests in early adulthood. It should be noted that the p.(Arg1139Trp)

variant, which has been described in two separate patients,[72,265] also exhibit progressive

cerebellar ataxia phenotypes albeit with discordant onset ages (due to differences in ethnicities)

and is thus considered PRCA-associated variant. Eye movement and gaze disorders, specifically

nystagmus and saccadic pursuits were found to be more prevalent in both the early- and adult-

onset forms of PRCA, collectively, (57%, 4/7) compared to all other patients in the cohort (25%,

4/16).

8.2.7 Spatial distribution of large deletions and intragenic *PUM1* variants

Large genomic deletions spanning multiple genes/loci, in addition to *PUM1*, made up the largest

class of variation; however, cases with deletions only comprised ~25% (16/62) of the cohort.

The size of these 16 large deletions ranged from ~9kb to ~5.6mb and collectively spanned a

region 5.7mb in size on chromosome 1 (1p35.3-1p35.1) that encompasses a total of 69 coding

genes. Among these 69 coding genes, 6 have known associations to disease-- *AK2* (MIM#

267500), *HPCA* (MIM# 224500), *LCK* (MIM# 615758), *EPB41* (MIM# 611804) and *MECR*

(MIM# 617282) however, disease associations are do not overlap with the known phenotype of

PUM1 (**Supplemental Figure 4**) nor were any of these reported disease features found in our

patients.. It was not possible to rule out additive effects from these genes and others

encompassed in the deletions, however PUM1 is the most intolerant to variation (RVIS score = -

1.02 (8.1%), **Supplemental Table 13**). There was no apparent association between clinical severity and size of deletions.

The majority of coding variants occurred outside of the RNA-binding domain specifically in clusters within the 3 repression domains (RD) domains[267] (**Figure 38, Figure 42A**). Missense variants within the RNA binding domain, (p.(Gln879Arg), p.(Met955Ile), p.(Thr1035Ser), p.(Glu1035Gln), p.(His1050Tyr) and p.(Arg1139Trp)), were found in patients with PRCA with the exception of p.(Asn973Asp) and p.(Arg1139Trp). The p.(Asn973Asp) variant is responsible for non-specific nucleotide binding with mRNA target through a van der Waals interactions over a distance of 3.4 Å[268] (**Figure 42C**). The effect of p.(Arg1147Trp) is unclear however the disproportionately high *de novo* recurrence rate at this position (7 independent alleles, P =) is highly indicative of pathogenicity.
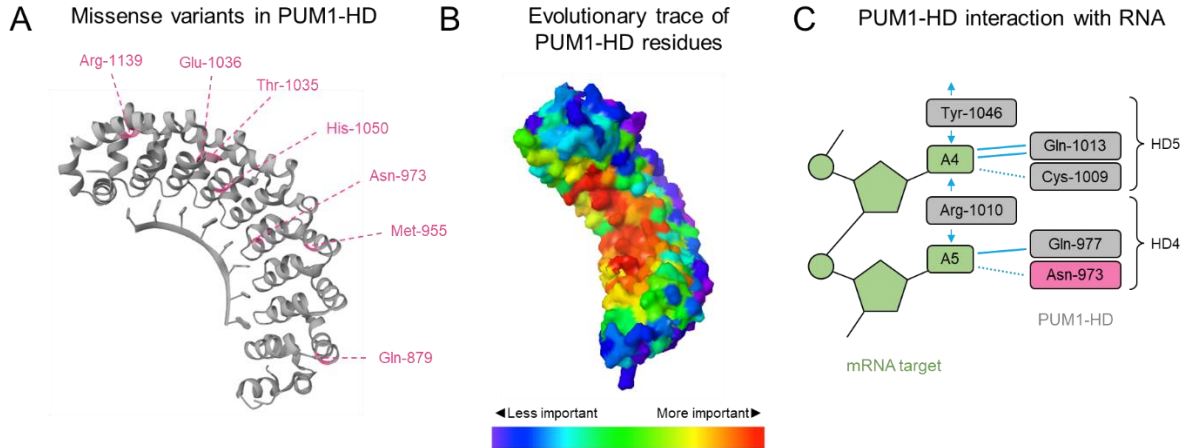
Figure 42. *Location of missense variants within the RNA-binding homology domain (HD) of PUM1.*[gg]

[gg] (A) residue positions of the seven missense variants within the crystal structure of PUM1-HD (3Q0P): p.(Arg1139Trp), p.(Glu1036Gln), p.(Thr1035Ser), p.(His1050Tyr), p.(Asn973Asp), p.(Met955Ile) and p.(Gln879Arg). (B) RNA-binding HD of PUM1 color coded by the relative rank of functional and structural "importance" among protein homologs sequence positions (Evolutionary Trace). (C) Schematic representation of interaction between residues (gray), including Asn-973 (pink), with a representative segment of an mRNA target (CycBreverse) in green. Stacking interactions are indicated by blue arrows, hydrogen bonds are indicated by solid blue lines and van der Waals interactions are indicated by dotted blue lines. This diagram is partially adapted from Gupta et al.

8.2.8 Discussion

Many of the results in this study are preliminary however several key findings have significant implications in the field of molecular diagnostics. Prior to this study, the genotypic profile of *PUM1* predominantly consisted of large genomic deletions. In this updated cohort, we now find that large deletions comprise ~25% of known pathogenic variation. The coding variants identified in the prior study also occurred within or in proximity to HD repeats suggesting possible disruption of RNA-binding function. While we did identify more variants within this domain, variants were also found in N-terminal low complexity regions. These N-terminal variants were not uniformly distributed but, interestingly, occurred in two large clusters in stretches of residues corresponding to possible domains recently identified to be involved in CNOT deadenylase-dependent repression.[267]

Missense variants identified in the HD repeats were largely associated with PRCA, including the substitution p.(Glu1036Gln) which occurs at the residue directly adjacent to the known PRCA variant, p.(Thr1035Ser). Although, none of these variants occur on the three residues of the tripartite recognition motif, the p.(Asn973Asp) has been shown to participate in non-specific binding with certain mRNA target molecules via van der Waals interactions. The individual with this variant (S24) was affected with a more severe developmental delay phenotype whereas those with missense variants situation far from the RNA-binding interface were found in individuals with milder, classical PRCA phenotypes. This genotype-phenotype correlation is consistent with the functional importance of RNA-interfacing residues relative to more distally situated residues in the HD repeat domains (**Figure 42B**).

Not all intragenic variants appear to be related to RNA-binding. One of the more interesting variants identified is the synonymous variant, p.(Gly647=). This variant was identified in a 69-year-old man and his brother who both affected by a recent-onset of ataxia. While the clinical phenotype and familial segregation of this variant is consistent with pathogenicity, the exact mechanism is not known. The role of synonymous variants in disease is becoming better understood with time. Mechanisms include phenomena such as codon usage or codon bias found in Alzheimer's disease. Synonymous variants can also alter critical nucleotide sequences which can affect splicing in pre-mRNA, particularly enhancer sequences near the boundaries of exons and introns. The p.(Gly647=) variant in this occurs in the middle of exon 13 at a relatively un-conserved nucleotide (phyloP100 = 2.662); however it does appear activate a cryptic splice donor site of equal (or greater) strength to the canonical 5' donor site. Splicing at this cryptic donor site would lead to a 143 bp excision (or truncation) of exon 13. The ultimate protein product (if any) would depend on several factors such as whether the effect shifts the reading frame possible resulting in an early terminal codon or a partially truncated but functional protein. Given the phenotype in the patients, the most likely outcome is a mixture of correctly spliced and mutant product which is corroborated by the sequencing data reported for this patient.

Perhaps the most interesting finding across the cohort is the incomplete penetrance and variable expressivity seen in a number of variants, particularly the LOF variants segregating in the 4 families. We cannot be certain about the variable expressivity of these variants without further functional analyses, but it is plausible that some may not result in a null allele. For instance, although the c.363+1G>A variant in S19 and S20 (**Figure 40A**) obliterates the canonical Exon 2 donor, an alternative splice donor exists 96 bp upstream of the exon junction (**Supplemental Figure 5**). The strength of this alternative splice donor (SpliceRover score = 99.1%) is

comparable to the canonical splice donor (SpliceRover score = 99.6%) and increases in strength

in the mutant (c.363+1G>A) sequence (SpliceRover score = 99.5%). It therefore plausible that

alternative splicing at this upstream donor rescues the splicing of (a moderately truncated) Exon

2 resulting in a functional (or partly function) PUM1 protein product. In another example, the

c.3267_3270del, p.(His1090Profs*16) variant is expected to result in a premature termination

codon (PTC) at the penultimate exon (exon 21 of 22) that is predicted to undergo nonsense

mediated decay (NMD), a quality control process by which a cell triages faulty transcripts

entering the cytoplasm.[269] However, its location in the protein may suggest that it may escape

NMD. According to the exon junction complex (EJC)-dependent model of NMD, any transcript

with a PTC 50-55 bp upstream of the final exon-exon junction is predicted to be NMD-sensitive

(NMD+).[270] On the other hand, transcripts within 50-55 bp of the final exon-exon junction or

anywhere in the last exon are predicted to be NMD-resistant (NMD-). The EJC-dependent

model is well-supported by a preponderance of experimental evidence, and indeed the 50 bp rule

accurately predicts NMD sensitivity in approximately 85% of cancer-related variants[271-274]

however a number of exceptions have been reported.[275] Specifically, NMD-resistant range is

extended to PTC across the entire penultimate exon which would certainly include the case of

c.3267_3270del, p.(His1090Profs*16) since it occurs within several nucleotides of this NMD-

resistant window (**Figure 43**). Escaping NMD could result in a mutant protein product with

potent gain-of-function activity although further studies are necessary to validate this.
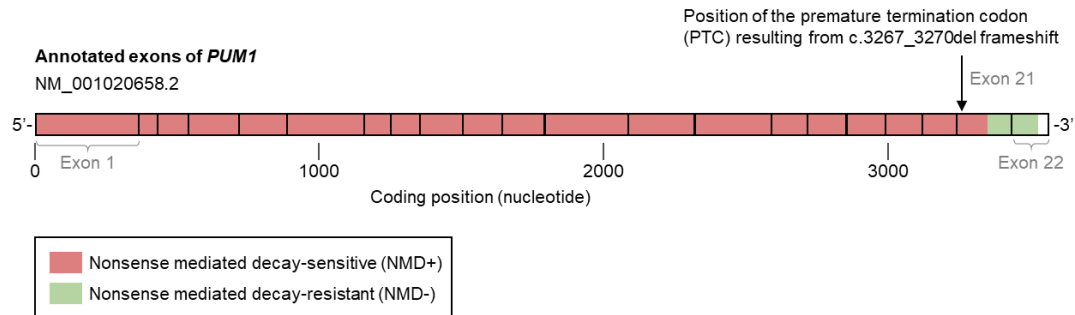
Figure 43. *Position of the premature truncation codon introduced by the c.3267_3270del, p.(His1090Profs*16) variant relative to the nonsense mediated decay-sensitive (NMD+) and -resistant (NMD-) portions of the post-processed PUM1 mRNA transcript (NM_001020658.2).[hh]*

This study has several limitations worth consideration. The pathogenicity of each variant needs to be confirmed by identifying more cases or functional studies (ideally both). While each of these variants were identified as the leading candidate gene, the accuracy of this is sensitive to the type of genetic analysis used in each case. There are also several of the variants in this study exhibit features that are inconsistent with pathogenicity. For instance, 3 missense variants are over-represented in gnomAD, p.(Lys151Arg) (8 heterozygotes), p.(Ala52Gly) (6 heterozygotes) and p.(Ile417Val) (6 heterozygotes). Given the mild phenotype and late-onset of symptoms, it is conceivable that the individuals in gnomAD may be erroneously classified as "healthy" which is a known caveat of such databases.[276] Given the additional penetrance in this disease, extra caution should be taken when considering the allele frequency of PUM1 variants at the population level.

In summary, we have established the largest repository of *PUM1* patients to date which consists of 62 individuals (from 54 families) from around the world. The genetic spectrum is quite large

---

[hh] This variant is predicted to undergo NMD according to its position outside of NMD- region. Given its position in the penultimate exon (exon 21) and relative proximity to the 50-55 bp from the final exon-exon junction, there is a possibility of NMD escape which explain the phenotype variability of patients harboring this variant in the study.

and associated with a higher-than-expected degree of heritability. While most variants occur *de novo*, many of these variants exhibit either incomplete penetrance or variable expressivity across multiple affected individuals within a family. Remarkably, some of the familial incompletely penetrant variants include expected LOF variants such as large deletions, canonical splice site and frameshift variants. In addition to PADDAS and PRCA, we also identified several other clinical entities, most notably an early onset form of PRCA. Additional analyses are needed however they results significantly expand both the genotypic and clinical profile of *PUM*1-associated disease.

8.3 Materials & Methods

8.3.1 Human Subjects

All study procedures were defined under protocol #AAAR7750 approved by the Institutional Review Board at Columbia University Irving Medical Center. Written informed consent was obtained from all subjects prior to study enrollment. The study adhered to tenets set out in the Declaration of Helsinki. Each patient underwent a full clinical examination by a neurologist and/or medical geneticist. Clinical data were directly abstracted from medical records provided by the referring clinician(s) which included behavioral assessment and electroencephalograms (EEG). Developmental delay and cognitive ability tests were performed in subjects using the following tests: Gesell Developmental Schedules, and Bayley Scales of Infant Development II. Magnetic resonance imaging records were obtained electronically (image file, DICOM) for most subjects and independently reviewed by investigators at CUIMC and referring co-investigators.

8.3.2 Sequencing and genetic analysis

All subjects underwent various diagnostic screening procedures including standard-of-care tests for neurodevelopmental diseases such as karyotyping, microarray or panel sequencing, as well as

targeted candidate gene testing: CGG repeat expansion of FMR1, direct sequencing of MECP2, among others. All identified variants were analyzed in accordance with ACMG Guidelines for variant interpretation and classification.[85] Minor allele frequencies (MAF) of all *PUM1* variants were obtained from The Genome Aggregation Database (gnomAD).   Functional annotation of variants was carried out with ANNOtate VARiation (ANNOVAR) using pathogenicity scores of M-CAP, REVEL, Eigen and CADD. As a general guideline, pathogenicity is predicted for variants with scores over 0.025 for M-CAP, 0.5 for REVEL, 0.5 for Eigen and 25 for CADD. The pathogenicity of large CNV assessed using the Structural Variant Classifier Trained on Variants Rare and Exonic (STRVCTVRE) algorithm.[277] Variants with STRVCTVRE scores ≥ 0.37 are predicted to be pathogenic. Structural analysis of the PUMILIO-homology domain was performed using the crystal structure resolved by Liu et al (PDB: 3Q0P).[278] Residues on this structure were color coded according to Evolutionary Trace (ET) scores.[279] The predicted effect on splicing was assessed using SpliceSiteFinder-like, MaxEntScan, NNSPLICE, GeneSplicer,[280] and ESEfinder[281] provided by the Interactive Biosoftware Alamut® software, SpliceAI and SpliceRover. Using SpliceRover, splice donors near the 3' end of Exon2 were identified in the following partial sequence:

GTGTTGGTGGAATGAGCGTTGCATGTGTCTTGAAGAGAAAAGCAGTGCTTTGGCAG
GACTCTTTCAGCCCCCACCTGAAACATCACCCTCAAGAACCAGCTAATCCCAACATG
CCTGTTGTTTTGACATCTGGAACAGGGTCGCAAGCGCAGCCACAACCAGCTGCAAAT
CAGGCTCTTGCAGCTGGGACTCACTCCAGCCCTGTCCCAGGATCTATAGGAGTTGCA
GGCCGTTCCCAGGACGACGCTATGGTGGACTACTTCTTTCAGAGGCAGCATGGTGAG
CAGCTTGGGGGAGGAGGAAGTGGAGGAGGCGGCTATAATAATAGCAAACATCGAT
GGCCTACTGGGGATAACATTCATGCAGAACATCAGGTAAAAAAAAGTTCACTATTTT

AGCTTTGACATTCCTTTATAATCACTGTCCACTTAGGATATCTGCTTTCAATGAATGT

CTTTAAAAACTTAAATTGCTGATTTCATTTGGTACAAACAGTTACTTGTTATAAAGGT

TGATCATTGTAATACTGTTACTGGGTCATTTATCATTCTGGTGCTAAATAGTAAATAG

CCAATTCATTCATTCATTCATTCATTTGAGACAGAGTCTTGCTCTGTCACCCAGGCTG

GAGTGTAGTCGCCTGATCTCGGCTTGATCTCGGCTCACTGCAGCTGCTGCCTCCCAG

GTTCAAGCAGTTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGACTACAGGTGTGCACC

ACCACGCCTGGCTGATTATGTATTTTCAGTAGAAACAGGGTTTCACCATGTTGTCCA

GGCTGGCCTCGAACTCCTGACCTCAAGTGATCTGCCCGCTTCGGCCTCCCAAAGTGC

TGGGATTACAGTTGTGAGCCACCCACGCCCGGCCTATAGAATGCATTTTTTTTTTTT

TTAAGACGGAGTCTCACTCTTTTGCCCAGGCTGGAGTGCACTGGCGCGATCTCTGCT

CACTGCAAGCTCCGCCTCCCAGGTTCATGCCATTCTCCTGCCTCAGCCTCCCGAGTA

GCTGGGACTACAGGTGCCCGCACCATGCCTGGCTAATTTTTTGTATTTTTTTGGTAG

AGACGGGGTTTCACTGTGTTAGCCAGGATGGTCTCTATCTCCTGACCTCGTGATCCA

CCCATCTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACTGCGCCCGGCCT

ACAACGCATTTTTAAGAACACTTTTTTTGGAACAAAGGATTCTTTTTTTACTGTTGA

CATAAGAACTGCTAGGGTTGTTTATAGGATAGATACATATTTTTGAATCTTTTCTCTT

GGGTGTGGGTGACTTGATTTTAAGATGTTGTAGTTGAAAGTAGGGGGAAGTTGATAA

CATCAGCTGCTCAAGTGAAATTTATTTGAAATGGGACTGATATATTGTGTCAGAGAG

AAGTGGTTGTGGCTTTTAGGTTTTAACCACCTTTGTCCTCTTGTTATTGTTTAATTTT

AATTTTAACATGTTCAGTCTTATTAATCTAAAACCATAGTGGGAATCATAGCATATTT

TACTGAGAAGAGAGGTTAGCCTTATCCATGTTTCCTGATTTCTTGATAATGACCAAC

CAAAAGATTAGACTGTTTTGATCAGGGTATTCTTTGAAATTGGGTATAGGTTGAATT

TGGAAAGGTTCATTTTGTGTCACCCGTCAGTACTGATTAGCGAGAACTGACATTTTT

CATTAGGCTTCTAGTGCTTCTGTTTTACTTCAGAAAAAATCAGGTACAAACATACCT

GCAAGTCTGATTTTATTATACTTCCTCTGATCCCCCCAAAAGTGAACCATAGTTGTTT

TTTTTTTTTTTTTTTTTTTTGAGATGGAGTCTC

The probability of exon skipping by the c.1941C>T, p.(Gly647=) variant was assessed by EX-SKIP.[98] Partial *PUM1* sequences compared using EX-SKIP were: WT (reference):

TAACAACCTGGCATCCAGTTCTTTCTACGG<u>C</u>AACAACTCTCTGAACAGCAATTCACA

GAGC and MUT (c.1941C>T):

TAACAACCTGGCATCCAGTTCTTTCTACGG<u>T</u>AACAACTCTCTGAACAGCAATTCACA

GAGC.

Chapter 9 Conclusions and Future Directions

Short-term impact: prognostication and treatment prospects

Genotype-phenotype correlation studies have broad applications in the clinic as a more precise tool for molecular diagnoses. The disease spectrum of ABCA4 disease is so wide that a molecular diagnosis offers very little relevant information regarding long-term symptom development. While the severity of ABCA4 disease does track with the class of variation found in patients, for instance, genotypes with LOF alleles give rise to the more severe disease phenotypes, the overwhelming majority of genotype are difficult to assess. Most *ABCA4* variants are missense alleles, which include all the mild variants (p.(Gly1961Glu) and hypomorphs) and many severe alleles. In fact, our results in ***Chapter 5*** found that many of these "severe" missense variant are indistinguishable from PVS1 (LOF) alleles in their long-term effect on phenotype. Not all missense alleles, especially those that are rare, have been accounted for, which is a limitation of this study. Nevertheless, the advantage of such studies is that they can be scaled up and built upon, as data from more patients are acquired to increase the rigor of predictions.
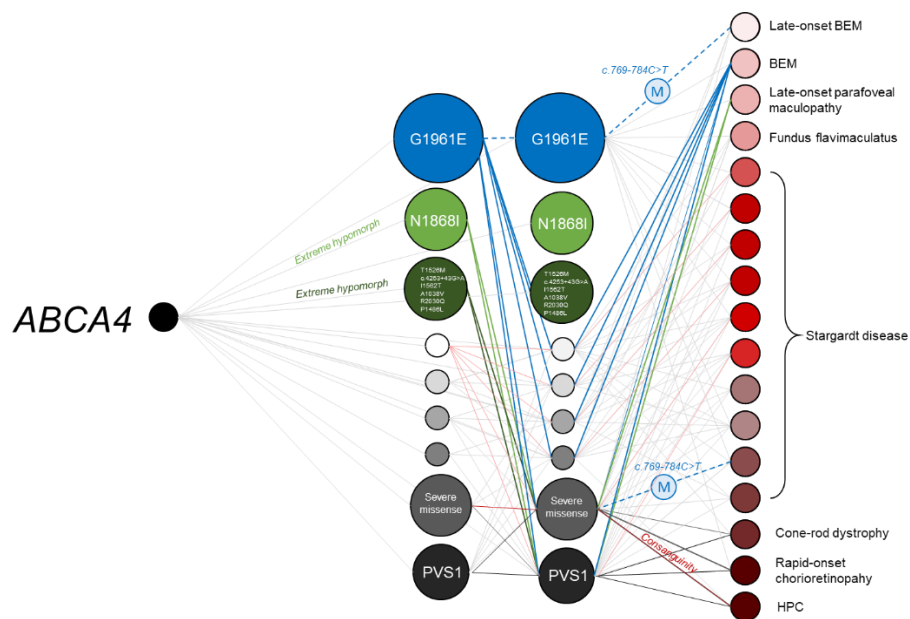
Figure 44. *Monogenic disease architecture of ABCA4*

ABCA4 disease has been an attractive target for therapeutic intervention due to it being the most

common form of retinal degeneration and the suitability of the eye for the implementation of

novel therapies.[ii] Nevertheless, approved therapy nor cure exists for ABCA4 disease. The are

several active or emerging clinical trials of therapeutic compounds designed to slow or arrest

disease progression in patients, specifically ALK-001 (NCT02402660) and LBS-008

(NCT05266014) which counteract the toxicity and/or accumulation of bisretinoids.[282,283] Neither

these compounds are not intended to reverse vision loss as photoreceptor (a type of neuron)

degeneration is an irreversible process[jj]. As such, both trials are required by the FDA to

demonstrate efficacy by comparing the progression rates of atrophic lesion growth in the retina

between treated and placebo groups. Due to relatively slow natural rate of progression of these

---

[ii] As an organ, the eye is well-suited for novel therapies because it is easily accessible, compartmentalized, and immune-privileged.

[jj] Reversing vision loss due to the degeneration of photoreceptors and/or RPE in the retina requires stem cell replacement therapy. Both Bone marrow-derived stem cell (NCT01920867) and human embryonic stem cell (hESC)-derived RPE cell (NCT02903576) have been attempted in recent years neither of which advanced beyond interim safety assessment.

lesions (~0.6-2.0 mm^2/year),[121,201,204,218] inclusion criteria favor "fast progressors" who are often younger patients in order to convincingly show efficacy within the trial timeframe of 1-3 years. Despite the heterogeneous spectrum of ABCA4 disease, none of these trials regard genotype as considerable factor. The genotype-phenotype framework we developed in *Chapter 5* allowed us to convincingly show (in *Chapter 6*) significant differences in growth rates between major genotypes which could skew results if treatment groups are not properly randomized. We also show that the current inclusion criteria will firmly select for study participants with mild genotypes which is not ideal due to the slow-progressing nature of these phenotypes and the possibility that these cases may not benefit from further reduction of lipofuscin accumulation.[22,284] The results in these two chapters therefore warrant urgent revisions to the design of several ongoing clinical trials for ABCA4 disease.

Similarly, our goal in *Chapter 7* was partially motivated by the fact that there is a known CERT inhibitor that could be therapeutically repurposed; however, this is contingent on several key pieces of theoretical knowledge which are (1) whether or not *CERT1* is indeed a monogenic disease and (2) if the disease mechanism involves an aberrantly constitutive activation of CERT function. We were able to answer this question simply by identifying the type of *CERT1* variation pathogenic in patients and their precise location within the CERT protein. We performed an unbiased and geographically unlimited search of patients all over the world. Across two years, we identified 31 patients in whom we were able to obtain comprehensive clinical information and confirm the causality of the underlying *CERT1* variant. Unlike other disorders such as ABCA4 disease, the variant profile consisted only of heterozygous missense variants. The absence of LOF variants in patients is an indicator of tolerance to loss of function,[260] which is also corroborated by the relatively high proportion of LOF variants in the

133

general population (LOEUF = 0.36[kk]).[285] The distribution of missense variants was also

revealing. Pathogenic variants often cluster in functional domains of genes in which LOF is the

primary genetic etiology.[260] In *ABCA4* for instance, ~70% of missense variants occur in either

the exocytomasmic domains (ECD) or nucleotide binding domains (NBD) (NBD1 = 17%, NBD2

= 13%, ECD1 = 26%, ECD2 = 13%).[286] In *CERT1*, 27/31 (87%) of the missense variants

occurred in several clusters within the unstructured intervening region of the protein rather than

in the important functional PH and START domains (**Figure 37**). The largest of these variant

clusters occurs specifically within the serine rich region (SRR), the regulatory domain where

CERT is deactivated by multisite phosphorylation of the serine residues contained within. We

found that variants in the SRR collectively underlie the most severe phenotypes across all

patients but, specifically, the recurring variant p.(Ser132Leu) is the most severe relative to other

SRR variants. The sequential phosphorylation of serine residues in the SRR by Casein kinase

$1\gamma2$ (CK1γ2)[287] requires the initial phosphorylation (or priming) of Ser-132 protein kinase D

(PKD).[64,288,289] Using these results, we teamed up with collaborators and conceived additional

experiments that led to a large study in which we resolved structurally this intervening region,

verified the pathogenicity (via disruption of CERT and sphingolipid homeostasis) of all missense

variants found in patients, generated a drosophila model that recapitulates the disease phenotype

and demonstrated pre-clinical efficacy of HPA-12 as a potential treatment strategy for CerTra

syndrome.

[kk]The LOEUF (Loss-Of-Function Expected Upper bound Fraction) score in gnomAD is a metric that estimates the proportion of loss-of-function (LoF) variants expected to be tolerated in a given gene. The LOEUF score ranges from 0 to 1, with higher scores indicating a greater expected tolerance to LoF variants. There is no absolute thresholds for LOEUF however studies have recommended using LOEUF <0.35 as a filter for genes intolerant to LOF.[Cummings et al.] Although the LOEUF for CERT1 is borderline, it is higher than genes like PUM1 (LOEUF = 0.13, pLI = 1) in which LOF intolerance has been validated by low Mendelian ratio of *Pum1*$^{-/-}$ mice.[Gennarino et al.]

Long-term impact: genetic penetrance and modifiers

Beyond disease relevance, our results in **Chapters 3** (*ABCA4*) and 8 (*PUM1*) highlight the prevalence of incomplete penetrance and variable expressivity in two very different Mendelian diseases. While the phenotypic discordance in *ABCA4* is not as etiologically distinct as in *PUM1*, the effect of the c.769-784C>T variant is still quite pronounced between individuals with and without this additional modifier variant. In Chapter 5, the effect of c.769-784C>T on the p.(Gly1961Glu) allele can be the deciding factor between rather favorable Prognosis 1 and quite grim Prognosis 3. The mild phenotype of p.(Gly1961Glu) in patients has been known for some time. The precise disease mechanism caused by this variant is still mostly unknown; however, collaborative studies quantifying fundus autofluorescence[103,219] from our department have shown that lipofuscin levels among p.(Gly1961Glu) patients are only marginally elevated (if at all), which suggests that treatment approaches aimed at reducing lipofuscin accumulation would be efficacious in such patients.[55] However, we now know that at least some specific p.(Gly1961Glu) genotypes; i.e. those modified by c.769-784C>T variant, could benefit from this type of treatment.

Our findings in the expanded PUM1 disease cohort in *Chapter 8* in some ways mirror the penetrance observations of *ABCA4* alleles (**Figure 44**) (**Figure 45**); however, understanding the underlying etiology in PUM1-associated disease is likely outside the scope of sequencing studies. Although ABCA4 has an important constituent role in the visual cycle, its function of transporting all-*trans*-retinal out of the outer segment disc lumen is largely independent and non-redundant (**Figure 2**). This explains why significant disease modifiers identified thus far have been intragenic (*cis*).[46,106] We recently identified common haplotype in PRPH2, a protein responsible for the structural integrity of outer segment discs,[290,291] as a potential *trans*-modifier

for ABCA4 disease. While PRPH2/ROM1 complex and ABCA4 share the same subcellular

localization and could, therefore, physically interact, the functional relationship in this case

remains unknown.[58] PUM1, on the other hand, functions within a complex interactome

containing multiple binding partners including its paralog PUM2.[292] These interactions along

with the large group mRNA targets regulated by PUM1 will require a multifaceted approach

involving both *in vitro* and *in vivo* studies currently being undertaken in the lab. Nevertheless,

our study identified several novel lines of inquiry. The precise function and structure of the N-

terminal (low-complexity) region of PUM1 has been largely unknown compared to the well-

characterized RNA-binding HD repeat domains. Our analysis identified clusters of putatively

pathogenic variants in this part of the protein that spatially correlates with a binding region

involved with CCR4-NOT-dependent repressive activity (**Figure 39**).[267]
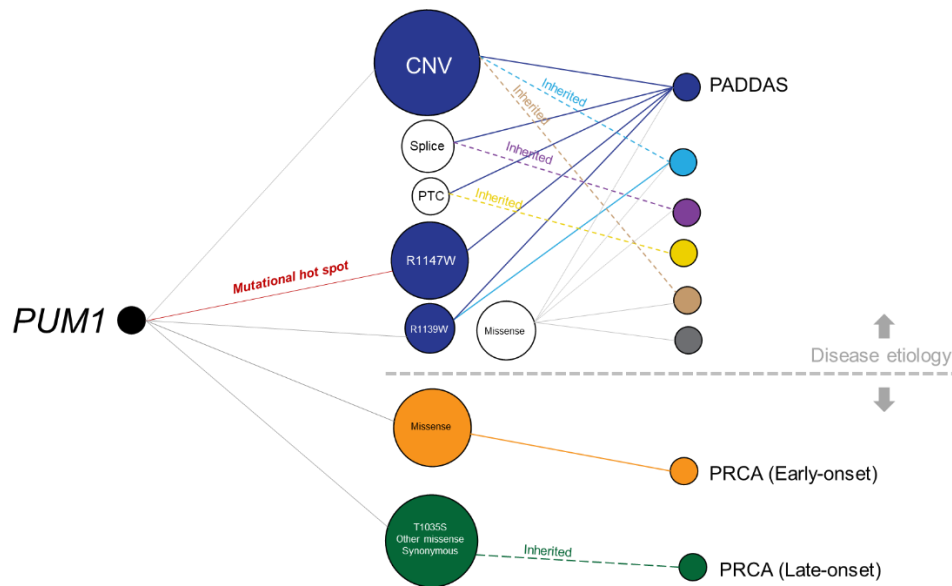
Figure 45. Monogenic disease architecture of PUM1

A new direction for genomic medicine

Like most other accounts of scientific research, the advancement of knowledge contributed by this dissertation is incremental across three very different and specific areas of medicine. Nevertheless, the broader impact of this dissertation points to the need for a consequential shift in genomic medicine. The practice of assigning of pathogenicity to variants identified by diagnostic sequencing should account for phenomena such as penetrance and known genetic modifiers. Gene-specific molecular diagnoses should also consider phenotypic differences at the variant level provide more informative results to clinicians and patients. While this new direction will indeed be challenging, it is necessary struggle if we wish to truly deliver on the promise of precision medicine.

References

1.   Allikmets, R.*, et al.* A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. *Nat Genet* **15**, 236-246 (1997).
2.   Cremers, F.P.M., Lee, W., Collin, R.W.J. & Allikmets, R. Clinical spectrum, genetic complexity and therapeutic approaches for retinal disease caused by ABCA4 mutations. *Prog Retin Eye Res*, 100861 (2020).
3.   Shuster, T.A., Nagy, A.K. & Farber, D.B. Nucleotide binding to the rod outer segment rim protein. *Exp Eye Res* **46**, 647-655 (1988).
4.   Sun, H. & Nathans, J. Stargardt's ABCR is localized to the disc membrane of retinal rod outer segments. *Nat Genet* **17**, 15-16 (1997).
5.   Molday, L.L., Rabin, A.R. & Molday, R.S. ABCR expression in foveal cone photoreceptors and its role in Stargardt macular dystrophy. *Nat Genet* **25**, 257-258 (2000).
6.   Papermaster, D.S., Reilly, P. & Schneider, B.G. Cone lamellae and red and green rod outer segment disks contain a large intrinsic membrane protein on their margins: an ultrastructural immunocytochemical study of frog retinas. *Vision Res* **22**, 1417-1428 (1982).
7.   Quazi, F., Lenevich, S. & Molday, R.S. ABCA4 is an N-retinylidene-phosphatidylethanolamine and phosphatidylethanolamine importer. *Nat Commun* **3**, 925 (2012).
8.   Quazi, F. & Molday, R.S. Differential phospholipid substrates and directional transport by ATP-binding cassette proteins ABCA1, ABCA7, and ABCA4 and disease-causing mutants. *J Biol Chem* **288**, 34414-34426 (2013).
9.   Fong, S.L.*, et al.* Characterization, localization, and biosynthesis of an interstitial retinol-binding glycoprotein in the human eye. *J Neurochem* **42**, 1667-1676 (1984).
10.   Xue, L., Gollapalli, D.R., Maiti, P., Jahng, W.J. & Rando, R.R. A palmitoylation switch mechanism in the regulation of the visual cycle. *Cell* **117**, 761-771 (2004).
11.   Jin, M., Li, S., Moghrabi, W.N., Sun, H. & Travis, G.H. Rpe65 is the retinoid isomerase in bovine retinal pigment epithelium. *Cell* **122**, 449-459 (2005).
12.   Moiseyev, G., Chen, Y., Takahashi, Y., Wu, B.X. & Ma, J.X. RPE65 is the isomerohydrolase in the retinoid visual cycle. *Proc Natl Acad Sci U S A* **102**, 12413-12418 (2005).
13.   Redmond, T.M.*, et al.* Mutation of key residues of RPE65 abolishes its enzymatic role as isomerohydrolase in the visual cycle. *Proc Natl Acad Sci U S A* **102**, 13658-13663 (2005).
14.   Yamamoto, H.*, et al.* Mutations in the gene encoding 11-cis retinol dehydrogenase cause delayed dark adaptation and fundus albipunctatus. *Nat Genet* **22**, 188-191 (1999).
15.   Sparrow, J.R.*, et al.* The bisretinoids of retinal pigment epithelium. *Prog Retin Eye Res* **31**, 121-135 (2012).
16.   Cideciyan, A.V.*, et al.* Mutations in ABCA4 result in accumulation of lipofuscin before slowing of the retinoid cycle: a reappraisal of the human disease sequence. *Hum Mol Genet* **13**, 525-534 (2004).
17.   Weng, J.*, et al.* Insights into the function of Rim protein in photoreceptors and etiology of Stargardt's disease from the phenotype in abcr knockout mice. *Cell* **98**, 13-23 (1999).

18. Sparrow, J.R. & Boulton, M. RPE lipofuscin and its role in retinal pathobiology. *Exp Eye Res* **80**, 595-606 (2005).
19. Sparrow, J.R., Yoon, K.D., Wu, Y. & Yamamoto, K. Interpretations of fundus autofluorescence from studies of the bisretinoids of the retina. *Invest Ophthalmol Vis Sci* **51**, 4351-4357 (2010).
20. Bosch, E., Horwitz, J. & Bok, D. Phagocytosis of outer segments by retinal pigment epithelium: phagosome-lysosome interaction. *J Histochem Cytochem* **41**, 253-263 (1993).
21. Young, R.W. The renewal of photoreceptor cell outer segments. *J Cell Biol* **33**, 61-72 (1967).
22. Burke, T.R.*, et al.* Quantitative fundus autofluorescence in recessive Stargardt disease. *Invest Ophthalmol Vis Sci* **55**, 2841-2852 (2014).
23. Duncker, T.*, et al.* Quantitative fundus autofluorescence distinguishes ABCA4-associated and non-ABCA4-associated bull's-eye maculopathy. *Ophthalmology* **122**, 345-355 (2015).
24. Paavo, M.*, et al.* Mutations in GPR143/OA1 and ABCA4 Inform Interpretations of Short-Wavelength and Near-Infrared Fundus Autofluorescence. *Invest Ophthalmol Vis Sci* **59**, 2459-2469 (2018).
25. Sparrow, J.R.*, et al.* Quantitative fundus autofluorescence in mice: correlation with HPLC quantitation of RPE lipofuscin and measurement of retina outer nuclear layer thickness. *Invest Ophthalmol Vis Sci* **54**, 2812-2820 (2013).
26. Charbel Issa, P.*, et al.* Fundus autofluorescence in the Abca4(-/-) mouse model of Stargardt disease--correlation with accumulation of A2E, retinal function, and histology. *Invest Ophthalmol Vis Sci* **54**, 5602-5612 (2013).
27. Makelainen, S.*, et al.* An ABCA4 loss-of-function mutation causes a canine form of Stargardt disease. *PLoS Genet* **15**, e1007873 (2019).
28. Zangerl, B., Lindauer, S.J., Acland, G.M. & Aguirre, G.D. Identification of genetic variation and haplotype structure of the canine ABCA4 gene for retinal disease association studies. *Mol Genet Genomics* **284**, 243-250 (2010).
29. Ekesten, B., Makelainen, S., Ellis, S., Kjellstrom, U. & Bergstrom, T.F. Abnormal Appearance of the Area Centralis in Labrador Retrievers With an ABCA4 Loss-of-function Mutation. *Transl Vis Sci Technol* **11**, 36 (2022).
30. Cornelis, S.S.*, et al.* In Silico Functional Meta-Analysis of 5,962 ABCA4 Variants in 3,928 Retinal Dystrophy Cases. *Hum Mutat* **38**, 400-408 (2017).
31. Bauwens, M.*, et al.* An augmented ABCA4 screen targeting noncoding regions reveals a deep intronic founder variant in Belgian Stargardt patients. *Hum Mutat* **36**, 39-42 (2015).
32. Sangermano, R.*, et al.* Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. *Genet Med* **21**, 1751-1760 (2019).
33. Sangermano, R.*, et al.* ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Res* **28**, 100-110 (2018).
34. Zernant, J.*, et al.* Analysis of the ABCA4 genomic locus in Stargardt disease. *Hum Mol Genet* **23**, 6797-6806 (2014).
35. Khan, M.*, et al.* Resolving the dark matter of ABCA4 for 1054 Stargardt disease probands through integrated genomics and transcriptomics. *Genet Med* **22**, 1235-1246 (2020).

36. Miller, D.E.*, et al.* Targeted long-read sequencing identifies missing disease-causing variation. *Am J Hum Genet* **108**, 1436-1449 (2021).

37. Lee, W.*, et al.* Complex inheritance of ABCA4 disease: four mutations in a family with multiple macular phenotypes. *Hum Genet* **135**, 9-19 (2016).

38. Stargardt, K. Über familiäre, progressive degeneration in der maculagegend des auges. *Albrecht von Graefes Arch Klin Ophthalmology* **71**, 534–550 (1909).

39. Franceschetti, A. A special form of tapetoretinal degeneration: fundus flavimaculatus. *Trans Am Acad Ophthalmol Otolaryngol* **69**, 1048-1053 (1965).

40. Franceschetti, A. & François, J. [Fundus flavimaculatus]. *Arch Ophtalmol Rev Gen Ophtalmol* **25**, 505-530 (1965).

41. Leng, T.*, et al.* Foveal cavitation as an optical coherence tomography finding in central cone dysfunction. *Retina* **32**, 1411-1419 (2012).

42. Noupuu, K., Lee, W., Zernant, J., Tsang, S.H. & Allikmets, R. Structural and genetic assessment of the ABCA4-associated optical gap phenotype. *Invest Ophthalmol Vis Sci* **55**, 7217-7226 (2014).

43. Bertelsen, M.*, et al.* Generalized choriocapillaris dystrophy, a distinct phenotype in the spectrum of ABCA4-associated retinopathies. *Invest Ophthalmol Vis Sci* **55**, 2766-2776 (2014).

44. Lee, W., Zernant, J., Nagasaki, T., Tsang, S.H. & Allikmets, R. Deep Scleral Exposure: A Degenerative Outcome of End-Stage Stargardt Disease. *Am J Ophthalmol* **195**, 16-25 (2018).

45. Lee, W.*, et al.* The external limiting membrane in early-onset Stargardt disease. *Invest Ophthalmol Vis Sci* **55**, 6139-6149 (2014).

46. Zernant, J.*, et al.* Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration. *J Med Genet* **54**, 404-412 (2017).

47. Zernant, J.*, et al.* Extremely hypomorphic and severe deep intronic variants in the ABCA4 locus result in varying Stargardt disease phenotypes. *Cold Spring Harb Mol Case Stud* **4**(2018).

48. Lee, W.*, et al.* Longitudinal Analysis of a Resolving Foveomacular Vitelliform Lesion in ABCA4 Disease. *Ophthalmol Retina* **6**, 847-860 (2022).

49. Bauwens, M.*, et al.* ABCA4-associated disease as a model for missing heritability in autosomal recessive disorders: novel noncoding splice, cis-regulatory, structural, and recurrent hypomorphic variants. *Genet Med* **21**, 1761-1771 (2019).

50. Runhart, E.H.*, et al.* The Common ABCA4 Variant p.Asn1868Ile Shows Nonpenetrance and Variable Expression of Stargardt Disease When Present in trans With Severe Variants. *Invest Ophthalmol Vis Sci* **59**, 3220-3231 (2018).

51. Aguirre-Lamban, J.*, et al.* Further associations between mutations and polymorphisms in the ABCA4 gene: clinical implication of allelic variants and their role as protector/risk factors. *Invest Ophthalmol Vis Sci* **52**, 6206-6212 (2011).

52. Maugeri, A.*, et al.* The ABCA4 2588G>C Stargardt mutation: single origin and increasing frequency from South-West to North-East Europe. *Eur J Hum Genet* **10**, 197-203 (2002).

53. Schulz, H.L.*, et al.* Mutation Spectrum of the ABCA4 Gene in 335 Stargardt Disease Patients From a Multicenter German Cohort-Impact of Selected Deep Intronic Variants and Common SNPs. *Invest Ophthalmol Vis Sci* **58**, 394-403 (2017).

54.   Webster, A.R.*, et al.* An analysis of allelic variation in the ABCA4 gene. *Invest Ophthalmol Vis Sci* **42**, 1179-1189 (2001).

55.   Burke, T.R.*, et al.* Retinal phenotypes in patients homozygous for the G1961E mutation in the ABCA4 gene. *Invest Ophthalmol Vis Sci* **53**, 4458-4467 (2012).

56.   Lee, W.*, et al.* Genotypic spectrum and phenotype correlations of ABCA4-associated disease in patients of south Asian descent. *Eur J Hum Genet* **25**, 735-743 (2017).

57.   Guymer, R.H.*, et al.* Variation of codons 1961 and 2177 of the Stargardt disease gene is not associated with age-related macular degeneration. *Arch Ophthalmol* **119**, 745-751 (2001).

58.   Zernant, J.*, et al.* Rare and common variants in ROM1 and PRPH2 genes trans-modify Stargardt/ABCA4 disease. *PLoS Genet* **18**, e1010129 (2022).

59.   Lee, W.*, et al.* Modification of the PROM1 disease phenotype by a mutation in ABCA4. *Ophthalmic Genet* **40**, 369-375 (2019).

60.   Lee, W.*, et al.* Simultaneous Expression of ABCA4 and GPR143 Mutations: A Complex Phenotypic Manifestation. *Invest Ophthalmol Vis Sci* **57**, 3409-3415 (2016).

61.   Hanada, K.*, et al.* Molecular machinery for non-vesicular trafficking of ceramide. *Nature* **426**, 803-809 (2003).

62.   Raya, A., Revert, F., Navarro, S. & Saus, J. Characterization of a novel type of serine/threonine kinase that specifically phosphorylates the human goodpasture antigen. *J Biol Chem* **274**, 12642-12649 (1999).

63.   Breslow, D.K. & Weissman, J.S. Membranes in balance: mechanisms of sphingolipid homeostasis. *Mol Cell* **40**, 267-279 (2010).

64.   Fugmann, T.*, et al.* Regulation of secretory transport by protein kinase D-mediated phosphorylation of the ceramide transfer protein. *J Cell Biol* **178**, 15-22 (2007).

65.   Kumagai, K.*, et al.* CERT mediates intermembrane transfer of various molecular species of ceramides. *J Biol Chem* **280**, 6488-6495 (2005).

66.   Gennarino, V.A.*, et al.* NUDT21-spanning CNVs lead to neuropsychiatric disease and altered MeCP2 abundance via alternative polyadenylation. *Elife* **4**(2015).

67.   Gennarino, V.A.*, et al.* Pumilio1 haploinsufficiency leads to SCA1-like neurodegeneration by increasing wild-type Ataxin1 levels. *Cell* **160**, 1087-1098 (2015).

68.   Zhang, B.*, et al.* A conserved RNA-binding protein that regulates sexual fates in the C. elegans hermaphrodite germ line. *Nature* **390**, 477-484 (1997).

69.   Nagase, T.*, et al.* Prediction of the coding sequences of unidentified human genes. III. The coding sequences of 40 new genes (KIAA0081-KIAA0120) deduced by analysis of cDNA clones from human cell line KG-1. *DNA Res* **2**, 37-43 (1995).

70.   Wang, X., McLachlan, J., Zamore, P.D. & Hall, T.M. Modular recognition of RNA by a human pumilio-homology domain. *Cell* **110**, 501-512 (2002).

71.   Rodrigues, D.C.*, et al.* MECP2 Is Post-transcriptionally Regulated during Human Neurodevelopment by Combinatorial Action of RNA-Binding Proteins and miRNAs. *Cell Rep* **17**, 720-734 (2016).

72.   Gennarino, V.A.*, et al.* A Mild PUM1 Mutation Is Associated with Adult-Onset Ataxia, whereas Haploinsufficiency Causes Developmental Delay and Seizures. *Cell* **172**, 924-936 e911 (2018).

73.   Miles, W.O., Tschop, K., Herr, A., Ji, J.Y. & Dyson, N.J. Pumilio facilitates miRNA regulation of the E2F3 oncogene. *Genes Dev* **26**, 356-368 (2012).

74.   Lee, S*., et al.* Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell* **164**, 69-80 (2016).

75.   Kedde, M*., et al.* A Pumilio-induced RNA structure switch in p27-3' UTR controls miR-221 and miR-222 accessibility. *Nat Cell Biol* **12**, 1014-1020 (2010).

76.   Ye, Y*., et al.* A de novo PUM1 Variant in a Girl With a Dravet-Like Syndrome: Case Report and Literature Review. *Front Pediatr* **10**, 759889 (2022).

77.   Voet, J., Ceulemans, B., Kooy, F. & Meuwissen, M.E.C. PUM1 haploinsufficiency is associated with syndromic neurodevelopmental delay and epilepsy. *Am J Med Genet A* **182**, 591-594 (2020).

78.   Bonnemason-Carrere, P*., et al.* PADDAS syndrome associated with hair dysplasia caused by a de novo missense variant of PUM1. *Am J Med Genet A* **179**, 1030-1033 (2019).

79.   Zernant, J*., et al.* Analysis of the ABCA4 gene by next-generation sequencing. *Invest Ophthalmol Vis Sci* **52**, 8479-8487 (2011).

80.   Wolock, C.J*., et al.* A case-control collapsing analysis identifies retinal dystrophy genes associated with ophthalmic disease in patients with no pathogenic ABCA4 variants. *Genet Med* **21**, 2336-2344 (2019).

81.   Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).

82.   Van Der Auwera, G.A. & O'Connor, B.D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*, (O'Reilly Media, 2020).

83.   Danecek, P*., et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158 (2011).

84.   Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).

85.   Richards, S*., et al.* 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* **17**, 405-424 (2015).

86.   Kryazhimskiy, S. & Plotkin, J.B. The population genetics of dN/dS. *PLoS Genet* **4**, e1000304 (2008).

87.   Kimura, M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**, 275-276 (1977).

88.   Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* **20**, 110-121 (2010).

89.   Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886-D894 (2019).

90.   Kircher, M*., et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-315 (2014).

91.   Ioannidis, N.M*., et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* **99**, 877-885 (2016).

92.   Ionita-Laza, I., McCallum, K., Xu, B. & Buxbaum, J.D. A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat Genet* **48**, 214-220 (2016).

93.   Jagadeesh, K.A*., et al.* M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat Genet* **48**, 1581-1586 (2016).

94.   Yeo, G. & Burge, C.B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* **11**, 377-394 (2004).

95.   Desmet, F.O.*, et al.* Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* **37**, e67 (2009).

96.   Xiong, H.Y.*, et al.* RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science* **347**, 1254806 (2015).

97.   Zuallaert, J.*, et al.* SpliceRover: interpretable convolutional neural networks for improved splice site prediction. *Bioinformatics* **34**, 4180-4188 (2018).

98.   Raponi, M.*, et al.* Prediction of single-nucleotide substitutions that result in exon skipping: identification of a splicing silencer in BRCA1 exon 6. *Hum Mutat* **32**, 436-444 (2011).

99.   Jaganathan, K.*, et al.* Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**, 535-548 e524 (2019).

100.   Shimazaki, N., Askary, A., Swanson, P.C. & Lieber, M.R. Mechanistic basis for RAG discrimination between recombination sites and the off-target sites of human lymphomas. *Mol Cell Biol* **32**, 365-375 (2012).

101.   Lee, W., de Prisco, N., Gennarino, V.A. & Buttery, S. How to expand the method details in your Cell Press paper with step-by-step STAR Protocols. *STAR Protoc* **3**, 101550 (2022).

102.   Lee, W., de Prisco, N. & Gennarino, V.A. Identifying patients and assessing variant pathogenicity for an autosomal dominant disease-driving gene. *STAR Protoc* **3**, 101150 (2022).

103.   Delori, F.*, et al.* Quantitative measurements of autofluorescence with the scanning laser ophthalmoscope. *Invest Ophthalmol Vis Sci* **52**, 9379-9390 (2011).

104.   Huang, D.*, et al.* Optical coherence tomography. *Science* **254**, 1178-1181 (1991).

105.   Wojtkowski, M.*, et al.* Ophthalmic imaging by spectral optical coherence tomography. *Am J Ophthalmol* **138**, 412-419 (2004).

106.   Lee, W.*, et al.* Cis-acting modifiers in the ABCA4 locus contribute to the penetrance of the major disease-causing variant in Stargardt disease. *Hum Mol Genet* **30**, 1293-1304 (2021).

107.   Hood, D.C.*, et al.* Thickness of receptor and post-receptor retinal layers in patients with retinitis pigmentosa measured with frequency-domain optical coherence tomography. *Invest Ophthalmol Vis Sci* **50**, 2328-2336 (2009).

108.   McCulloch, D.L.*, et al.* ISCEV extended protocol for the stimulus-response series for light-adapted full-field ERG. *Doc Ophthalmol* **138**, 205-215 (2019).

109.   Lois, N., Holder, G.E., Bunce, C., Fitzke, F.W. & Bird, A.C. Phenotypic subtypes of Stargardt macular dystrophy-fundus flavimaculatus. *Arch Ophthalmol* **119**, 359-369 (2001).

110.   Tanaka, K.*, et al.* The Rapid-Onset Chorioretinopathy Phenotype of ABCA4 Disease. *Ophthalmology* **125**, 89-99 (2018).

111.   Sun, H., Molday, R.S. & Nathans, J. Retinal stimulates ATP hydrolysis by purified and reconstituted ABCR, the photoreceptor-specific ATP-binding cassette transporter responsible for Stargardt disease. *J Biol Chem* **274**, 8269-8281 (1999).

112.   Sun, H., Smallwood, P.M. & Nathans, J. Biochemical defects in ABCR protein variants associated with human retinopathies. *Nat Genet* **26**, 242-246 (2000).

113. Albert, S., *et al.* Identification and Rescue of Splice Defects Caused by Two Neighboring Deep-Intronic ABCA4 Mutations Underlying Stargardt Disease. *Am J Hum Genet* **102**, 517-527 (2018).

114. Garces, F., *et al.* Correlating the Expression and Functional Activity of ABCA4 Disease Variants With the Phenotype of Patients With Stargardt Disease. *Invest Ophthalmol Vis Sci* **59**, 2305-2315 (2018).

115. Fadaie, Z., *et al.* Identification of splice defects due to noncanonical splice site or deep-intronic variants in ABCA4. *Hum Mutat* (2019).

116. Garanto, A., Duijkers, L., Tomkiewicz, T.Z. & Collin, R.W.J. Antisense Oligonucleotide Screening to Optimize the Rescue of the Splicing Defect Caused by the Recurrent Deep-Intronic ABCA4 Variant c.4539+2001G>A in Stargardt Disease. *Genes (Basel)* **10**(2019).

117. Sangermano, R., *et al.* Photoreceptor Progenitor mRNA Analysis Reveals Exon Skipping Resulting from the ABCA4 c.5461-10T-->C Mutation in Stargardt Disease. *Ophthalmology* **123**, 1375-1385 (2016).

118. Weleber, R.G. Stargardt's macular dystrophy. *Arch Ophthalmol* **112**, 752-754 (1994).

119. Cella, W., *et al.* G1961E mutant allele in the Stargardt disease gene ABCA4 causes bull's eye maculopathy. *Exp Eye Res* **89**, 16-24 (2009).

120. Cukras, C.A., *et al.* Centrifugal expansion of fundus autofluorescence patterns in Stargardt disease over time. *Arch Ophthalmol* **130**, 171-179 (2012).

121. Fujinami, K., *et al.* A longitudinal study of Stargardt disease: quantitative assessment of fundus autofluorescence, progression, and genotype correlations. *Invest Ophthalmol Vis Sci* **54**, 8181-8190 (2013).

122. Sparrow, J.R., *et al.* Flecks in Recessive Stargardt Disease: Short-Wavelength Autofluorescence, Near-Infrared Autofluorescence, and Optical Coherence Tomography. *Invest Ophthalmol Vis Sci* **56**, 5029-5039 (2015).

123. Kellner, S., *et al.* Lipofuscin- and melanin-related fundus autofluorescence in patients with ABCA4-associated retinal dystrophies. *Am J Ophthalmol* **147**, 895-902, 902 e891 (2009).

124. Audere, M., Rutka, K., Sepetiene, S. & Lace, B. Presentation of Complex Homozygous Allele in ABCA4 Gene in a Patient with Retinitis Pigmentosa. *Case Rep Ophthalmol Med* **2015**, 452068 (2015).

125. Cremers, F.P., *et al.* Autosomal recessive retinitis pigmentosa and cone-rod dystrophy caused by splice site mutations in the Stargardt's disease gene ABCR. *Hum Mol Genet* **7**, 355-362 (1998).

126. Fukui, T., *et al.* ABCA4 gene mutations in Japanese patients with Stargardt disease and retinitis pigmentosa. *Invest Ophthalmol Vis Sci* **43**, 2819-2824 (2002).

127. Klevering, B.J., Deutman, A.F., Maugeri, A., Cremers, F.P. & Hoyng, C.B. The spectrum of retinal phenotypes caused by mutations in the ABCA4 gene. *Graefes Arch Clin Exp Ophthalmol* **243**, 90-100 (2005).

128. Martinez-Mir, A., *et al.* A new locus for autosomal recessive retinitis pigmentosa (RP19) maps to 1p13-1p21. *Genomics* **40**, 142-146 (1997).

129. Mullins, R.F., *et al.* Autosomal recessive retinitis pigmentosa due to ABCA4 mutations: clinical, pathologic, and molecular characterization. *Invest Ophthalmol Vis Sci* **53**, 1883-1894 (2012).

130. Rozet, J.M.*, et al.* Mutations of the retinal specific ATP binding transporter gene (ABCR) in a single family segregating both autosomal recessive retinitis pigmentosa RP19 and Stargardt disease: evidence of clinical heterogeneity at this locus. *J Med Genet* **36**, 447-451 (1999).
131. Rudolph, G., Kalpadakis, P., Haritoglou, C., Rivera, A. & Weber, B.H. [Mutations in the ABCA4 gene in a family with Stargardt's disease and retinitis pigmentosa (STGD1/RP19)]. *Klin Monbl Augenheilkd* **219**, 590-596 (2002).
132. Shroyer, N.F., Lewis, R.A., Yatsenko, A.N. & Lupski, J.R. Null missense ABCR (ABCA4) mutations in a family with stargardt disease and retinitis pigmentosa. *Invest Ophthalmol Vis Sci* **42**, 2757-2761 (2001).
133. Huang, X.*, et al.* Identification of a Novel Mutation in the ABCA4 Gene in a Chinese Family with Retinitis Pigmentosa Using Exome Sequencing. *Biosci Rep* **38**(2018).
134. Joo, K., Seong, M.W., Park, K.H., Park, S.S. & Woo, S.J. Genotypic profile and phenotype correlations of ABCA4-associated retinopathy in Koreans. *Mol Vis* **25**, 679-690 (2019).
135. Rong, W.N., Wang, X.G. & Sheng, X.L. [ABCA4 mutations and phenotype of different hereditary retinopathies in 3 pedigrees]. *Zhonghua Yan Ke Za Zhi* **54**, 775-781 (2018).
136. Tracewska, A.M.*, et al.* Genetic Spectrum of ABCA4-Associated Retinal Degeneration in Poland. *Genes (Basel)* **10**(2019).
137. Wiszniewski, W.*, et al.* ABCA4 mutations causing mislocalization are found frequently in patients with severe retinal dystrophies. *Hum Mol Genet* **14**, 2769-2778 (2005).
138. Singh, H.P., Jalali, S., Narayanan, R. & Kannabiran, C. Genetic analysis of Indian families with autosomal recessive retinitis pigmentosa by homozygosity screening. *Invest Ophthalmol Vis Sci* **50**, 4065-4071 (2009).
139. Ferrucci, S., Anderson, S.F. & Townsend, J.C. Retinitis pigmentosa inversa. *Optom Vis Sci* **75**, 560-570 (1998).
140. Dias, M.F.*, et al.* Molecular genetics and emerging therapies for retinitis pigmentosa: Basic research and clinical perspectives. *Prog Retin Eye Res* **63**, 107-131 (2018).
141. Verbakel, S.K.*, et al.* Non-syndromic retinitis pigmentosa. *Prog Retin Eye Res* **66**, 157-186 (2018).
142. Maw, M.A.*, et al.* A frameshift mutation in prominin (mouse)-like 1 causes human retinal degeneration. *Hum Mol Genet* **9**, 27-34 (2000).
143. Zhang, Q.*, et al.* Severe retinitis pigmentosa mapped to 4p15 and associated with a novel mutation in the PROM1 gene. *Hum Genet* **122**, 293-299 (2007).
144. Michaelides, M.*, et al.* The PROM1 mutation p.R373C causes an autosomal dominant bull's eye maculopathy associated with rod, rod-cone, and macular dystrophy. *Invest Ophthalmol Vis Sci* **51**, 4771-4780 (2010).
145. Meindl, A.*, et al.* A gene (RPGR) with homology to the RCC1 guanine nucleotide exchange factor is mutated in X-linked retinitis pigmentosa (RP3). *Nat Genet* **13**, 35-42 (1996).
146. Mears, A.J.*, et al.* Remapping of the RP15 locus for X-linked cone-rod degeneration to Xp11.4-p21.1, and identification of a de novo insertion in the RPGR exon ORF15. *Am J Hum Genet* **67**, 1000-1003 (2000).
147. Farrar, G.J.*, et al.* A three-base-pair deletion in the peripherin-RDS gene in one form of retinitis pigmentosa. *Nature* **354**, 478-480 (1991).

148. Wells, J*., et al.* Mutations in the human retinal degeneration slow (RDS) gene can cause either retinitis pigmentosa or macular dystrophy. *Nat Genet* **3**, 213-218 (1993).

149. Hoyng, C.B*., et al.* Autosomal dominant central areolar choroidal dystrophy caused by a mutation in codon 142 in the peripherin/RDS gene. *Am J Ophthalmol* **121**, 623-629 (1996).

150. Akahori, M*., et al.* Dominant mutations in RP1L1 are responsible for occult macular dystrophy. *Am J Hum Genet* **87**, 424-429 (2010).

151. Davidson, A.E*., et al.* RP1L1 variants are associated with a spectrum of inherited retinal diseases including retinitis pigmentosa and occult macular dystrophy. *Hum Mutat* **34**, 506-514 (2013).

152. Petrukhin, K*., et al.* Identification of the gene responsible for Best macular dystrophy. *Nat Genet* **19**, 241-247 (1998).

153. Davidson, A.E*., et al.* Missense mutations in a retinal pigment epithelium protein, bestrophin-1, cause retinitis pigmentosa. *Am J Hum Genet* **85**, 581-592 (2009).

154. Sohocki, M.M*., et al.* Prevalence of AIPL1 mutations in inherited retinal degenerative disease. *Mol Genet Metab* **70**, 142-150 (2000).

155. Kelsell, R.E*., et al.* Mutations in the retinal guanylate cyclase (RETGC-1) gene in dominant cone-rod dystrophy. *Hum Mol Genet* **7**, 1179-1184 (1998).

156. Perrault, I*., et al.* Retinal-specific guanylate cyclase gene mutations in Leber's congenital amaurosis. *Nat Genet* **14**, 461-464 (1996).

157. Freund, C.L*., et al.* De novo mutations in the CRX homeobox gene associated with Leber congenital amaurosis. *Nat Genet* **18**, 311-312 (1998).

158. Freund, C.L*., et al.* Cone-rod dystrophy due to mutations in a novel photoreceptor-specific homeobox gene (CRX) essential for maintenance of the photoreceptor. *Cell* **91**, 543-553 (1997).

159. Riveiro-Alvarez, R*., et al.* Outcome of ABCA4 disease-associated alleles in autosomal recessive retinal dystrophies: retrospective analysis in 420 Spanish families. *Ophthalmology* **120**, 2332-2337 (2013).

160. Papaioannou, M*., et al.* An analysis of ABCR mutations in British patients with recessive retinal dystrophies. *Invest Ophthalmol Vis Sci* **41**, 16-19 (2000).

161. Fishman, G.A. Fundus flavimaculatus. A clinical classification. *Arch Ophthalmol* **94**, 2061-2067 (1976).

162. Fishman, G.A. Historical evolution in the understanding of Stargardt macular dystrophy. *Ophthalmic Genet* **31**, 183-189 (2010).

163. Kajiwara, K., Berson, E.L. & Dryja, T.P. Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. *Science* **264**, 1604-1608 (1994).

164. Costa, K.A., Salles, M.V., Whitebirch, C., Chiang, J. & Sallum, J.M.F. Gene panel sequencing in Brazilian patients with retinitis pigmentosa. *Int J Retina Vitreous* **3**, 33 (2017).

165. Ma, C.J*., et al.* Late-onset pattern macular dystrophy mimicking ABCA4 and PRPH2 disease is caused by a homozygous frameshift mutation in ROM1. *Cold Spring Harb Mol Case Stud* **5**(2019).

166. Haider, N.B*., et al.* Mutation of a nuclear receptor gene, NR2E3, causes enhanced S cone syndrome, a disorder of retinal cell fate. *Nat Genet* **24**, 127-131 (2000).

167.   Coppieters, F*., et al.* Recurrent mutation in the first zinc finger of the orphan nuclear receptor NR2E3 causes autosomal dominant retinitis pigmentosa. *Am J Hum Genet* **81**, 147-157 (2007).

168.   Escher, P*., et al.* Mutations in NR2E3 can cause dominant or recessive retinal degenerations in the same family. *Hum Mutat* **30**, 342-351 (2009).

169.   Fishman, G.A., Anderson, R.J. & Lourenco, P. Prevalence of posterior subcapsular lens opacities in patients with retinitis pigmentosa. *Br J Ophthalmol* **69**, 263-266 (1985).

170.   Fujiwara, K*., et al.* Risk Factors for Posterior Subcapsular Cataract in Retinitis Pigmentosa. *Invest Ophthalmol Vis Sci* **58**, 2534-2537 (2017).

171.   Vasavada, A.R., Mamidipudi, P.R. & Sharma, P.S. Morphology of and visual performance with posterior subcapsular cataract. *J Cataract Refract Surg* **30**, 2097-2104 (2004).

172.   Gwon, A., Mantras, C., Gruber, L. & Cunanan, C. Concanavalin A-induced posterior subcapsular cataract: a new model of cataractogenesis. *Invest Ophthalmol Vis Sci* **34**, 3483-3488 (1993).

173.   Abalem, M.F., Omari, A.A., Schlegel, D., Khan, N.W. & Jayasundera, T. Macular hyperpigmentary changes in ABCA4-Stargardt disease. *Int J Retina Vitreous* **5**, 9 (2019).

174.   Zhao, P.Y*., et al.* Peripheral Pigmented Retinal Lesions in Stargardt Disease. *Am J Ophthalmol* **188**, 104-110 (2018).

175.   Schuerch, K., Marsiglia, M., Lee, W., Tsang, S.H. & Sparrow, J.R. Multimodal Imaging of Disease-Associated Pigmentary Changes in Retinitis Pigmentosa. *Retina* **36 Suppl 1**, S147-S158 (2016).

176.   Flannery, J.G., Farber, D.B., Bird, A.C. & Bok, D. Degenerative changes in a retina affected with autosomal dominant retinitis pigmentosa. *Invest Ophthalmol Vis Sci* **30**, 191-211 (1989).

177.   Woods, C.G*., et al.* Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. *Am J Hum Genet* **78**, 889-896 (2006).

178.   Bittles, A.H. & Black, M.L. Evolution in health and medicine Sackler colloquium: Consanguinity, human evolution, and complex diseases. *Proc Natl Acad Sci U S A* **107 Suppl 1**, 1779-1786 (2010).

179.   Collins, R.L*., et al.* A structural variation reference for medical and population genetics. *Nature* **581**, 444-451 (2020).

180.   McCulloch, D.L*., et al.* ISCEV Standard for full-field clinical electroretinography (2015 update). *Doc Ophthalmol* **130**, 1-12 (2015).

181.   Liu, X., Jian, X. & Boerwinkle, E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat* **32**, 894-899 (2011).

182.   Liu, X., Li, C., Mou, C., Dong, Y. & Tu, Y. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Genome Med* **12**, 103 (2020).

183.   Cremers, F.P.M., Lee, W., Collin, R.W.J. & Allikmets, R. Clinical spectrum, genetic complexity and therapeutic approaches for retinal disease caused by ABCA4 mutations. *Prog Retin Eye Res* **79**, 100861 (2020).

184.   Maugeri, A*., et al.* The 2588G-->C mutation in the ABCR gene is a mild frequent founder mutation in the Western European population and allows the classification of ABCR mutations in patients with Stargardt disease. *Am J Hum Genet* **64**, 1024-1035 (1999).

185. Runhart, E.H.*, et al.* Late-Onset Stargardt Disease Due to Mild, Deep-Intronic ABCA4 Alleles. *Invest Ophthalmol Vis Sci* **60**, 4249-4256 (2019).

186. Lewis, R.A.*, et al.* Genotype/Phenotype analysis of a photoreceptor-specific ATP-binding cassette transporter gene, ABCR, in Stargardt disease. *Am J Hum Genet* **64**, 422-434 (1999).

187. Fujinami, K.*, et al.* A longitudinal study of stargardt disease: clinical and electrophysiologic assessment, progression, and genotype correlations. *Am J Ophthalmol* **155**, 1075-1088 e1013 (2013).

188. Di Iorio, V.*, et al.* ASSOCIATION BETWEEN GENOTYPE AND DISEASE PROGRESSION IN ITALIAN STARGARDT PATIENTS: A Retrospective Natural History Study. *Retina* **39**, 1399-1409 (2019).

189. Del Pozo-Valero, M.*, et al.* Genotype-Phenotype Correlations in a Spanish Cohort of 506 Families With Biallelic ABCA4 Pathogenic Variants. *Am J Ophthalmol* **219**, 195-204 (2020).

190. Zhang, N.*, et al.* Protein misfolding and the pathogenesis of ABCA4-associated retinal degenerations. *Hum Mol Genet* **24**, 3220-3237 (2015).

191. Fakin, A.*, et al.* The Effect on Retinal Structure and Function of 15 Specific ABCA4 Mutations: A Detailed Examination of 82 Hemizygous Patients. *Invest Ophthalmol Vis Sci* **57**, 5963-5973 (2016).

192. Cideciyan, A.V.*, et al.* ABCA4 disease progression and a proposed strategy for gene therapy. *Hum Mol Genet* **18**, 931-941 (2009).

193. Shroyer, N.F., Lewis, R.A., Yatsenko, A.N., Wensel, T.G. & Lupski, J.R. Cosegregation and functional analysis of mutant ABCR (ABCA4) alleles in families that manifest both Stargardt disease and age-related macular degeneration. *Hum Mol Genet* **10**, 2671-2678 (2001).

194. Yatsenko, A.N., Wiszniewski, W., Zaremba, C.M., Jamrich, M. & Lupski, J.R. Evolution of ABCA4 proteins in vertebrates. *J Mol Evol* **60**, 72-80 (2005).

195. Schindler, E.I.*, et al.* Deducing the pathogenic contribution of recessive ABCA4 alleles in an outbred population. *Hum Mol Genet* **19**, 3693-3701 (2010).

196. Green, J.S.*, et al.* The genetic architecture of Stargardt macular dystrophy (STGD1): a longitudinal 40-year study in a genetic isolate. *Eur J Hum Genet* **28**, 925-937 (2020).

197. Klevering, B.J.*, et al.* Phenotypic variations in a family with retinal dystrophy as result of different mutations in the ABCR gene. *Br J Ophthalmol* **83**, 914-918 (1999).

198. Martinez-Mir, A.*, et al.* Retinitis pigmentosa caused by a homozygous mutation in the Stargardt disease gene ABCR. *Nat Genet* **18**, 11-12 (1998).

199. Ervin, A.M.*, et al.* A Workshop on Measuring the Progression of Atrophy Secondary to Stargardt Disease in the ProgStar Studies: Findings and Lessons Learned. *Transl Vis Sci Technol* **8**, 16 (2019).

200. Sears, A.E.*, et al.* Towards Treatment of Stargardt Disease: Workshop Organized and Sponsored by the Foundation Fighting Blindness. *Transl Vis Sci Technol* **6**, 6 (2017).

201. Chen, B., Tosha, C., Gorin, M.B. & Nusinowitz, S. Analysis of autofluorescent retinal images and measurement of atrophic lesion growth in Stargardt disease. *Exp Eye Res* **91**, 143-152 (2010).

202. Pfau, M.*, et al.* Photoreceptor degeneration in ABCA4-associated retinopathy and its genetic correlates. *JCI Insight* **7**(2022).

203. Strauss, R.W.*, et al.* The Natural History of the Progression of Atrophy Secondary to Stargardt Disease (ProgStar) Studies: Design and Baseline Characteristics: ProgStar Report No. 1. *Ophthalmology* **123**, 817-828 (2016).

204. Strauss, R.W.*, et al.* Progression of Stargardt Disease as Determined by Fundus Autofluorescence Over a 12-Month Period: ProgStar Report No. 11. *JAMA Ophthalmol* **137**, 1134-1145 (2019).

205. Strauss, R.W.*, et al.* The Progression of the Stargardt Disease Type 4 (ProgStar-4) Study: Design and Baseline Characteristics (ProgStar-4 Report No. 1). *Ophthalmic Res* **60**, 185-194 (2018).

206. Strauss, R.W.*, et al.* Progression of Stargardt Disease as Determined by Fundus Autofluorescence in the Retrospective Progression of Stargardt Disease Study (ProgStar Report No. 9). *JAMA Ophthalmol* **135**, 1232-1241 (2017).

207. Strauss, R.W.*, et al.* Incidence of Atrophic Lesions in Stargardt Disease in the Progression of Atrophy Secondary to Stargardt Disease (ProgStar) Study: Report No. 5. *JAMA Ophthalmol* **135**, 687-695 (2017).

208. Sparrow, J.R., Duncker, T., Schuerch, K., Paavo, M. & de Carvalho, J.R.L., Jr. Lessons learned from quantitative fundus autofluorescence. *Prog Retin Eye Res* **74**, 100774 (2020).

209. Fritsche, L.G.*, et al.* A subgroup of age-related macular degeneration is associated with mono-allelic sequence variants in the ABCA4 gene. *Invest Ophthalmol Vis Sci* **53**, 2112-2118 (2012).

210. De La Paz, M.A.*, et al.* Analysis of the Stargardt disease gene (ABCR) in age-related macular degeneration. *Ophthalmology* **106**, 1531-1536 (1999).

211. Lindner, M.*, et al.* Differential Disease Progression in Atrophic Age-Related Macular Degeneration and Late-Onset Stargardt Disease. *Invest Ophthalmol Vis Sci* **58**, 1001-1007 (2017).

212. Keenan, T.D.*, et al.* Progression of Geographic Atrophy in Age-related Macular Degeneration: AREDS2 Report Number 16. *Ophthalmology* **125**, 1913-1928 (2018).

213. Thiele, S.*, et al.* Multimodal Imaging Patterns for Development of Central Atrophy Secondary to Age-Related Macular Degeneration. *Invest Ophthalmol Vis Sci* **59**, AMD1-AMD11 (2018).

214. Fleckenstein, M.*, et al.* The Progression of Geographic Atrophy Secondary to Age-Related Macular Degeneration. *Ophthalmology* **125**, 369-390 (2018).

215. Fleckenstein, M.*, et al.* Distinct Genetic Risk Profile of the Rapidly Progressing Diffuse-Trickling Subtype of Geographic Atrophy in Age-Related Macular Degeneration (AMD). *Invest Ophthalmol Vis Sci* **57**, 2463-2471 (2016).

216. Mauschitz, M.M.*, et al.* Topography of geographic atrophy in age-related macular degeneration. *Invest Ophthalmol Vis Sci* **53**, 4932-4939 (2012).

217. Holz, F.G.*, et al.* Progression of geographic atrophy and impact of fundus autofluorescence patterns in age-related macular degeneration. *Am J Ophthalmol* **143**, 463-472 (2007).

218. Strauss, R.W.*, et al.* The progression of Stargardt Disease as determined by fundus autofluorescence over a 24-month period (ProgStar Report No. 17). *Am J Ophthalmol* (2023).

219. Greenberg, J.P.*, et al.* Quantitative fundus autofluorescence in healthy eyes. *Invest Ophthalmol Vis Sci* **54**, 5684-5693 (2013).

220. Dunn, T.M., Tifft, C.J. & Proia, R.L. A perilous path: the inborn errors of sphingolipid metabolism. *J. Lipid Res.* **60**, 475-483 (2019).
221. Hannun, Y.A. & Obeid, L.M. Sphingolipids and their metabolism in physiology and disease. *Nat. Rev. Mol. Cell Biol.* **19**, 175-191 (2018).
222. Ozkara, H.A. Recent advances in the biochemistry and genetics of sphingolipidoses. *Brain Dev* **26**, 497-505 (2004).
223. Dunn, T.M., Tifft, C.J. & Proia, R.L. A perilous path: the inborn errors of sphingolipid metabolism. *J Lipid Res* **60**, 475-483 (2019).
224. Sabourdy, F.*, et al.* Monogenic neurological disorders of sphingolipid metabolism. *Biochim Biophys Acta* **1851**, 1040-1051 (2015).
225. Breslow, D.K. & Weissman, J.S. Membranes in balance: mechanisms of sphingolipid homeostasis. *Mol. Cell* **40**, 267-279 (2010).
226. Hanada, K.*, et al.* Molecular machinery for non-vesicular trafficking of ceramide. *Nature* **426**, 803-809 (2003).
227. Fugmann, T.*, et al.* Regulation of secretory transport by protein kinase D-mediated phosphorylation of the ceramide transfer protein. *J. Cell Biol.* **178**, 15-22 (2007).
228. de Ligt, J.*, et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* **367**, 1921-1929 (2012).
229. De Rubeis, S.*, et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209-215 (2014).
230. Hamdan, F.F.*, et al.* De novo mutations in moderate or severe intellectual disability. *PLoS Genet.* **10**, e1004772 (2014).
231. Helbig, K.L.*, et al.* Diagnostic exome sequencing provides a molecular diagnosis for a significant proportion of patients with epilepsy. *Genet Med* **18**, 898-905 (2016).
232. Iossifov, I.*, et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216-221 (2014).
233. Murakami, H.*, et al.* Intellectual disability-associated gain-of-function mutations in CERT1 that encodes the ceramide transport protein CERT. *PLoS One* **15**, e0243980 (2020).
234. Study, T.D.D.D. & The Deciphering Developmental Disorders, S. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* **519**, 223-228 (2015).
235. Takata, A.*, et al.* Integrative Analyses of De Novo Mutations Provide Deeper Biological Insights into Autism Spectrum Disorder. *Cell Rep.* **22**, 734-747 (2018).
236. Wang, T.*, et al.* De novo genic mutations among a Chinese autism spectrum disorder cohort. *Nat Commun* **7**, 13316 (2016).
237. de Ligt, J.*, et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* **367**, 1921-1929 (2012).
238. Iossifov, I.*, et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216-221 (2014).
239. Hamdan, F.F.*, et al.* De novo mutations in moderate or severe intellectual disability. *PLoS Genet* **10**, e1004772 (2014).
240. Breslow, D.K.*, et al.* Orm family proteins mediate sphingolipid homeostasis. *Nature* **463**, 1048-1053 (2010).
241. Clarke, B.A.*, et al.* The Ormdl genes regulate the sphingolipid synthesis pathway to ensure proper myelination and neurologic function in mice. *Elife* **8**(2019).

242. Karsai, G.*, et al.* DEGS1-associated aberrant sphingolipid metabolism impairs nervous system function in humans. *J. Clin. Invest.* **129**, 1229-1239 (2019).
243. Pant, D.C., Aguilera-Albesa, S. & Pujol, A. Ceramide signalling in inherited and multifactorial brain metabolic diseases. *Neurobiol. Dis.* **143**, 105014 (2020).
244. Edvardson, S.*, et al.* Deficiency of the alkaline ceramidase ACER3 manifests in early childhood by progressive leukodystrophy. *J. Med. Genet.* **53**, 389-396 (2016).
245. Bhuiyan, R.H.*, et al.* Loss of Enzyme Activity in Mutated B4GALNT1 Gene Products in Patients with Hereditary Spastic Paraplegia Results in Relatively Mild Neurological Disorders: Similarity with Phenotypes of B4galnt1 Knockout Mice. *Neuroscience* **397**, 94-106 (2019).
246. Boccuto, L.*, et al.* A mutation in a ganglioside biosynthetic enzyme, ST3GAL5, results in salt & pepper syndrome, a neurocutaneous disorder with altered glycolipid and glycoprotein glycosylation. *Hum Mol Genet* **23**, 418-433 (2014).
247. Harlalka, G.V.*, et al.* Mutations in B4GALNT1 (GM2 synthase) underlie a new disorder of ganglioside biosynthesis. *Brain* **136**, 3618-3624 (2013).
248. Simpson, E.P., Henry, Y.K., Henkel, J.S., Smith, R.G. & Appel, S.H. Increased lipid peroxidation in sera of ALS patients: a potential biomarker of disease burden. *Neurology* **62**, 1758-1765 (2004).
249. Wakil, S.M.*, et al.* Novel B4GALNT1 mutations in a complicated form of hereditary spastic paraplegia. *Clin Genet* **86**, 500-501 (2014).
250. Wang, H.*, et al.* Early growth and development impairments in patients with ganglioside GM3 synthase deficiency. *Clin Genet* **89**, 625-629 (2016).
251. Grob, J.J., Collet-Villette, A.M. & Bonerandi, J.J. [Piezogenic cellulo-adipose hernia of the leg]. *Ann Dermatol Venereol* **114**, 1567-1569 (1987).
252. Lone, M.A.*, et al.* SPTLC1 variants associated with ALS produce distinct sphingolipid signatures through impaired interaction with ORMDL proteins. *J Clin Invest* **132**(2022).
253. Mohassel, P.*, et al.* Childhood amyotrophic lateral sclerosis caused by excess sphingolipid synthesis. *Nat Med* (2021).
254. Srivastava, S.*, et al.* SPTSSA variants alter sphingolipid synthesis and cause a complex hereditary spastic paraplegia. *Brain* (2023).
255. Kawano, M., Kumagai, K., Nishijima, M. & Hanada, K. Efficient trafficking of ceramide from the endoplasmic reticulum to the Golgi apparatus requires a VAMP-associated protein-interacting FFAT motif of CERT. *J Biol Chem* **281**, 30279-30288 (2006).
256. Kumagai, K., Kawano, M., Shinkai-Ouchi, F., Nishijima, M. & Hanada, K. Interorganelle trafficking of ceramide is regulated by phosphorylation-dependent cooperativity between the PH and START domains of CERT. *J Biol Chem* **282**, 17758-17766 (2007).
257. Firth, H.V.*, et al.* DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am. J. Hum. Genet.* **84**, 524-533 (2009).
258. Sobreira, N., Schiettecatte, F., Valle, D. & Hamosh, A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum. Mutat.* **36**, 928-930 (2015).
259. Richards, S.*, et al.* 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. *Genetics in Medicine* **17**, 405-423 (2015).

260. Karczewski, K.J., *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).
261. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
262. Jagadeesh, K.A., *et al.* M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat. Genet.* **48**, 1581-1586 (2016).
263. Ioannidis, N.M., *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* **99**, 877-885 (2016).
264. Ionita-Laza, I., McCallum, K., Xu, B. & Buxbaum, J.D. A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat. Genet.* **48**, 214-220 (2016).
265. Lai, K.L., *et al.* Investigating PUM1 mutations in a Taiwanese cohort with cerebellar ataxia. *Parkinsonism Relat Disord* **66**, 220-223 (2019).
266. Imaizumi, T., Mogami, Y., Okamoto, N., Yamamoto-Shimojima, K. & Yamamoto, T. De novo 1p35.2 microdeletion including PUM1 identified in a patient with sporadic West syndrome. *Congenit Anom (Kyoto)* **59**, 193-194 (2019).
267. Enwerem, III, *et al.* Human Pumilio proteins directly bind the CCR4-NOT deadenylase complex to regulate the transcriptome. *RNA* **27**, 445-464 (2021).
268. Gupta, Y.K., Nair, D.T., Wharton, R.P. & Aggarwal, A.K. Structures of human Pumilio with noncognate RNAs reveal molecular mechanisms for binding promiscuity. *Structure* **16**, 549-557 (2008).
269. Trcek, T., Sato, H., Singer, R.H. & Maquat, L.E. Temporal and spatial characterization of nonsense-mediated mRNA decay. *Genes Dev* **27**, 541-551 (2013).
270. Nagy, E. & Maquat, L.E. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci* **23**, 198-199 (1998).
271. Hu, Z., Yau, C. & Ahmed, A.A. A pan-cancer genome-wide analysis reveals tumour dependencies by induction of nonsense-mediated decay. *Nat Commun* **8**, 15943 (2017).
272. Lappalainen, T., *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506-511 (2013).
273. Lindeboom, R.G., Supek, F. & Lehner, B. The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat Genet* **48**, 1112-1118 (2016).
274. Rivas, M.A., *et al.* Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science* **348**, 666-669 (2015).
275. Coban-Akdemir, Z., *et al.* Identifying Genes Whose Mutant Transcripts Cause Dominant Disease Traits by Potential Gain-of-Function Alleles. *Am J Hum Genet* **103**, 171-187 (2018).
276. Gudmundsson, S., *et al.* Addendum: The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **597**, E3-E4 (2021).
277. Sharo, A.G., Hu, Z., Sunyaev, S.R. & Brenner, S.E. StrVCTVRE: A supervised learning method to predict the pathogenicity of human genome structural variants. *Am J Hum Genet* **109**, 195-209 (2022).
278. Lu, G. & Hall, T.M. Alternate modes of cognate RNA recognition by human PUMILIO proteins. *Structure* **19**, 361-367 (2011).
279. Lichtarge, O., Bourne, H.R. & Cohen, F.E. An evolutionary trace method defines binding surfaces common to protein families. *J Mol Biol* **257**, 342-358 (1996).

280. Pertea, M., Lin, X. & Salzberg, S.L. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res* **29**, 1185-1190 (2001).

281. Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q. & Krainer, A.R. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* **31**, 3568-3571 (2003).

282. Charbel Issa, P., Barnard, A.R., Herrmann, P., Washington, I. & MacLaren, R.E. Rescue of the Stargardt phenotype in Abca4 knockout mice through inhibition of vitamin A dimerization. *Proc Natl Acad Sci U S A* **112**, 8415-8420 (2015).

283. Petrukhin, K. Pharmacological inhibition of lipofuscin accumulation in the retina as a therapeutic strategy for dry AMD treatment. *Drug Discov Today Ther Strateg* **10**, e11-e20 (2013).

284. Lee, W.*, et al.* A genotype-phenotype correlation matrix for ABCA4 disease based on long-term prognostic outcomes. *JCI Insight* **7**(2022).

285. Cummings, B.B.*, et al.* Transcript expression-aware annotation improves rare variant interpretation. *Nature* **581**, 452-458 (2020).

286. Molday, R.S., Garces, F.A., Scortecci, J.F. & Molday, L.L. Structure and function of ABCA4 and its role in the visual cycle and Stargardt macular degeneration. *Prog Retin Eye Res* **89**, 101036 (2022).

287. Tomishige, N., Kumagai, K., Kusuda, J., Nishijima, M. & Hanada, K. Casein kinase Igamma2 down-regulates trafficking of ceramide in the synthesis of sphingomyelin. *Mol Biol Cell* **20**, 348-357 (2009).

288. Kumagai, K. & Hanada, K. Structure, functions and regulation of CERT, a lipid-transfer protein for the delivery of ceramide at the ER-Golgi membrane contact sites. *FEBS Lett* **593**, 2366-2377 (2019).

289. Pinna, L.A. & Ruzzene, M. How do protein kinases recognize their substrates? *Biochim Biophys Acta* **1314**, 191-225 (1996).

290. Conley, S.M.*, et al.* Prph2 initiates outer segment morphogenesis but maturation requires Prph2/Rom1 oligomerization. *Hum Mol Genet* **28**, 459-475 (2019).

291. Zulliger, R., Conley, S.M., Mwoyosvi, M.L., Al-Ubaidi, M.R. & Naash, M.I. Oligomerization of Prph2 and Rom1 is essential for photoreceptor outer segment formation. *Hum Mol Genet* **27**, 3507-3518 (2018).

292. Botta, S.d.P., N.; Chemiakine, A.; Cabaj, M.; Patel, P.; Soni, R.K.; Gennarino, V.A. Dosage sensitivity in Pumilio1-associated diseases involves two distinct mechanisms. *bioRXiv* (2021).

293. Rackham, O.J., Shihab, H.A., Johnson, M.R. & Petretto, E. EvoTol: a protein-sequence based evolutionary intolerance framework for disease-gene prioritization. *Nucleic Acids Res* **43**, e33 (2015).

294. Petrovski, S.*, et al.* The Intolerance of Regulatory Sequence to Genetic Variation Predicts Gene Dosage Sensitivity. *PLoS Genet* **11**, e1005492 (2015).

295. Samocha, K.E.*, et al.* A framework for the interpretation of de novo mutation in human disease. *Nat Genet* **46**, 944-950 (2014).

296. Lek, M.*, et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291 (2016).

297. Li, C., Zhi, D., Wang, K. & Liu, X. MetaRNN: differentiating rare pathogenic and rare benign missense SNVs and InDels using deep learning. *Genome Med* **14**, 115 (2022).

298. Landrum, M.J.*, et al.* ClinVar: improvements to accessing data. *Nucleic Acids Res* **48**, D835-D844 (2020).

299. Cideciyan, A.V.*, et al.* ABCA4-associated retinal degenerations spare structure and function of the human parapapillary retina. *Invest Ophthalmol Vis Sci* **46**, 4739-4746 (2005).

300. Ernest, J.T. & Krill, A.E. Fluorescein studies in fundus flavimaculatus and drusen. *Am J Ophthalmol* **62**, 1-6 (1966).

301. Duncker, T.*, et al.* Quantitative Fundus Autofluorescence and Optical Coherence Tomography in PRPH2/RDS- and ABCA4-Associated Disease Exhibiting Phenotypic Overlap. *Invest Ophthalmol Vis Sci* **56**, 3159-3170 (2015).

302. Battaglia Parodi, M., De Benedetto, U., Knutsson, K.A. & Bandello, F.M. Bilateral choroidal neovascularization associated with bilateral ABCA4 gene mutation. *Eur J Ophthalmol* **22**, 485-487 (2012).

303. Essilfie, J.O., Huynh, A., Gorin, M.B. & Sadda, S.R. A Case of Intraretinal Peripapillary Neovascularization in Abca4-Related Retinopathy. *Retin Cases Brief Rep* **15**, 5-9 (2021).

304. Cohen, J.*, et al.* Retinal pigment epithelial detachment in ABCA4-associated Stargardt's disease. *Ophthalmic Surg Lasers Imaging Retina* **44**, 401-404 (2013).

305. Rizzo, S.*, et al.* Macular hole in Stargardt disease: Clinical and ultra-structural observation. *Ophthalmic Genet* **38**, 486-489 (2017).

306. Sodi, A., Bini, A., Passerini, I., Menchini, U. & Torricelli, F. Occurrence of full-thickness macular hole complicating Stargardt disease with ABCR mutation. *Eur J Ophthalmol* **16**, 335-338 (2006).

307. Aleman, T.S.*, et al.* Spinocerebellar ataxia type 7 (SCA7) shows a cone-rod dystrophy phenotype. *Exp Eye Res* **74**, 737-745 (2002).

308. Park, S.J., Woo, S.J., Park, K.H., Hwang, J.M. & Chung, H. Morphologic photoreceptor abnormality in occult macular dystrophy on spectral-domain optical coherence tomography. *Invest Ophthalmol Vis Sci* **51**, 3673-3679 (2010).

309. Thiadens, A.A.*, et al.* Progressive loss of cones in achromatopsia: an imaging study using spectral-domain optical coherence tomography. *Invest Ophthalmol Vis Sci* **51**, 5952-5957 (2010).

310. Carroll, J.*, et al.* The effect of cone opsin mutations on retinal structure and the integrity of the photoreceptor mosaic. *Invest Ophthalmol Vis Sci* **53**, 8006-8015 (2012).

311. Sisk, R.A. & Leng, T. Multimodal imaging and multifocal electroretinography demonstrate autosomal recessive Stargardt disease may present like occult macular dystrophy. *Retina* **34**, 1567-1575 (2014).

312. Sergouniotis, P.I.*, et al.* High-resolution optical coherence tomography imaging in KCNV2 retinopathy. *Br J Ophthalmol* **96**, 213-217 (2012).

313. Ahn, S.J.*, et al.* Clinical and genetic characteristics of Korean occult macular dystrophy patients. *Invest Ophthalmol Vis Sci* **54**, 4856-4863 (2013).

314. Watkins, W.M., Schoenberger, S.D., Lavin, P. & Agarwal, A. Circumscribed outer foveolar defects in spinocerebellar ataxia type 7. *Retin Cases Brief Rep* **7**, 294-296 (2013).

315. Greenberg, J.P.*, et al.* Spectral-domain optical coherence tomography staging and autofluorescence imaging in achromatopsia. *JAMA Ophthalmol* **132**, 437-445 (2014).

316. Sundaram, V.*, et al.* Retinal structure and function in achromatopsia: implications for gene therapy. *Ophthalmology* **121**, 234-245 (2014).

317. Mukherjee, R., *et al.* A detailed phenotypic description of autosomal dominant cone dystrophy due to a de novo mutation in the GUCY2D gene. *Eye (Lond)* **28**, 481-487 (2014).

318. Roosing, S., *et al.* Disruption of the basal body protein POC1B results in autosomal-recessive cone-rod dystrophy. *Am J Hum Genet* **95**, 131-142 (2014).

319. Doshi, R.R., Fortun, J.A., Kim, B.T., Dubovy, S.R. & Rosenfeld, P.J. Pseudocystic foveal cavitation in tamoxifen retinopathy. *Am J Ophthalmol* **157**, 1291-1298 e1293 (2014).

320. Kohl, S., *et al.* Mutations in the unfolded protein response regulator ATF6 cause the cone dysfunction disorder achromatopsia. *Nat Genet* **47**, 757-765 (2015).

321. Maertz, J., Gloeckle, N., Nentwich, M.M. & Rudolph, G. [Genotype-phenotype correlation in patients with PRPH2-mutations]. *Klin Monbl Augenheilkd* **232**, 266-274 (2015).

322. Palejwala, N.V., *et al.* Insights into Autosomal Dominant Stargardt-Like Macular Dystrophy through Multimodality Diagnostic Imaging. *Retina* **36**, 119-130 (2016).

323. Manes, G., *et al.* Cone dystrophy or macular dystrophy associated with novel autosomal dominant GUCA1A mutations. *Mol Vis* **23**, 198-209 (2017).

324. Lee, S., Kim, H.A. & Yoon, Y.H. OCT Angiography Findings of Tamoxifen Retinopathy: Similarity with Macular Telangiectasia Type 2. *Ophthalmol Retina* **3**, 681-689 (2019).

325. Simonelli, F., *et al.* New ABCR mutations and clinical phenotype in Italian patients with Stargardt disease. *Invest Ophthalmol Vis Sci* **41**, 892-897 (2000).

326. Maugeri, A., *et al.* Mutations in the ABCA4 (ABCR) gene are the major cause of autosomal recessive cone-rod dystrophy. *Am J Hum Genet* **67**, 960-966 (2000).

327. Xu, Y., *et al.* Mutations of 60 known causative genes in 157 families with retinitis pigmentosa based on exome sequencing. *Hum Genet* **133**, 1255-1271 (2014).

328. Fujinami, K., *et al.* Clinical and molecular characteristics of childhood-onset Stargardt disease. *Ophthalmology* **122**, 326-334 (2015).

329. Jiang, F., *et al.* Screening of ABCA4 Gene in a Chinese Cohort With Stargardt Disease or Cone-Rod Dystrophy With a Report on 85 Novel Mutations. *Invest Ophthalmol Vis Sci* **57**, 145-152 (2016).

330. Maggi, J., *et al.* Long-Range PCR-Based NGS Applications to Diagnose Mendelian Retinal Diseases. *Int J Mol Sci* **22**(2021).

331. Lee, W., *et al.* A genotype-phenotype correlation matrix for Stargardt/ABCA4 disease based on long-term prognostic outcomes. *JCI Insight* (2021).

332. Maia-Lopes, S., *et al.* ABCA4 mutations in Portuguese Stargardt patients: identification of new mutations and their phenotypic analysis. *Mol Vis* **15**, 584-591 (2009).

333. Lambertus, S., *et al.* Highly sensitive measurements of disease progression in rare disorders: Developing and validating a multimodal model of retinal degeneration in Stargardt disease. *PLoS One* **12**, e0174020 (2017).

334. Nassisi, M., *et al.* Expanding the Mutation Spectrum in ABCA4: Sixty Novel Disease Causing Variants and Their Associated Phenotype in a Large French Stargardt Cohort. *Int J Mol Sci* **19**(2018).

335. Salles, M.V., *et al.* Variants in the ABCA4 gene in a Brazilian population with Stargardt disease. *Mol Vis* **24**, 546-559 (2018).

336. Fujinami, K.*, et al.* Detailed genetic characteristics of an international large cohort of patients with Stargardt disease: ProgStar study report 8. *Br J Ophthalmol* **103**, 390-397 (2019).

337. Rozet, J.M.*, et al.* Spectrum of ABCR gene mutations in autosomal recessive macular dystrophies. *Eur J Hum Genet* **6**, 291-295 (1998).

338. Fishman, G.A.*, et al.* Variation of clinical expression in patients with Stargardt dystrophy and sequence variations in the ABCR gene. *Arch Ophthalmol* **117**, 504-510 (1999).

339. Briggs, C.E.*, et al.* Mutations in ABCR (ABCA4) in patients with Stargardt macular degeneration or cone-rod degeneration. *Invest Ophthalmol Vis Sci* **42**, 2229-2236 (2001).

340. Gerth, C.*, et al.* Phenotypes of 16 Stargardt macular dystrophy/fundus flavimaculatus patients with known ABCA4 mutations and evaluation of genotype-phenotype correlation. *Graefes Arch Clin Exp Ophthalmol* **240**, 628-638 (2002).

341. Hargitai, J.*, et al.* Correlation of clinical and genetic findings in Hungarian patients with Stargardt disease. *Invest Ophthalmol Vis Sci* **46**, 4402-4408 (2005).

342. Fujinami, K.*, et al.* The clinical effect of homozygous ABCA4 alleles in 18 patients. *Ophthalmology* **120**, 2324-2331 (2013).

343. Cideciyan, A.V.*, et al.* Macular function in macular degenerations: repeatability of microperimetry as a potential outcome measure for ABCA4-associated retinopathy trials. *Invest Ophthalmol Vis Sci* **53**, 841-852 (2012).

344. Sciezynska, A.*, et al.* Next-generation sequencing of ABCA4: High frequency of complex alleles and novel mutations in patients with retinal dystrophies from Central Europe. *Exp Eye Res* **145**, 93-99 (2016).

345. Zaneveld, J.*, et al.* Comprehensive analysis of patients with Stargardt macular dystrophy reveals new genotype-phenotype correlations and unexpected diagnostic revisions. *Genet Med* **17**, 262-270 (2015).

346. Oh, J.K.*, et al.* Optical Gap Biomarker in Cone-Dominant Retinal Dystrophy. *Am J Ophthalmol* **218**, 40-53 (2020).

347. Oh, K.T.*, et al.* Electroretinographic findings in patients with Stargardt disease and fundus flavimaculatus. *Retina* **24**, 920-928 (2004).

348. Fingert, J.H.*, et al.* Case of Stargardt disease caused by uniparental isodisomy. *Arch Ophthalmol* **124**, 744-745 (2006).

349. Hwang, J.C.*, et al.* Peripapillary atrophy in Stargardt disease. *Retina* **29**, 181-186 (2009).

350. Chacon-Camacho, O.F., Granillo-Alvarez, M., Ayala-Ramirez, R. & Zenteno, J.C. ABCA4 mutational spectrum in Mexican patients with Stargardt disease: Identification of 12 novel mutations and evidence of a founder effect for the common p.A1773V mutation. *Exp Eye Res* **109**, 77-82 (2013).

351. Kousal, B., Zahlava, J., Vejvalkova, S., Hejtmankova, M. & Liskova, P. [The molecular genetic and clinical findings in two probands with Stargardt disease]. *Cesk Slov Oftalmol* **70**, 228-233 (2014).

352. Stone, E.M.*, et al.* Clinically Focused Molecular Investigation of 1000 Consecutive Families with Inherited Retinal Disease. *Ophthalmology* **124**, 1314-1331 (2017).

353. Khan, M.*, et al.* Cost-effective molecular inversion probe-based ABCA4 sequencing reveals deep-intronic variants in Stargardt disease. *Hum Mutat* **40**, 1749-1759 (2019).

354. Sun, Z.*, et al.* Clinical and genetic analysis of the ABCA4 gene associated retinal dystrophy in a large Chinese cohort. *Exp Eye Res* **202**, 108389 (2021).

355.	Klevering, B.J*., et al.* Phenotypic spectrum of autosomal recessive cone-rod dystrophies caused by mutations in the ABCA4 (ABCR) gene. *Invest Ophthalmol Vis Sci* **43**, 1980-1985 (2002).

356.	Kitiratschky, V.B*., et al.* ABCA4 gene analysis in patients with autosomal recessive cone and cone rod dystrophies. *Eur J Hum Genet* **16**, 812-819 (2008).

357.	Cideciyan, A.V., Swider, M., Schwartz, S.B., Stone, E.M. & Jacobson, S.G. Predicting Progression of ABCA4-Associated Retinal Degenerations Based on Longitudinal Measurements of the Leading Disease Front. *Invest Ophthalmol Vis Sci* **56**, 5946-5955 (2015).

358.	Birtel, J*., et al.* Clinical and genetic characteristics of 251 consecutive patients with macular and cone/cone-rod dystrophy. *Sci Rep* **8**, 4824 (2018).

359.	Simonelli, F*., et al.* Genotype-phenotype correlation in Italian families with Stargardt disease. *Ophthalmic Res* **37**, 159-167 (2005).

360.	Duncker, T*., et al.* Correlations among near-infrared and short-wavelength autofluorescence and spectral-domain optical coherence tomography in recessive Stargardt disease. *Invest Ophthalmol Vis Sci* **55**, 8134-8143 (2014).

361.	Thiadens, A.A*., et al.* Clinical course, genetic etiology, and visual outcome in cone and cone-rod dystrophy. *Ophthalmology* **119**, 819-826 (2012).

362.	Rivera, A*., et al.* A comprehensive survey of sequence variation in the ABCA4 (ABCR) gene in Stargardt disease and age-related macular degeneration. *Am J Hum Genet* **67**, 800-813 (2000).

363.	Zolnikova, I.V*., et al.* Stargardt disease-associated mutation spectrum of a Russian Federation cohort. *Eur J Med Genet* **60**, 140-147 (2017).

364.	Marquardt, A*., et al.* Mutations in a novel gene, VMD2, encoding a protein of unknown properties cause juvenile-onset vitelliform macular dystrophy (Best's disease). *Hum Mol Genet* **7**, 1517-1525 (1998).

365.	Francois, J., De Rouck, A. & Fernandez-Sasso, D. Electro-oculography in vitelliform degeneration of the macula. *Arch Ophthalmol* **77**, 726-733 (1967).

366.	Caldwell, G.M*., et al.* Bestrophin gene mutations in patients with Best vitelliform macular dystrophy. *Genomics* **58**, 98-101 (1999).

367.	Testa, F*., et al.* A normal electro-oculography in a family affected by best disease with a novel spontaneous mutation of the BEST1 gene. *Br J Ophthalmol* **92**, 1467-1470 (2008).

368.	Lima de Carvalho, J.R., Jr*., et al.* Multimodal Imaging in Best Vitelliform Macular Dystrophy. *Invest Ophthalmol Vis Sci* **60**, 2012-2022 (2019).

369.	Duncker, T*., et al.* Quantitative fundus autofluorescence and optical coherence tomography in best vitelliform macular dystrophy. *Invest Ophthalmol Vis Sci* **55**, 1471-1482 (2014).

370.	Greenstein, V.C*., et al.* A Comparison of En Face Optical Coherence Tomography and Fundus Autofluorescence in Stargardt Disease. *Invest Ophthalmol Vis Sci* **58**, 5227-5236 (2017).

371.	Sunness, J.S., Ifrah, A., Wolf, R., Applegate, C.A. & Sparrow, J.R. Abnormal Visual Function Outside the Area of Atrophy Defined by Short-Wavelength Fundus Autofluorescence in Stargardt Disease. *Invest Ophthalmol Vis Sci* **61**, 36 (2020).

372.	Felbor, U., Schilling, H. & Weber, B.H. Adult vitelliform macular dystrophy is frequently associated with mutations in the peripherin/RDS gene. *Hum Mutat* **10**, 301-309 (1997).

373. Manes, G.*, et al.* Mutations in IMPG1 cause vitelliform macular dystrophies. *Am J Hum Genet* **93**, 571-578 (2013).
374. Meunier, I.*, et al.* Frequency and clinical pattern of vitelliform macular dystrophy caused by mutations of interphotoreceptor matrix IMPG1 and IMPG2 genes. *Ophthalmology* **121**, 2406-2414 (2014).
375. Parodi, M.B., Iacono, P. & Bandello, F. Pseudoxanthoma elasticum associated with vitelliform macular lesion. *Ophthalmic Surg Lasers Imaging Retina* **46**, 287-288 (2015).
376. Freund, K.B.*, et al.* Acquired Vitelliform Lesions: correlation of clinical findings and multiple imaging analyses. *Retina* **31**, 13-25 (2011).
377. Birtel, J.*, et al.* Peripapillary Sparing in Autosomal Recessive Bestrophinopathy. *Ophthalmol Retina* **4**, 523-529 (2020).
378. Sparrow, J.R., Parish, C.A., Hashimoto, M. & Nakanishi, K. A2E, a lipofuscin fluorophore, in human retinal pigmented epithelial cells in culture. *Invest Ophthalmol Vis Sci* **40**, 2988-2995 (1999).
379. Bakall, B.*, et al.* Enhanced accumulation of A2E in individuals homozygous or heterozygous for mutations in BEST1 (VMD2). *Exp Eye Res* **85**, 34-43 (2007).
380. O'Gorman, S., Flaherty, W.A., Fishman, G.A. & Berson, E.L. Histopathologic findings in Best's vitelliform macular dystrophy. *Arch Ophthalmol* **106**, 1261-1268 (1988).
381. Weingeist, T.A., Kobrin, J.L. & Watzke, R.C. Histopathology of Best's macular dystrophy. *Arch Ophthalmol* **100**, 1108-1114 (1982).
382. Zhang, Q., Small, K.W. & Grossniklaus, H.E. Clinicopathologic findings in Best vitelliform macular dystrophy. *Graefes Arch Clin Exp Ophthalmol* **249**, 745-751 (2011).
383. Arnold, J.J., Sarks, J.P., Killingsworth, M.C., Kettle, E.K. & Sarks, S.H. Adult vitelliform macular degeneration: a clinicopathological study. *Eye (Lond)* **17**, 717-726 (2003).
384. Chen, K.C.*, et al.* Intraretinal Hyperreflective Foci in Acquired Vitelliform Lesions of the Macula: Clinical and Histologic Study. *Am J Ophthalmol* **164**, 89-98 (2016).
385. Dubovy, S.R.*, et al.* Adult-onset foveomacular pigment epithelial dystrophy: clinicopathologic correlation of three cases. *Retina* **20**, 638-649 (2000).
386. Patrinely, J.R., Lewis, R.A. & Font, R.L. Foveomacular vitelliform dystrophy, adult type. A clinicopathologic study including electron microscopic observations. *Ophthalmology* **92**, 1712-1718 (1985).
387. Voigt, M.*, et al.* Analysis of retinal flecks in fundus flavimaculatus using high-definition spectral-domain optical coherence tomography. *Am J Ophthalmol* **150**, 330-337 (2010).
388. Dupas, B., Tadayoni, R., Erginay, A., Massin, P. & Gaudric, A. Subfoveal deposits secondary to idiopathic epiretinal membranes. *Ophthalmology* **116**, 1794-1798 (2009).
389. Constable, P.A.*, et al.* ISCEV Standard for clinical electro-oculography (2017 update). *Doc Ophthalmol* **134**, 1-9 (2017).

Appendix A: Gene-Specific Pathogenicity Score Calibration Protocol

Summary: Identifying a disease gene and determining its causality in patients can be challenging. Here we present an approach to predicting the pathogenicity of deletions and missense variants for an autosomal dominant gene. We provide online resources for identifying patients and determining constraint metrics to isolate the causal gene among several candidates encompassed in a shared region of deletion. We also provide instructions for optimizing functional annotation programs that may be otherwise inaccessible to a non-expert or novice in computational approaches.[101,102]

The text and figures from this protocol have been recently published (Lee et al. *STAR Protocols*. 2022 Feb 2;3(1):101150).

Assess population constraint metrics and mutation tolerance sensitivity scores for all genes encompassed in the overlapping region of deletions

In this step, users will use a series of web-based tools to obtain constraint metrics and dosage sensitivity scores for a group of genes within a shared region encompassed by large deletion variants identified in patients. Comparing pathogenicity metrics will assist users in distinguishing which among the deleted candidates is the most likely shared causal gene in all patients.

1. Obtain pre-calculated scores using each of the URL provided below:

    a. pLI and Missense Z-score

        i. Go to URL: https://gnomad.broadinstitute.org/

        ii. Choose ExAC in the dropdown menu

iii.   Enter single Entrez gene ID

iv.   Scores provided under "Constraint"

v.   Repeat for each respective gene

b. EvoTol (evolutionary intolerance)[293]

i.   Go to URL: http://www.evotol.co.uk/

ii.   Enter comma delimited list of Entrez gene ID for all genes of interest

iii.   "EvoTol percentile" scores will be tabulated and graphed. (Note: Score may not be available for genes that are not expressed above)

c. RVIS (residual variation intolerance score)[294]

i.   Go to URL: http://genic-intolerance.org/

ii.   Enter comma delimited list of Entrez gene ID for all genes of interest

iii.   RVIS scores provided with percentiles in parentheses

Each gene can be assessed as to whether its score falls within the predicted range for pathogenicity. For instance, genes with pLI scores ≥ 0.9 are interpreted as extremely intolerant to loss-of-function [295,296]. A description of each score and recommended pathogenicity thresholds are provided in **Appendix Table A1**. In addition, the scores for all genes can be directly compared to one another and ranked relative to each other, especially in situations where multiple genes are predicted to be pathogenic.

Returning to the example in Gennarino et al. (2018), *PUM1* was determined to be the most likely

causal gene within the deleted regions (**Appendix Figure A2**) because (1) its loss of function

measures consistently surpassed the pathogenicity threshold, and, more importantly, (2) its effect

was greater than that of any of the other 15 deleted genes (**Appendix Figure 1B-1F**).

Use ANNOVAR and dbnsfp (version 4.2a) dataset to annotate genetic variants with

pathogenicity scores

In this step, PC/Apple iOS users will use a Perl programming language based genetic annotation

program, ANNOVAR[84] (through an installed or intrinsic Perl interpreter), to obtain functional

and pathogenic predictions for a list of variants found in patients. Rather than downloading the

entire dataset of numerous functional prediction scores for annotation, which would require local

disk space beyond the average capacity of individual work station (>100 TB), annotations will

come from a "lightweight" but comprehensive dataset, dbnsfa [181,297] (~50-60 Gb), containing all

potential nonsynonymous and splice-site SNVs in the human genome.

Note: the scripts provided for annotating the p.(Arg1147Trp), p.(Arg1139Trp), p.(Thr1035Ser)

and gnomAD singleton variants are for demonstration purposes. Users should modify and adapt

this script according to their own needs.

2. Install (or activate) Perl language interpreter
   a. Windows/PC users:
      i. Visit the URL: https://strawberryperl.com/ and install program according to
         operating system (64 bit or 32 bit)

ii. Verify installation:

```
perl -v
```

b. Apple iOS users: a Perl interpreter comes pre-installed. Perl scripts can run by

directly invoking the interpreter:

```
perl myprogram.pl
```

3. Install ANNOVAR annotation program

   a. Visit the URL:

      https://www.openbioinformatics.org/annovar/annovar_download_form.php

   b. Submit registration form and a download link will be emailed to you

   c. Unzip file into your base directory

4. Download dbnsfp 4.2a dataset into the humandb folder

   a. Open the command (CMD) prompt

   b. Set working directory to *annovar* program folder

   c. Running the following command will download 2 zipped folders:

      i. hg19_dbnsfp42a.txt.gz

      ii. hg19_dbnsfp42a.txt.idx.gz

```
perl annotate_variation.pl -buildver hg19 -downdb -webfrom annovar
```

Critical: the hg19 assembly was used here in accordance with the procedures in Gennarino et al.

the -buildver argument in annotate_variation.pl command can be changed to

`hg18` and `hg38` accordingly.

    d. Manually unzip folder using any generic decompression program (see **Key Resources Table**)

    e. Move the respective files in each folder (*hg19_dbnsfp42a.txt* and *hg19_dbnsfp42a.txt.idx*) into the *humandb* folder

5. Create .txt file listing the genomic location and nucleotide change of each patient variant. Required (tab-separated) column values for each variant are: <u>chromosome number</u>, <u>start position</u>, <u>end position</u>, <u>reference nucleotide</u>, *alternate nucleotide*. Headers should not be included in the input file. Depending on the number of variants, this may be done manually using any text editing software (see **Key Resources Table**).

       Note: ensure that the coordinates are consistent with the reference genome build of the annotation data—which in this case is hg19.

6. Place the .txt input file into the *annovar* folder. (For this example, we will name the file "*PUM1_subject_variants.txt*" (see **Data S1**)

7. Run the table_annovar.pl script to annotate this file:

```
perl table_annovar.pl PUM1_subject_variants.txt humandb/ -buildver hg19 -
out PUM1_subject_variants_output -remove -protocol refGene,dbnsfp42a -
operation gx,f -nastring . -csvout -polish -xref example/gene_xref.txt
```

Critical: the hg19 assembly was used here in accordance with the procedures in Gennarino et al. the `-buildver` argument in `table_annovar.pl` command can be changed to `hg18` and `hg38` accordingly.

a. The file "*PUM1_subject_variants_output.hg19_multianno.csv*" (**Data S2**) will be

generated in the user's working directory

Extract and annotate singleton missense variants in *PUM1* from the healthy population for comparison to patient variant scores

8. Extract a list of all singleton variants (allele count = 1) from population database

    a. Go to URL:

    https://gnomad.broadinstitute.org/gene/ENSG00000134644?dataset=gnomad_r2_1

    b. Download a .csv file of all *PUM1* variants by clicking the "Export variants to

    CSV" button on this URL page. (**Note**: downloaded file name is date- and time-

    stamped and therefore will vary for users. For this protocol, a sample file named:

    "*gnomAD_v2.1.1_ENSG00000134644_2021_10_17_00_08_13.csv*" will be used

    and is provided as **Data S3**)

    c. Singleton missense variants can be manually extracted from this file using any

    spreadsheet application (Excel) or text editor by setting the following column

    filters:

        i. VEP Annotation = "missense_variant"

        ii. Allele count = 1

9. After filtering, the values from the Chromosome, Position, Reference and Alternate

    columns in this file can then be extracted to create a (table_annovar.pl) input file in the

format described above.

10. Alternatively, this entire process can be performed in a single step using the statistical

    computing program R and the scripts provided:

    a. Install the latest version of R and RStudio (version 4.1.1 recommended, see Key

       Resources Table) and open RStudio application.

    b. Import the .csv file into RStudio in one of two ways:

       i. File>Import Database>From Text(base)…

       ii. Run the following read.csv command (**Note**: adjust the relative path for

           "*read.csv*" function to the directory containing the

           "*gnomAD_v2.1.1_ENSG00000134644_2021_10_17_00_08_13.csv file*",

           Data S3):

```
gnomAD_v2.1.1_ENSG00000134644_2021_10_17_00_08_13 <-
read.csv("~/gnomAD_v2.1.1_ENSG00000134644_2021_10_17_00_08_13.csv")
```

    c. The following script will extract all singleton missense variants and generate a .txt

       input file (in this case, named

       "*PUM1_gnomad_singleton_missense_annovar_input.txt*") (**Data S4**) in working

       directory of RStudio ready for annotation with ANNOVAR:

```
PUM1_gnomad_singleton_missense_annovar_input <-
    subset(
        gnomAD_v2.1.1_ENSG00000134644_2021_10_17_00_08_13,
        Allele.Count == 1 & VEP.Annotation == "missense_variant",
        select = c(
            Chromosome,
            Position,
            Position,
            Reference,
            Alternate))
write.table(
            PUM1_gnomad_singleton_missense_annovar_input,
            file = "PUM1_gnomad_singleton_missense_annovar_input.txt",
            sep = "\t",
            row.names = FALSE,
            col.names = FALSE,
            quote = FALSE)
```

    d.   The file "*PUM1_gnomad_singleton_missense_annovar_input.txt*" (**Data S4**) will

be generated in the working directory

    e.   Place this file in the "*annovar"* folder

11. Annotate this file with the following table_annovar.pl command:

```
perl table_annovar.pl PUM1_gnomad_singleton_missense_annovar_input.txt
humandb/ -buildver hg19 -out PUM1_gnomad_singleton_missense_annovar_output
-remove -protocol refGene,dbnsfp42a -operation gx,f -nastring . -csvout -
polish -xref example/gene_xref.txt
```

    a.   The file

"*PUM1_gnomad_singleton_missense_annovar_output.hg19_multianno.csv"*

(**Data S5**) will be generated in the user's working directory

Quantification and Statistical Analysis

Functional predictions, based simply on sequencing results from an individual, can help distinguish likely causal variants from non-causal (benign) variants. The current protocol provides users access to these resources and bioinformatic tools to gauge the pathogenicity of variants among a cohort of patients with shared clinical features.

All functional prediction scores have recommended pathogenicity thresholds (i.e., numerical cut-offs beyond which variants are predicted to be likely pathogenic or benign). The manner in which these thresholds are defined varies according to each algorithm. Methods such as REVEL[91], are random forest classification algorithms trained to distinguish variants as pathogenic or benign based on data deposited in known clinical databases (ClinVar)[298]. Many algorithms base their pathogenicity scores on loss-of-function variants, which of course is not the only mutation type. Statistically comparing the score of one or more variants of an autosomal dominant gene to the confidence interval or inter-quartile range of the distribution of scores of singleton variants from the (presumably unaffected) general population provides a conservative measure for users to assess pathogenicity.

Limitations

Computational predictions, of course, have their limits, as they represent statistical simulations of biological effects. Different algorithms can produce contradictory predictions for the same variant. Predictions therefore need to be considered in conjunction with other factors such as functional studies and the nature of the phenotype. Please refer to de Prisco et al. STAR Protocols (submitted, STAR-PROTOCOLS-D-21-00615R1) for an approach to functional studies.

Appendix Table A1. Description and recommended pathogenicity thresholds for gene-based constraint metrics and intolerance scores for the assessment of multiple genes encompassed with large copy-number variants in patients.

Algorithm	Description	Score	Threshold interpretation	Reference
pLI	Dichotomous metric that reflects the probability of loss of function intolerance based on predicted protein truncating variation: nonsense, splice acceptor and splice donor variation	0 to 1	pLI ≥ 0.9 = Extremely intolerant	(Samocha et al., 2014) (Lek et al., 2016)
Missense Z-score	Standard deviation of the # of <u>observed</u> rare (<1% MAF) missense SNP's from the mean of the predicted number of <u>expected</u> rare (<1% MAF) missense SNP's for a given gene. (Synonymous variants typically have Z-score close to 0 and o/e close to 1)	-5 to 5	Z-score > 0 = More Intolerant Z-score < 0 = Less intolerant	(Samocha et al., 2014) (Lek et al., 2016)
EvoTol percentile	Percentile ranking of a gene's evolutionary intolerance relative to other genes based on the number of damaging versus non-damaging variants in dbSNP.	Percentile (priority ranking)	Top 25th percentile = Intolerant Top 1 percentile = Most intolerant	(Rackham et al., 2015)
RVIS percentile	Percentile ranking of a gene's intolerance to functional variation relative to other genes based on observed versus expected frequency of loss-of-function variants from the NHLBI-ESP6500 data set in ExAC (release 0.3).	Percentile (priority ranking)	Bottom 25th percentile = Intolerant *(Genes are ranked from most to least intolerant)*	(Petrovski et al., 2013)
pcGERP percentile	Percentile estimate reflecting how conserved the protein-coding sequence of a gene relative to other genes	Percentile estimate	Lower percentile = More intolerance (increased conservation) *(Genes are ranked from most to least intolerant)*	(Davydov et al., 2010) (Cooper et al., 2005) (Petrovski et al., 2015)

Abbreviation: pLI, probability of being Loss-of-function Intolerant; EvoTol, Evolutionary Intolerance; RVIS, Residual Variation Intolerance Score; pcGERP, protein-coding Genomic Evolutionary Rate Profiling; MAF, minor allele frequency;
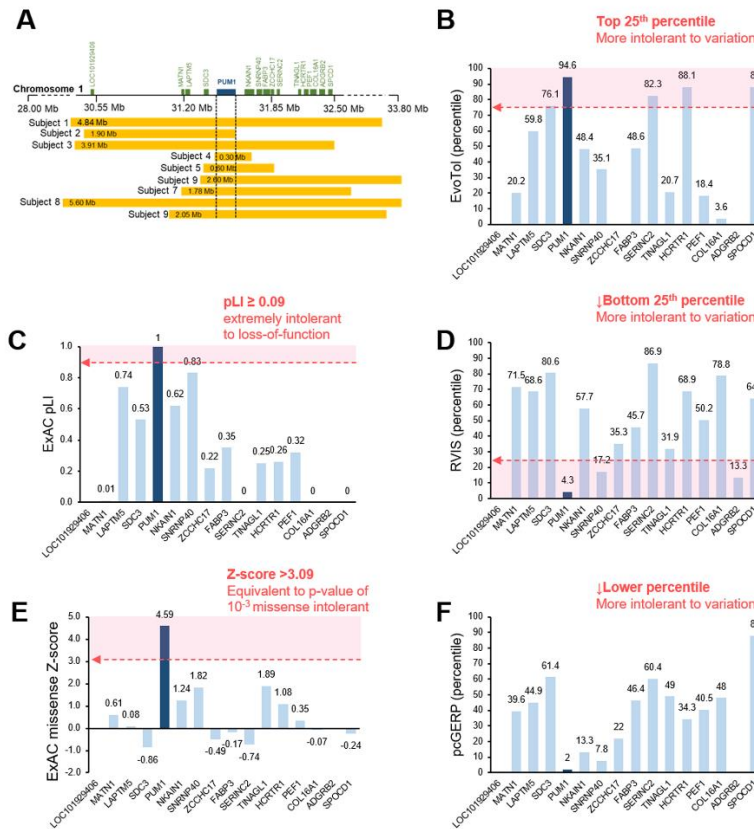
Appendix Figure A1: Comparison of constraint metrics and mutational intolerance scores amongst genes encompassed in the overlapping regions of deletions. (A) Alignment of deletions from nine patients with NDD spanning the PUM1 and other loci on chromosome 1p35.2; Mb, megabases. Dashed lines indicate the minimal region spanning PUM1, figure adapted from (Gennarino et al., 2018). Bar graphs of constraint metrics and variant intolerance algorithms: (B) EvoTol (C) ExAC pLI, (D) RVIS (E) ExAC missense Z-score and (F) pcGERP. The dotted pink arrow (and shaded region) delineates the recommended threshold in which variants are predicted to be deleterious.
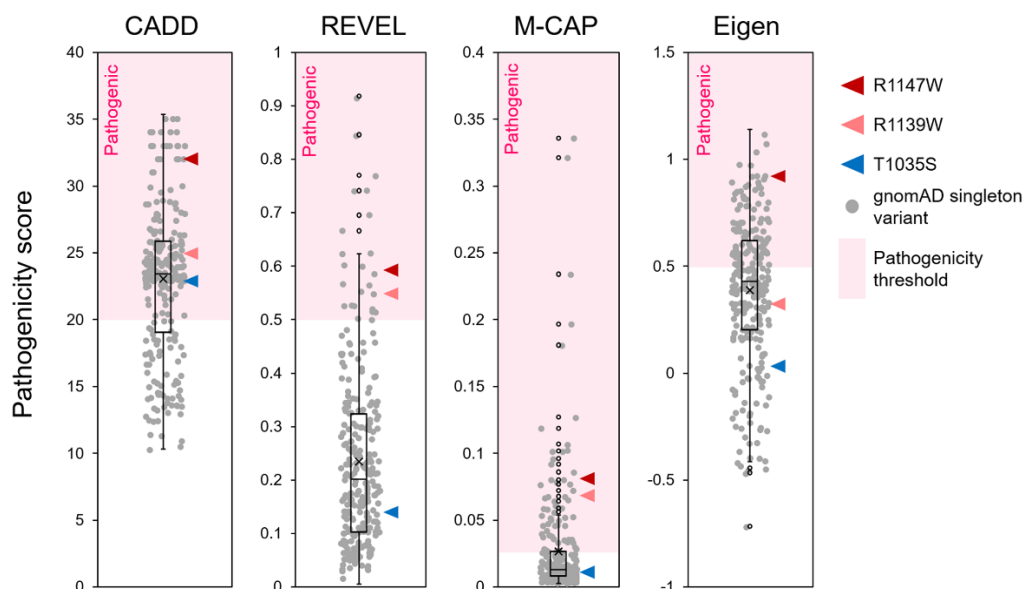
Appendix Figure A2: Re-calibrated pathogenicity thresholds for CADDv1.6, REVEL, M-CAP and Eigen based on the distribution of singleton missense variants in gnomAD.
Pathogenicity scores for disease-associated *PUM1* variants. CADD-PHRED (v1.6), REVEL, M-CAP and Eigen pathogenicity scores of *PUM1* variants, p.(Arg1147Trp) (red arrowhead), p.(Arg1139Trp) (pink arrowhead) and p.(Thr1035Ser) (blue arrowhead), relative to the distribution (box plot) of PUM1 singleton (ultra-rare) variant scores found in gnomAD (gray circles). The dotted pink arrow (and shaded region) delineates the recommended threshold in which variants are predicted to be deleterious.

Appendix B: Supplemental Figures and Tables

Supplemental Figure 1 (*Chapter 3*): Pedigrees of unaffected homozygous p.(Gly1961Glu) cases in two ABCA4 families.[ll]

Unaffected homozygous p.(Gly1961Glu) cases
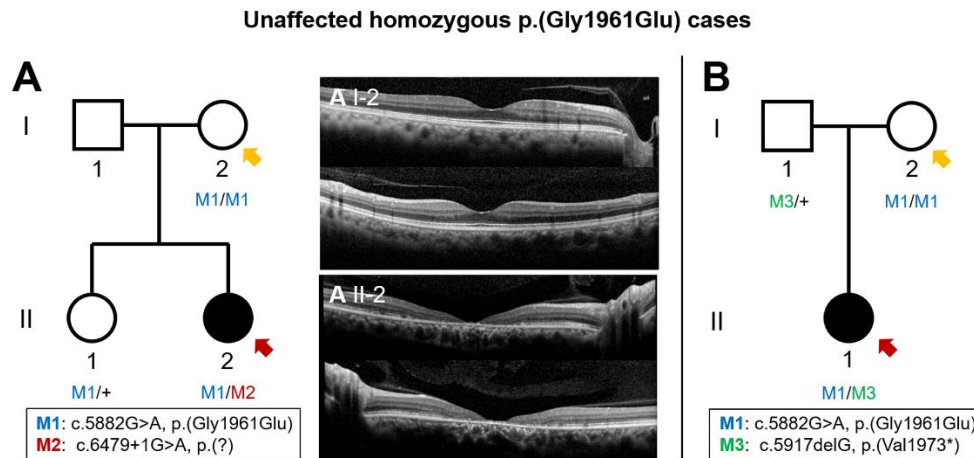
[ll] (A) The proband in Family A (red arrow), a 31-year old woman, was bi-allelic for p.(Gly1961Glu) and a canonical splice site variant, c.6479+1G>A (SpliceAI Δscore = 0.94, donor loss). Spectral domain-optical coherence tomography (SD-OCT) revealed a lesion of outer retinal atrophy in the central macula (SD-OCT, A II-2) although her mother (yellow arrow), a 56-year-old woman homozygous for p.(Gly1961Glu), reported no history of ocular symptoms and exhibited no apparent disease changes in the fundus at the time of examination. (B) The proband (red arrow) in Family B was a 12-year-old girl recently diagnosed with Stargardt disease. The proband was compound heterozygous for p.(Gly1961Glu) and a single nucleotide deletion, c.5917delG, that interrupts the reading frame and results in a premature truncation, p.(V1973*). The mother (yellow arrow) was confirmed to be homozygous for p.(Gly1961Glu) but reported no history of ocular symptoms and likewise, exhibited no apparent disease changes on fundus examination at 50 years of age.

Supplemental Figure 2 (*Chapter 5*): Retinal phenotype of the rare hypomorph p.(Arg2030Gln) variant of ABCA4 disease.[mm]

Supplemental Figure 3 (*Chapter 5*): Retinal phenotype of the rare hypomorph p.(Ile1562Thr) variant of ABCA4 disease.[nn]

Supplemental Figure 4 (*Chapter 8*): Gene-disease heat map of the recurrent disease-associated genes encompassed by large genomic deletions in patients with PUM1-associated neurodevelopmental disease.[oo]
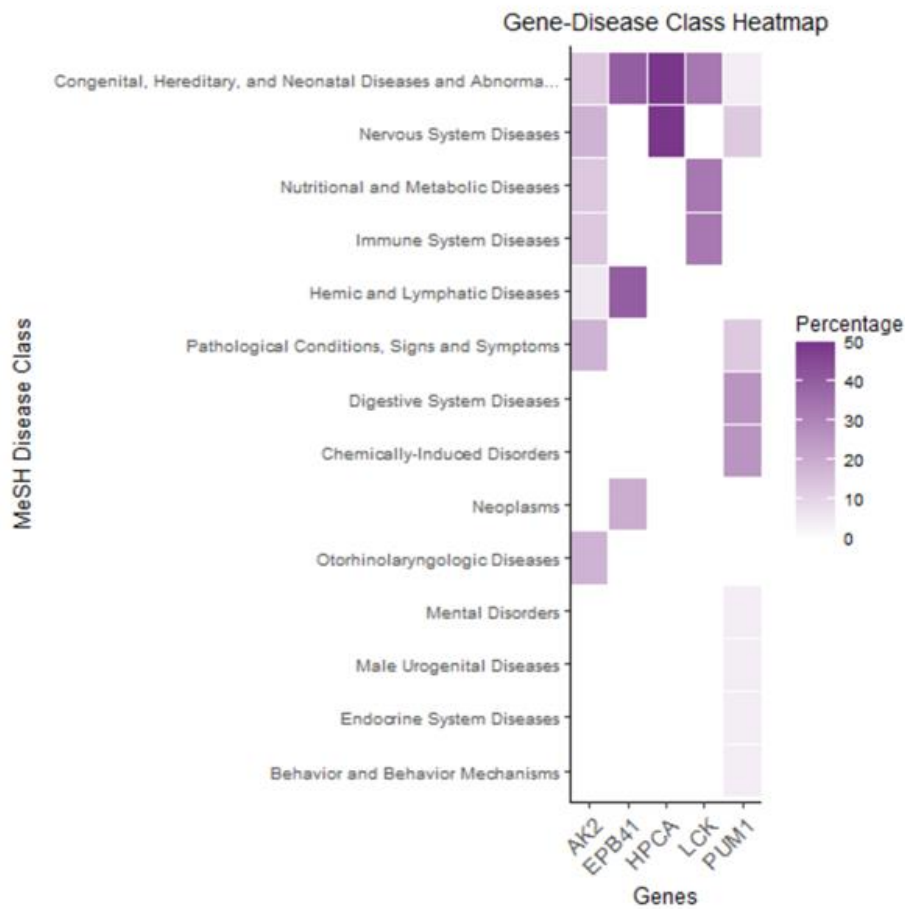
[oo] Diseases on the y-axis are grouped by MeSH disease classes. Color scale is proportional to the percentage of diseases in each MeSH class.

174

Supplemental Figure 5 (*Chapter 8*): Predicted cryptic splice donor in Exon 2 of *PUM1*.[pp]

WT sequence	Nucleotides	Position	SpliceRover score
Canonical splice donor	GT	+1	0.980828
Cryptic splice donor	GT	-96	0.9959

c.363+1G>A	Nucleotides	Position	SpliceRover score
Canonical splice donor	AT	+1	0
Cryptic splice donor	GT	-96	0.995306
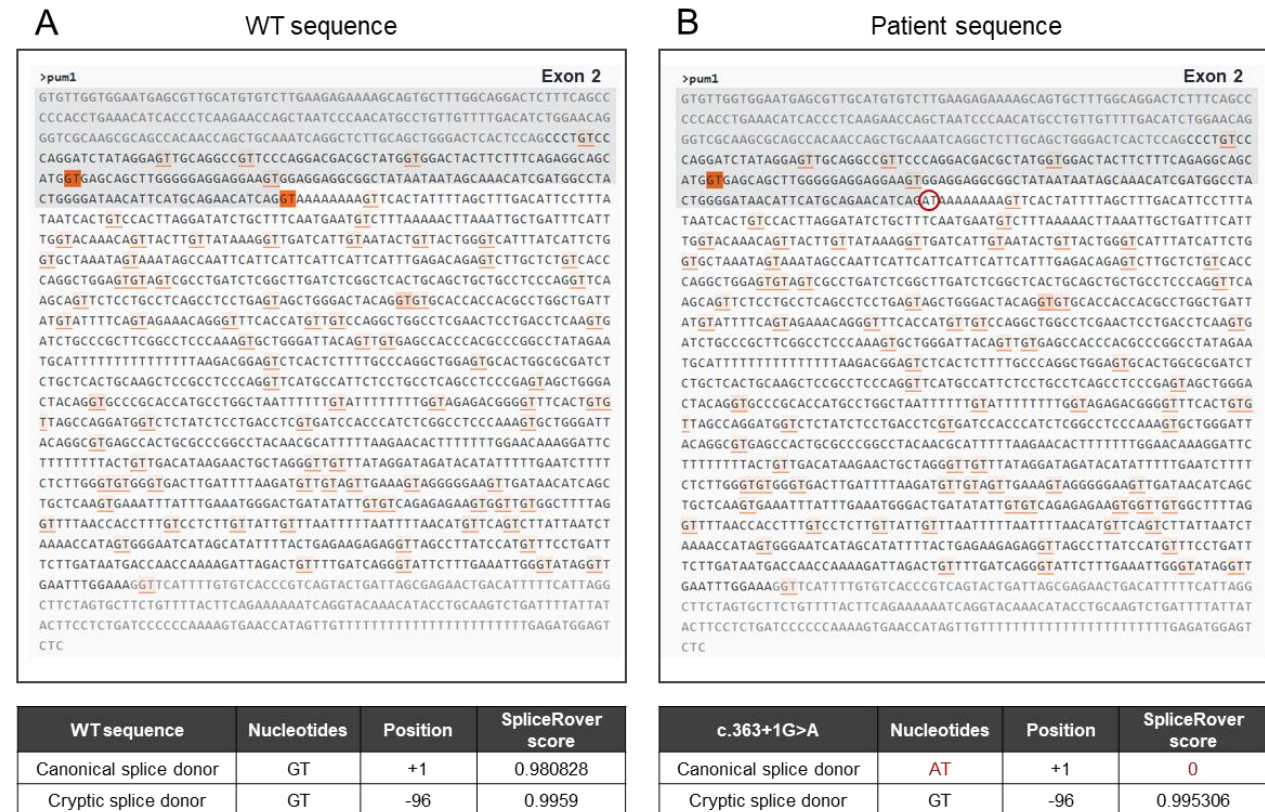
[pp] (A) Partial wild-type (WT) *PUM1* sequence of Exon 2 (gray shading) and Intron 2 showing the canonical splice donor (c.363+1G) at the 3' end of Exon 2 (SpliceRover score = 99.6%) and the cryptic splice donor (SpliceRover score = 98.1%). (B) Partial *PUM1* sequence showing the loss of the canonical splice donor due to the c.363+1G>A substitution (circled in red) and the increased strength of the upstream cryptic donor (SpliceRover score = 99.5%).

175

Supplemental Table 1 (*Chapter 3*): Cohort of 150 STGD1 patients with homozygous and compound heterozygous p.(Gly1961Glu) alleles.

Patient #	c.769-784 C<T	Allele 1		Allele 2		2nd allele deleterious/ expressed with hypomorph
		cDNA	Protein	cDNA	Protein	
1	+/-	c.5882G>A	p.(Gly1961Glu)	c.5882G>A	p.(Gly1961Glu)	
2	+/-	c.5882G>A	p.(Gly1961Glu)	c.5882G>A	p.(Gly1961Glu)	
3	+/-	c.5882G>A	p.(Gly1961Glu)	c.5882G>A	p.(Gly1961Glu)	
4	+/-	c.5882G>A	p.(Gly1961Glu)	c.5882G>A	p.(Gly1961Glu)	
5	+/+	c.5882G>A	p.(Gly1961Glu)	c.5882G>A	p.(Gly1961Glu)	
6	+/-	c.5882G>A	p.(Gly1961Glu)	c.[3758C>T;5882G>A]	p.([Thr1253Met;Gly1961Glu])	
7	-/-	c.[2549A>G;5882G>A]	p.([Tyr850Cys;Gly1961Glu])	c.[2549A>G;5882G>A]	p.([Tyr850Cys;Gly1961Glu])	
8	-/-	c.[5242G>A;5882G>A]	p.([Gly1748Arg;Gly1961Glu])	c.5882G>A	p.(Gly1961Glu)	
9	-/-	c.5882G>A	p.(Gly1961Glu)	c.5882G>A	p.(Gly1961Glu)	
10	-/-	c.5882G>A	p.(Gly1961Glu)	c.5882G>A	p.(Gly1961Glu)	
11	+/-	c.5882G>A	p.(Gly1961Glu)	c.4139C>T	p.(Pro1380Leu)	
12	+/-	c.5882G>A	p.(Gly1961Glu)	c.4139C>T	p.(Pro1380Leu)	
13	+/-	c.5882G>A	p.(Gly1961Glu)	c.4139C>T	p.(Pro1380Leu)	
14	+/-	c.5882G>A	p.(Gly1961Glu)	c.4139C>T	p.(Pro1380Leu)	
15	+/-	c.5882G>A	p.(Gly1961Glu)	c.4139C>T	p.(Pro1380Leu)	
16	+/-	c.5882G>A	p.(Gly1961Glu)	c.4139C>T	p.(Pro1380Leu)	
17	-/-	c.[3758C>T;5882G>A]	p.([Thr1253Met;Gly1961Glu])	c.4139C>T	p.(Pro1380Leu)	
18	-/-	c.[3758C>T;5882G>A]	p.([Thr1253Met;Gly1961Glu])	c.4139C>T	p.(Pro1380Leu)	
19	-/-	c.[3758C>T;5882G>A]	p.([Thr1253Met;Gly1961Glu])	c.4139C>T	p.(Pro1380Leu)	
20	-/-	c.5882G>A	p.(Gly1961Glu)	c.4139C>T	p.(Pro1380Leu)	
21	-/-	c.5882G>A	p.(Gly1961Glu)	c.4139C>T	p.(Pro1380Leu)	
22	-/-	c.5882G>A	p.(Gly1961Glu)	c.4139C>T	p.(Pro1380Leu)	
23	-/-	c.5882G>A	p.(Gly1961Glu)	c.4139C>T	p.(Pro1380Leu)	
24	-/-	c.5882G>A	p.(Gly1961Glu)	c.4139C>T	p.(Pro1380Leu)	
25	-/-	c.5882G>A	p.(Gly1961Glu)	c.4139C>T	p.(Pro1380Leu)	
26	-/-	c.5882G>A	p.(Gly1961Glu)	c.4139C>T	p.(Pro1380Leu)	
27	-/-	c.5882G>A	p.(Gly1961Glu)	c.4139C>T	p.(Pro1380Leu)	
28	+/-	c.5882G>A	p.(Gly1961Glu)	c.5603A>T	p.(Asn1868Ile)	
29	+/-	c.5882G>A	p.(Gly1961Glu)	c.4070C>T	p.(Ala1357Val)	
30	+/-	c.5882G>A	p.(Gly1961Glu)	c.6448T>C	p.(Cys2150Arg)	
31	+/-	c.5882G>A	p.(Gly1961Glu)	c.286A>G	p.(Asn96Asp)	
32	+/-	c.5882G>A	p.(Gly1961Glu)	c.183G>A	p.(Met61Ile)	
33	+/-	c.5882G>A	p.(Gly1961Glu)	c.[4594G>A;5603A>T]	p.([Asp1532Asn;Asn1868Ile])	
34	-/-	c.[3758C>T;5882G>A]	p.([Thr1253Met;Gly1961Glu])	c.[2588G>C;5603A>T]	p.([Gly863Ala;Asn1868Ile])	
35	-/-	c.[3758C>T;5882G>A]	p.([Thr1253Met;Gly1961Glu])	c.1938-619A>G	p.([Phe647Alafs*22,Phe647Serfs*22,=])	
36	-/-	c.5882G>A	p.(Gly1961Glu)	c.[4594G>A;5603A>T]	p.([Asp1532Asn;Asn1868Ile])	
37	-/-	c.5882G>A	p.(Gly1961Glu)	c.1908G>T	p.(Gln636His)	
38	-/-	c.5882G>A	p.(Gly1961Glu)	c.5087G>A	p.(Ser1696Asn)	
39	-/-	c.5882G>A	p.(Gly1961Glu)	c.1208A>T	p.(Asp403Val)	
40	-/-	c.5882G>A	p.(Gly1961Glu)	c.1964T>G	p.(Phe655Cys)	
41	-/-	c.5882G>A	p.(Gly1961Glu)	c.2453G>T	p.(Gly818Val)	
42	-/-	c.5882G>A	p.(Gly1961Glu)	c.5114G>T	p.(Arg1705Leu)	
43	-/-	c.5882G>A	p.(Gly1961Glu)	c.3364G>A	p.(Glu1122Lys)	
44	-/-	c.5882G>A	p.(Gly1961Glu)	c.160+5G>C	p.([=,Ile23Alafs*24,His55Asnfs*63])	
45	-/-	c.5882G>A	p.(Gly1961Glu)	c.2919-826T>A	p.([Leu973Phefs*1,=])	

46	-/-	c.5882G>A	p.(Gly1961Glu)	c.634C>T	p.(Arg212Cys)	
47	-/-	c.5882G>A	p.(Gly1961Glu)	c.[4216C>T;5603A>T]	p.([His1406Tyr;Asn1868Ile])	
48	-/-	c.5882G>A	p.(Gly1961Glu)	c.2947A>G	p.(Thr983Ala)	
49	-/-	c.5882G>A	p.(Gly1961Glu)	c.3385C>T	p.(Arg1129Cys)	
50	-/-	c.5882G>A	p.(Gly1961Glu)	c.6449G>A	p.(Cys2150Tyr)	
51	-/-	c.5882G>A	p.(Gly1961Glu)	c.[4594G>A;5603A>T]	p.([Asp1532Asn;Asn1868Ile])	
52	-/-	c.5882G>A	p.(Gly1961Glu)	c.6098T>A	p.(Leu2033His)	
53	-/-	c.5882G>A	p.(Gly1961Glu)	c.1819G>T	p.(Gly607Trp)	
54	-/-	c.5882G>A	p.(Gly1961Glu)	c.[4216C>T;5603A>T]	p.([His1406Tyr;Asn1868Ile])	
55	-/-	c.5882G>A	p.(Gly1961Glu)	c.6079C>T	p.(Leu2027Phe)	
56	-/-	c.5882G>A	p.(Gly1961Glu)	c.634C>T	p.(Arg212Cys)	
57	-/-	c.5882G>A	p.(Gly1961Glu)	c.634C>T	p.(Arg212Cys)	
58	-/-	c.5882G>A	p.(Gly1961Glu)	c.71G>A(;)6079C>T	p.(Arg24His(;)Leu2027Phe)	
59	-/-	c.5882G>A	p.(Gly1961Glu)	c.3407G>T	p.(Gly1136Val)	
60	-/-	c.5882G>A	p.(Gly1961Glu)	c.634C>T	p.(Arg212Cys)	
61	-/-	c.5882G>A	p.(Gly1961Glu)	c.4793C>A	p.(Ala1598Asp)	
62	-/-	c.5882G>A	p.(Gly1961Glu)	c.641T>A	p.(Ile214Asn)	
63	-/-	c.5882G>A	p.(Gly1961Glu)	c.[2588G>C;5603A>T]	p.([Gly863Ala;Asn1868Ile])	
64	-/-	c.5882G>A	p.(Gly1961Glu)	c.3065A>G	p.(Glu1022Gly)	
65	-/-	c.5882G>A	p.(Gly1961Glu)	c.4217A>G	p.(His1406Arg)	
66	+/-	c.5882G>A	p.(Gly1961Glu)	c.[1A>G;6089G>A]	p.(?)	yes
67	+/-	c.5882G>A	p.(Gly1961Glu)	c.3050+5G>A	p.(Leu973_His1017delinsPhe)	yes
68	+/-	c.5882G>A	p.(Gly1961Glu)	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	yes
69	+/-	c.5882G>A	p.(Gly1961Glu)	c.4841_4843del	p.(Asn1614del)	yes
70	+/-	c.5882G>A	p.(Gly1961Glu)	c.161G>A	p.(Cys54Tyr)	yes
71	-/-	c.5882G>A	p.(Gly1961Glu)	c.3322C>T	p.(Arg1108Cys)	yes
72	-/-	c.5882G>A	p.(Gly1961Glu)	c.6342G>A	p.(?)	yes
73	-/-	c.5882G>A	p.(Gly1961Glu)	c.2382+1G>A	p.(?)	yes
74	-/-	c.5882G>A	p.(Gly1961Glu)	c.5316G>A	p.(Trp1772*)	yes
75	-/-	c.5882G>A	p.(Gly1961Glu)	c.161G>A	p.(Cys54Tyr)	yes
76	-/-	c.5882G>A	p.(Gly1961Glu)	c.[5461-10T>C;5603A>T]	p.(Thr1821Aspfs*6)	yes
77	-/-	c.5882G>A	p.(Gly1961Glu)	c.5917del	p.(Val1973*)	yes
78	-/-	c.5882G>A	p.(Gly1961Glu)	c.3322C>T	p.(Arg1108Cys)	yes
79	-/-	c.5882G>A	p.(Gly1961Glu)	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	yes
80	-/-	c.5882G>A	p.(Gly1961Glu)	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	yes
81	-/-	c.5882G>A	p.(Gly1961Glu)	c.564del	p.(Glu189Serfs*12)	yes
82	-/-	c.5882G>A	p.(Gly1961Glu)	c.4234C>T	p.(Gln1412*)	yes
83	-/-	c.5882G>A	p.(Gly1961Glu)	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	yes
84	-/-	c.5882G>A	p.(Gly1961Glu)	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	yes
85	-/-	c.5882G>A	p.(Gly1961Glu)	c.45G>A	p.(Trp15*)	yes
86	-/-	c.5882G>A	p.(Gly1961Glu)	c.6229C>T	p.(Arg2077Trp)	yes
87	-/-	c.5882G>A	p.(Gly1961Glu)	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	yes
88	-/-	c.5882G>A	p.(Gly1961Glu)	c.4734del	p.(Leu1580*)	yes
89	-/-	c.5882G>A	p.(Gly1961Glu)	c.4537dup	p.(Gln1513Profs*42)	yes
90	-/-	c.5882G>A	p.(Gly1961Glu)	c.1622T>C	p.(Leu541Pro)	yes
91	-/-	c.5882G>A	p.(Gly1961Glu)	c.[5461-10T>C;5603A>T]	p.(Thr1821Aspfs*6)	yes
92	-/-	c.5882G>A	p.(Gly1961Glu)	c.4539+1G>T	p.(?)	yes
93	-/-	c.5882G>A	p.(Gly1961Glu)	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	yes
94	-/-	c.5882G>A	p.(Gly1961Glu)	c.454C>T	p.(Arg152*)	yes

177

95	-/-	c.5882G>A	p.(Gly1961Glu)	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	yes
96	-/-	c.5882G>A	p.(Gly1961Glu)	c.5196+1056A>G	p.(Met1733Valfs*2)	yes
97	-/-	c.5882G>A	p.(Gly1961Glu)	c.[5461-10T>C;5603A>T]	p.(Thr1821Aspfs*6)	yes
98	-/-	c.5882G>A	p.(Gly1961Glu)	c.3050+5G>A	p.(Leu973_His1017delinsPhe)	yes
99	-/-	c.5882G>A	p.(Gly1961Glu)	c.1804C>T	p.(Arg602Trp)	yes
100	-/-	c.5882G>A	p.(Gly1961Glu)	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	yes
101	-/-	c.5882G>A	p.(Gly1961Glu)	c.4234C>T	p.(Gln1412*)	yes
102	-/-	c.5882G>A	p.(Gly1961Glu)	c.6005+1G>T	p.(?)	yes
103	-/-	c.5882G>A	p.(Gly1961Glu)	c.6387-1G>T	p.(?)	yes
104	-/-	c.5882G>A	p.(Gly1961Glu)	c.[3608G>A;4537del]	p.([Gly1203Glu;Gln1513Argfs*13])	yes
105	-/-	c.5882G>A	p.(Gly1961Glu)	c.1988G>A	p.(Trp663*)	yes
106	-/-	c.5882G>A	p.(Gly1961Glu)	c.160+2T>C	p.(?)	yes
107	-/-	c.5882G>A	p.(Gly1961Glu)	c.5018+2T>C	p.(?)	yes
108	-/-	c.5882G>A	p.(Gly1961Glu)	c.[4469G>A;5603A>T]	p.([Cys1490Tyr;Asn1868Ile])	yes
109	-/-	c.5882G>A	p.(Gly1961Glu)	c.[1140T>A;1622T>C;4328G>A]	p.([Asn380Lys;Leu541Pro;Arg1443His])	yes
110	-/-	c.5882G>A	p.(Gly1961Glu)	c.1100-1G>A	p.(?)	yes
111	-/-	c.5882G>A	p.(Gly1961Glu)	c.4540-2A>G	p.(?)	yes
112	-/-	c.5882G>A	p.(Gly1961Glu)	c.6729+5_+19del	p.(Phe2161Cysfs*3)	yes
113	-/-	c.[2549A>G;5882G>A]	p.([Tyr850Cys;Gly1961Glu])	c.[5461-10T>C;5603A>T]	p.(Thr1821Aspfs*6)	yes
114	-/-	c.5882G>A	p.(Gly1961Glu)	c.[4469G>A;5603A>T]	p.([Cys1490Tyr;Asn1868Ile])	yes
115	-/-	c.5882G>A	p.(Gly1961Glu)	c.6148-698_ c.6670del	p.(?)	yes
116	-/-	c.5882G>A	p.(Gly1961Glu)	c.6119G>A	p.(Arg2040*)	yes
117	-/-	c.[3758C>T;5882G>A]	p.([Thr1253Met;Gly1961Glu])	c.3050+5G>A	p.(Leu973_His1017delinsPhe)	yes
118	-/-	c.5882G>A	p.(Gly1961Glu)	c.[5461-10T>C;5603A>T]	p.(Thr1821Aspfs*6)	yes
119	-/-	c.5882G>A	p.(Gly1961Glu)	c.4462T>C	p.(Cys1488Arg)	yes
120	-/-	c.[3758C>T;5882G>A]	p.([Thr1253Met;Gly1961Glu])	c.3007C>T	p.(Gln1003*)	yes
121	-/-	c.5882G>A	p.(Gly1961Glu)	c.2191_2199del	p.(Phe731_Leu733del)	yes
122	-/-	c.5882G>A	p.(Gly1961Glu)	c.6479+1G>A	p.(?)	yes
123	-/-	c.5882G>A	p.(Gly1961Glu)	c.1957C>T	p.(Arg653Cys)	yes
124	-/-	c.5882G>A	p.(Gly1961Glu)	c.67-1845_c.160+345del	p.(?)	yes
125	-/-	c.5882G>A	p.(Gly1961Glu)	c.6729+5_+19del	p.(Phe2161Cysfs*3)	yes
126	-/-	c.5882G>A	p.(Gly1961Glu)	c.5917del	p.(Val1973*)	yes
127	-/-	c.5882G>A	p.(Gly1961Glu)	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	yes
128	-/-	c.5882G>A	p.(Gly1961Glu)	c.3995_4018dup(;)2791G>A(;)4000C>G	p.(Gln1332_Cys1339dup)(;)(Val931Met)(;)(Pro1334Ala)	yes
129	-/-	c.5882G>A	p.(Gly1961Glu)	c.5196+1056A>G	p.(Met1733Valfs*2)	yes
130	-/-	c.5882G>A	p.(Gly1961Glu)	c.5318C>T	p.(Ala1773Val)	yes
131	-/-	c.5882G>A	p.(Gly1961Glu)	c.3050+5G>A	p.(Leu973_His1017delinsPhe)	yes
132	-/-	c.5882G>A	p.(Gly1961Glu)	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	yes
133	-/-	c.5882G>A	p.(Gly1961Glu)	c.1622T>C	p.(Leu541Pro)	yes
134	-/-	c.5882G>A	p.(Gly1961Glu)	c.5196+1G>A	p.(?)	yes
135	-/-	c.5882G>A	p.(Gly1961Glu)	c.4947del	p.(Glu1650Argfs*12)	yes
136	-/-	c.5882G>A	p.(Gly1961Glu)	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	yes
137	-/-	c.5882G>A	p.(Gly1961Glu)	c.[4469G>A;5603A>T]	p.([Cys1490Tyr;Asn1868Ile])	yes
138	-/-	c.5882G>A	p.(Gly1961Glu)	c.2894A>G	p.(Asn965Ser)	yes
139	-/-	c.5882G>A	p.(Gly1961Glu)	c.4234C>T	p.(Gln1412*)	yes
140	-/-	c.5882G>A	p.(Gly1961Glu)	c.3050+5G>A	p.(Leu973_His1017delinsPhe)	yes
141	-/-	c.5882G>A	p.(Gly1961Glu)	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	yes
142	-/-	c.5882G>A	p.(Gly1961Glu)	c.4253+5G>T	p.([Ile1377Hisfs*3,=])	yes

143	-/-	c.5882G>A	p.(Gly1961Glu)	c.5044_5058del	p.(Val1682_Val1686del)	yes
144	-/-	c.5882G>A	p.(Gly1961Glu)	c.1909_1922del	p.(Met637Leufs*12)	yes
145	-/-	c.5882G>A	p.(Gly1961Glu)	c.[5461-10T>C;5603A>T]	p.(Thr1821Aspfs*6)	yes
146	-/-	c.5882G>A	p.(Gly1961Glu)	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	yes
147	-/-	c.5882G>A	p.(Gly1961Glu)	c.3050+5G>A	p.(Leu973_His1017delinsPhe)	yes
148	-/-	c.[3758C>T;5882G>A]	p.([Thr1253Met;Gly1961Glu])	c.3050+5G>A	p.(Leu973_His1017delinsPhe)	yes
149	-/-	c.5882G>A	p.(Gly1961Glu)	c.2564G>A	p.(Trp855*)	yes
150	-/-	c.5882G>A	p.(Gly1961Glu)	c.[5461-10T>C;5603A>T]	p.(Thr1821Aspfs*6)	yes

Nucleotide positions and protein translation correspond to CCDS747.1 and NP_000341.2, respectively.

Supplemental Table 2 (*Chapter 3*): Demographic, clinical and genetic characteristics of patients homozygous for c.5882G>A p.(Gly1961Glu) and compound heterozygous for p.(Gly1961Glu) and c.4139C>T p.(Pro1380Leu) in *ABCA4*.

Patient #	Gender	Age at examination (years)	Age of onset (years)	Disease Duration	BCVA OD	BCVA OS	ffERG group	Allele 1	Allele 2
					OD	OS			
1	F	69	43	18	20/200	20/400	1	c.[769-784C>T;5882G>A]	c.5882G>A
2	F	21	n/a		20/25	20/40	1	c.[769-784C>T;5882G>A]	c.5882G>A
3	M	60	60	0*	20/70	20/70	n/a	c.[769-784C>T;5882G>A]	c.5882G>A
4	F	62	48	15	20/200	20/150	1	c.[769-784C>T;5882G>A]	c.5882G>A
5	M	45	38	7	20/800	20/25	n/a	c.[769-784C>T;5882G>A]	c.[769-784C>T;5882G>A]
6	M	42	n/a		20/25	20/30	n/a	c.[769-784C>T;5882G>A]	c.[3758C>T;5882G>A]
7	M	45	35	10	20/50	20/25	1	c.[2549A>G;5882G>A]	c.[2549A>G;5882G>A]
8	F	48	n/a		n/a	n/a	n/a	c.[5242G>A;5882G>A]	c.5882G>A
9	M	32	21	11	20/80	20/100	1	c.5882G>A	c.5882G>A
10	F	51	51	0*	20/30	20/30	1	c.5882G>A	c.5882G>A
11	M	45	14	31	20/125	20/125	1	c.[769-784C>T;5882G>A]	c.4139C>T
12	M	47	18	29	20/400	20/400	1	c.[769-784C>T;5882G>A]	c.4139C>T
13	M	63	10	53	20/300	20/150	1	c.[769-784C>T;5882G>A]	c.4139C>T
14	M	57	16	41	20/400	20/200	1	c.[769-784C>T;5882G>A]	c.4139C>T
15	M	34	31	3	20/40	20/50	n/a	c.[769-784C>T;5882G>A]	c.4139C>T
16	n/a	n/a	n/a		n/a	n/a	n/a	c.[769-784C>T;5882G>A]	c.4139C>T
17	F	59	39	20	20/70	20/100	1	c.[3758C>T;5882G>A]	c.4139C>T
18	F	53	48	5	20/50	20/50	1	c.[3758C>T;5882G>A]	c.4139C>T
19	M	27	n/a		n/a	n/a	n/a	c.[3758C>T;5882G>A]	c.4139C>T
20	M	60	n/a		n/a	n/a	n/a	c.5882G>A	c.4139C>T
21	M	13	9	4	20/125	20/100	1	c.5882G>A	c.4139C>T
22	F	17	15	2	20/150	20/150	1	c.5882G>A	c.4139C>T
23	F	23	18	5	20/40	20/30	n/a	c.5882G>A	c.4139C>T
24	F	69	10	59	20/200	20/200	1	c.5882G>A	c.4139C>T
25	F	57	15	42	20/125	20/200	1	c.5882G>A	c.4139C>T
26	F	33	24	9	20/150	20/150	1	c.5882G>A	c.4139C>T
27	M	29	20	9	20/60	20/50	n/a	c.5882G>A	c.4139C>T

n/a = not available; predicted amino acid substitution for cis modifying variants = c.769-784C>T p.([=,Leu257Aspfs*3], c.2549A>G p.(Tyr850Cys), c.3758C>T p.(Thr1253Met), c.5242G>A p.(Gly1748Arg). *Disease duration (DD) were calculated by subtracting the age at examination from the patient's self-reported age at which symptoms manifested.

Supplemental Table 3 (*Chapter 4*): Demographic, clinical and genetic summary of patients with ABCA4-associated retinitis pigmentosa 19 (RP19, #601718).

Patient	Sex	Age (years)	Age of onset (years)	Race/ Ethnicity	BCVA OD	BCVA OS	Consanguinity	Total ROH (Mb)	*ABCA4* allele 1 cDNA	*ABCA4* allele 1 Protein	*ABCA4* allele 2 cDNA	*ABCA4* allele 2 Protein	Oculomotor	Anterior segment	Photophobia	Ocular discomfort
1	F	26	9	Yemen	NLP	NLP	Yes	535.94	c.6112C>T	p.Arg2038Trp	c.6112C>T	p.Arg2038Trp	Horizontal nystagmus (exotropia)	Clear/Quiet	-	No
2	F	35	10	Indian	HM	HM	Yes	326.31	c.93G>A	p.Trp31*	c.93G>A	p.Trp31*	WNL	Clear/Quiet	-	N/A
3	F	19	6	Russian	20/800	CF	Yes	(not assessed)	c.5318C>T	p.Ala1773Val	c.5318C>T	p.Ala1773Val	Nystagmus + R(x)T	Mild PSC OU	-	N/A
4	M	29	7	Indian	CF	20/300	No	39.9	c.[66G>A;859-9T>C]	p.?	c.6658C>T	p.Gln2220*	R(x)T	Clear/Quiet	Yes	Ocular migraines
5	F	32	9	Iranian	HM	CF	Yes	(not assessed)	c.5018+2T>C	p.?	c.5018+2T>C	p.?	Nystagmus	Clear/Quiet	Yes	Sharp pain
6	F	45	9	Haitian	CF@1'	CF@1'	No	8.89	c.2947A>G	p.Thr983Ala	c.5186T>C	p.Leu1729Pro	-	Clear/Quiet	Yes	-
7	F	44	9	Hispanic	HM	HM	Yes	176.41	c.1804C>T	p.Arg602Trp	c.1804C>T	p.Arg602Trp	Nystagmus (horizontal)	Clear/Quiet	-	-
8	F	33	11	Caucasian	HM	HM	No	(not assessed)	c.4139C>T	p.Pro1380Leu	c.4919G>A	p.Arg1640Gln	Nystagmus (horizontal)	Clear/Quiet	Yes	N/A
9	F	39	8	Russian	20/400	HM	Yes	152.25	c.4667G>A	p.Arg1556Lys	c.4667G>A	p.Arg1556Lys	Exotropia	PSC OU	Yes	Intermittent pain; photopsias; fatigue
10	M	44	7	Egyptian	HM	CF	Yes	264.06	c.3988G>T	p.Glu1330*	c.3988G>T	p.Glu1330*	Nystagmus	PSC OU	-	N/A
11	F	15	7	Egyptian	20/400	20/400	Yes	(not assessed)	c.3988G>T	p.Glu1330*	c.3988G>T	p.Glu1330*	-	N/A	-	N/A
12	M	17	7	Egyptian	20/400	20/400	Yes	(not assessed)	c.3988G>T	p.Glu1330*	c.3988G>T	p.Glu1330*	-	N/A	-	N/A
13	F	60	13	Indian	20/150	20/400	Yes	(not assessed)	c.2894A>G	p.Asn965Ser	c.2894A>G	p.Asn965Ser	-	Clear/Quiet	-	-
14	M	26	7	Caucasian	20/400	CF@2'	No	(not assessed)	c.571-1G>T	p.?	c.2160+1G>C	p.?	No	Clear/Quiet	Yes	Pressure behind eyes
15	F	22	9	Caucasian	CF@2'	HM	No	71.71	c.571-1G>T	p.?	c.2160+1G>C	p.?	Nystagmus (horizontal)	PSC OU	Yes	Severe ocular migraine

181

16	F	14		Polish	20/400	20/400	No	92.38	c.[1622T>C;3113C>T]	p.[Leu541Pro;Ala1038Val]		p.E1087K	Strabismus	Clear/Quiet	-	-
17	F	22	7	El Salvador / German	HM	HM	No	(not assessed)	c.3210_3211dup	p.Ser1071Cysfs*14	c.4849-1G>A	p.?	Rotary nystagmus + exotropia	PSC OU	-	-
18	F	30	8	Indian	HM	HM	Yes	(not assessed)	c.6088C>T	p.Arg2030*	c.6088C>T	p.Arg2030*	Horizontal nystagmus	PSC OU	-	-
19	F	13	8	Dominican	20/300	20/200	Yes	(not assessed)	c.223T>G	p.Cys75Gly	c.223T>G	p.Cys75Gly	No		-	-
20	F	46	13	Italian	CF@2'	20/200	Yes	182.1	c.571-1G>T	p.?	c.571-1G>T	p.?	-	Clear/Quiet	-	Photopsias (blue/yellow)
21	M	9		Polish	20/200	20/200	No	99.84	c.[1622T>C;3113C>T]	p.[Leu541Pro;Ala1038Val]	c.[1622T>C;3113C>T]	p.[Leu541Pro;Ala1038Val]	-	-	-	-
22	M	17		Polish	20/200	20/200	No	(not assessed)	c.[1622T>C;3113C>T]	p.[Leu541Pro;Ala1038Val]	c.[1622T>C;3113C>T]	p.[Leu541Pro;Ala1038Val]	-	-	-	-
23	F	50	13	African-American	20/400	20/800	Yes	154.06	c.[4594G>A;5603A>T]	p.([Asp1532Asn;Asn1868Ile])	c.[4594G>A;5603A>T]	p.([Asp1532Asn;Asn1868Ile])	-	Clear/Quiet	-	-

Supplemental Table 4 (*Chapter 4*): List of identified variants from whole exome sequencing identified in the affected son, P1.1, (retinitis pigmentosa) of Patient 1.

Gene	cDNA	Protein	Zygosity	VEP annotation	dbSNP	Allele frequency	M-CAP	REVEL	Eigen	CADD	SpliceAI	ACMG classification	OMIM phenotypes (Inheritance)
ABCA4	c.6112C>T	p.(Arg2038Trp)	Heterozygous	Missense	rs61750643	1.59E-05	0.331	0.941	0.75	31.0	-	Pathogenic	Cone-rod dystrophy 3 (AR) Fundus flavimaculatus (AR) Retinal dystrophy, early-onset severe (AR) Retinitis pigmentosa 19 (AR) Stargardt disease 1 (AR)
NR2E3	c.119-2A>C	p.Val41Aalfs*23	Heterozygous	Splicing	-	5.00E-04	-	-	0.61	24.0	Exon 2 skipping (Haider et al.)	Pathogenic	Enhanced S-cone syndrome (AR) Retinitis pigmentosa 37 (AR, AD)
SEMA4A	c.1451G>A	p.(Gly484Ala)	Heterozygous	Missense	rs148744804	6.00E-04	0.016	0.671	0.35	23.4	-	Likely benign	Cone-rod dystrophy 10 (AR) Retinitis pigmentosa 35 (AD, AR)
PROM1	c.869G>A	p.(Arg290Gln)	Heterozygous	Missense	rs748149698	1.63E-05	0.007	0.018	-1.10	7.9	-	Likely benign	Cone-rod dystrophy 12 (AD, AR) Macular dystrophy, retinal, 2 (AD) Retinitis pigmentosa 41 (AD) Stargardt disease 4 (AD)
ADAM9	c.1501A>G	p.(Asn501Asp)	Heterozygous	Missense	rs140137562	5.69E-05	0.009	0.092	-0.21	22.5	-	Likely benign	Cone-rod dystrophy 9 (AR)
WDR19	c.2648C>T	p.(Ala883Val)	Heterozygous	Missense	-	(absent)	0.015	0.102	-0.44	21.7	-	Likely benign	Cranioectodermal dysplasia 4 (AR) Nephronophthisis 13 (AR) Senior-Loken syndrome 8 (AR) Short-rib thoracic dysplasia 5 with or without polydactyly (AR)

Supplemental Table 5 (*Chapter 5*): Demographic, clinical and genetic characteristics of individual ABCA4/Stargardt patients in the study cohort.

Patient	Sex	Age	AO	ABCA4 allele 1		*ABCA4* allele 2		ffERG	BCVA (Snellen		Clinical severity
				DNA change	protein change	DNA change	protein change		OD	OS	
1	M	60	55	c.5882G>A	p.(Gly1961Glu)	c.[769-784C>T;5882G>A]	p.([Leu257Aspfs*3,Gly1961Glu])		20/70	20/70	Prognosis 1
2	M	50	19	c.5141T>C	p.(Phe1714Ser)	c.4577C>T	p.(Thr1526Met)	Group I	20/40	20/70	Prognosis 1
3	M	65	48	c.3607G>A	p.(Gly1203Arg)	c.5603A>T	p.(Asn1868Ile)	Group I	20/20	20/25	Prognosis 1
4	F	54	48	c.[3758C>T;5882G>A]	p.([Thr1253Met;Gly1961Glu])	c.4139C>T	p.(Pro1380Leu)	Group I	20/50	20/50	Prognosis 1
5	F	50	49	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	c.5882G>A	p.(Gly1961Glu)		20/25	20/40	Prognosis 1
6	F	60	43	c.4249_4251del	p.(Phe1417del)	c.5603A>T	p.(Asn1868Ile)	Group I	20/150	20/150	Prognosis 1
7	F	52	35	c.52C>T	p.(Arg18Trp)	c.2966T>C	p.(Val989Ala)	Group I	20/20	20/20	Prognosis 1
8	M	53	46	c.[5461-10T>C;5603A>T]	p.([Thr1821Valfs*13,Thr1821Aspfs*6])	c.5603A>T	p.(Asn1868Ile)	Group I	20/20	20/20	Prognosis 1
9	M	53	42	c.4670A>G	p.(Tyr1557Cys)	c.5603A>T	p.(Asn1868Ile)	Group I	20/200	20/150	Prognosis 1
10	M	51	44	c.[769-784C>T;5882G>A]	p.([Leu257Aspfs*3,Gly1961Glu])	c.[769-784C>T;5882G>A]	p.([Leu257Aspfs*3,Gly1961Glu])	Group I	20/800	20/25	Prognosis 1
11	F	67	15	c.4139C>T	p.(Pro1380Leu)	c.5882G>A	p.(Gly1961Glu)	Group I	20/125	20/200	Prognosis 1
12	F	68	48	c.5882G>A	p.(Gly1961Glu)	c.[769-784C>T;5882G>A]	p.([Leu257Aspfs*3,Gly1961Glu])	Group I	20/80	20/80	Prognosis 1
13	F	74	35	c.6320G>A	p.(Arg2107His)	c.6320G>A	p.(Arg2107His)		20/100	20/100	Prognosis 1
14	F	70	55	c.[4594G>A;5603A>T]	p.([Asp1532Asn;Asn1868Ile])	c.5882G>A	p.(Gly1961Glu)	Group I	20/100	20/100	Prognosis 1
15	M	52	30	c.3050+5G>A	p.(Leu973_His1017delinsPhe)	c.5882G>A	p.(Gly1961Glu)	Group I	20/100	20/80	Prognosis 1
16	M	51	30	c.6229C>G	p.(Arg2077Gly)	c.6320G>A	p.(Arg2107His)				Prognosis 1
17	F	54	18	c.641T>A	p.(Ile214Asn)	c.5882G>A	p.(Gly1961Glu)		20/200	20/200	Prognosis 1
18	M	54	35	c.4139C>T	p.(Pro1380Leu)	c.5603A>T	p.(Asn1868Ile)	Group I	20/20	20/20	Prognosis 1
19	M	56	51	c.3995_4018dup	p.(Gln1332_Cys1339dup)	c.5882G>A	p.(Gly1961Glu)		20/30	20/30	Prognosis 1
20	F	71	18	c.4139C>T	p.(Pro1380Leu)	c.5882G>A	p.(Gly1961Glu)	Group I	20/200	20/200	Prognosis 1
21	F	50	14	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	c.5882G>A	p.(Gly1961Glu)	Group I			Prognosis 1
22	F	51	50	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	c.5603A>T	p.(Asn1868Ile)	Group I	20/50	20/25	Prognosis 1
23	F	60	39	c.[3758C>T;5882G>A]	p.([Thr1253Met;Gly1961Glu])	c.4139C>T	p.(Pro1380Leu)	Group I	20/70	20/100	Prognosis 1
24	M	51	30	c.5116del_5117delinsTA	p.(Val1706*)	c.5882G>A	p.(Gly1961Glu)	Group I	20/150	20/150	Prognosis 1
25	F	68	43	c.5882G>A	p.(Gly1961Glu)	c.[769-784C>T;5882G>A]	p.([Leu257Aspfs*3,Gly1961Glu])	Group I	20/200	20/400	Prognosis 1
26	M	66	50	c.1988G>A	p.(Trp663*)	c.4253+43G>A	p.([=,Ile1377Hisfs*3])	Group I	20/40	20/20	Prognosis 1
27	F	57	45	c.[4222T>C;4918C>T]	p.([Trp1408Arg;Arg1640Trp])	c.5603A>T	p.(Asn1868Ile)		20/20	20/20	Prognosis 1
28	F	56	50	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	c.5603A>T	p.(Asn1868Ile)		20/20	20/30	Prognosis 1
29	M	64	30	c.2041C>T	p.(Arg681*)	c.4685T>C	p.(Ile1562Thr)	Group I	20/25	20/25	Prognosis 2
30	F	70	22	c.1988G>A	p.(Trp663*)	c.3113C>T	p.(Ala1038Val)	Group I	20/30	20/40	Prognosis 2
31	F	54	31	c.1921T>C	p.(Cys641Arg)	c.5603A>T	p.(Asn1868Ile)	Group I	20/150	20/400	Prognosis 2
32	M	79	19	c.4539+1G>T	p.(?)	c.5882G>A	p.(Gly1961Glu)	Group I	20/400	20/400	Prognosis 2
33	F	52	45	c.161G>A	p.(Cys54Tyr)	c.4253+43G>A	p.([=,Ile1377Hisfs*3])	Group I	20/40	20/400	Prognosis 2
34	M	56	42	c.1804C>T	p.(Arg602Trp)	c.6320G>A	p.(Arg2107His)	Group I	20/150	20/400	Prognosis 2
35	F	50	36	c.4538A>G	p.(Gln1513Arg)	c.6316C>T	p.(Arg2106Cys)	Group II	CF	CF	Prognosis 2
36	M	52	45	c.2966T>C	p.(Val989Ala)	c.3211_3212insGT	p.(Ser1071Cysfs*13)	Group I	20/125	20/125	Prognosis 2
37	M	52	48	c.1522C>T	p.(Arg508Cys)	c.4223G>A	p.(Trp1408*)	Group I	20/25	20/25	Prognosis 2
38	M	68	50	c.4139C>T	p.(Pro1380Leu)	c.5603A>T	p.(Asn1868Ile)	Group I	CF	20/150	Prognosis 2
39	F	60	58	c.6089G>A	p.(Arg2030Gln)	c.6286G>A	p.(Glu2096Lys)	Group I	20/25	20/30	Prognosis 2

#	Sex	Age		Allele 1 (cDNA)	Allele 1 (protein)	Allele 2 (cDNA)	Allele 2 (protein)	Group			Prognosis
40	M	61	45	c.4139C>T	p.(Pro1380Leu)	c.5087G>A	p.(Ser1696Asn)	Group I	20/80	20/100	Prognosis 2
41	M	50	46	c.2023G>A	p.(Val675Ile)	c.5646G>A	p.(Met1882Ile)	Group II	20/80	20/100	Prognosis 2
42	M	62	60	c.6079C>T	p.(Leu2027Phe)	c.6089G>A	p.(Arg2030Gln)		20/20	20/20	Prognosis 2
43	F	65	32	c.5882G>A	p.(Gly1961Glu)	c.6449G>A	p.(Cys2150Tyr)	Group I	20/200	20/400	Prognosis 2
44	F	52	45	c.2543C>A	p.(Ala848Asp)	c.3113C>T	p.(Ala1038Val)	Group I	20/200	20/200	Prognosis 2
45	M	74	68	c.5603A>T	p.(Asn1868Ile)	c.6658C>T	p.(Gln2220*)		20/25	20/25	Prognosis 2
46	F	72	47	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	c.5603A>T	p.(Asn1868Ile)	Group I	20/400	20/400	Prognosis 2
47	M	72	32	c.5196+1056A>G	p.(Met1733Valfs*2)	c.5882G>A	p.(Gly1961Glu)		20/400	20/200	Prognosis 2
48	M	75	40	c.5882G>A	p.(Gly1961Glu)	c.6445C>T	p.(Arg2149*)	Group I	20/200	20/200	Prognosis 2
49	F	57	30	c.42_48del	p.(Trp15_Thr16delfs*23)	c.3113C>T	p.(Ala1038Val)	Group I	20/100	20/80	Prognosis 2
50	F	57	48	c.2382+179G>A	p.(?)	c.4577C>T	p.(Thr1526Met)	Group I	20/25	20/30	Prognosis 2
51	M	82	65	c.1100-6T>A	p.(Thr367Serfs*6)	c.6316C>T	p.(Arg2106Cys)		CF	20/400	Prognosis 2
52	M	58	30	c.4457C>T	p.(Pro1486Leu)	c.6079C>T	p.(Leu2027Phe)	Group I	20/20	20/20	Prognosis 2
53	F	56	45	c.4253+5G>T	p.(Ile1377Hisfs*3)	c.4685T>C	p.(Ile1562Thr)	Group I	20/70	CF	Prognosis 2
54	M	53	26	c.262G>A	p.(Gly88Arg)	c.5603A>T	p.(Asn1868Ile)	Group I	20/100	20/100	Prognosis 2
55	F	61	52	c.247_250dup	p.(Ser84Thrfs*15)	c.5603A>T	p.(Asn1868Ile)	Group I	20/125	20/20	Prognosis 2
56	F	78	20	c.5882G>A	p.(Gly1961Glu)	c.6342G>A	p.([Val2114_Ser2129delfs*5,=])	Group I	20/400	20/400	Prognosis 2
57	M	62	30	c.3050+5G>A	p.(Leu973_His1017delinsPhe)	c.5882G>A	p.(Gly1961Glu)	Group I	20/125	20/150	Prognosis 2
58	F	51	45	c.570+331_768+4523del	p.(?)	c.4685T>C	p.(Ile1562Thr)	Group I	20/20	20/20	Prognosis 2
59	F	56	50	c.5603A>T	p.(Asn1868Ile)	c.6229C>T	p.(Arg2077Trp)	Group I	20/40	20/30	Prognosis 2
60	M	53	48	c.1805G>A	p.(Arg602Gln)	c.[4469G>A;5603A>T]	p.([Cys1490Tyr;Asn1868Ile])	Group II	20/25	CF	Prognosis 3
61	M	68	57	c.161G>A	p.(Cys54Tyr)	c.4685T>C	p.(Ile1562Thr)		20/30	20/30	Prognosis 3
62	F	50	25	c.1868A>G	p.(Gln623Arg)	c.2041C>T	p.(Arg681*)		20/25	20/20	Prognosis 3
63	M	51	10	c.4577C>T	p.(Thr1526Met)	c.5714+5G>A	p.([=,Glu1863Leufs*33])				Prognosis 3
64	F	66	14	c.3287C>T	p.(Ser1096Leu)	c.6108_6130dup	p.(Ala2044Valfs*25)		20/400	20/400	Prognosis 3
65	F	55	38	c.1964T>G	p.(Phe655Cys)	c.[2588G>C;5603A>T]	p.([Gly863Ala,Gly863del;Asn1868Ile])		20/400	20/125	Prognosis 3
66	M	53		c.770T>G	p.(Leu257Arg)	c.5714+5G>A	p.([=,Glu1863Leufs*33])		20/400	20/300	Prognosis 3
67	F	50	40	c.4457C>T	p.(Pro1486Leu)	c.4793C>A	p.(Ala1598Asp)	Group II	20/25	CF	Prognosis 3
68	F	54	30	c.2587+2T>C	p.(?)	c.6320G>A	p.(Arg2107His)		CF	20/30	Prognosis 3
69	M	59	35	c.5113C>T	p.(Arg1705Trp)	c.5603A>T	p.(Asn1868Ile)	Group II	CF	20/20	Prognosis 3
70	M	54	15	c.3323G>A	p.(Arg1108His)	c.4139C>T	p.(Pro1380Leu)	Group I	20/150	20/150	Prognosis 3
71	M	70	55	c.5018+2T>C	p.(?)	c.5603A>T	p.(Asn1868Ile)	Group I	20/100	20/400	Prognosis 3
72	F	74	13	c.[4594G>A;5603A>T]	p.([Asp1532Asn;Asn1868Ile])	c.[4594G>A;5603A>T]	p.([Asp1532Asn;Asn1868Ile])	Group II	20/400	20/800	Prognosis 3
73	F	50	8	c.2966T>C	p.(Val989Ala)	c.5289del	p.(Val1764Trpfs*14)				Prognosis 3
74	F	58	23	c.[2588G>C;5603A>T]	p.([Gly863Ala,Gly863del;Asn1868Ile])	c.3335C>A	p.(Thr1112Asn)		CF	CF	Prognosis 3
75	M	68	62	c.1957C>T	p.(Arg653Cys)	c.5603A>T	p.(Asn1868Ile)	Group I	20/20	20/20	Prognosis 3
76	F	52	48	c.1015T>G	p.(Trp339Gly)	c.6320G>A	p.(Arg2107His)	Group I	20/30	20/40	Prognosis 3
77	M	76		c.[4594G>A;5603A>T]	p.([Asp1532Asn;Asn1868Ile])	c.[769-784C>T;5882G>A]	p.([Leu257Aspfs*3,Gly1961Glu])	Group I	20/60	20/40	Prognosis 3
78	M	63	10	c.4139C>T	p.(Pro1380Leu)	c.[769-784C>T;5882G>A]	p.([Leu257Aspfs*3,Gly1961Glu])	Group I	20/300	20/150	Prognosis 3
79	M	57	16	c.4139C>T	p.(Pro1380Leu)	c.[769-784C>T;5882G>A]	p.([Leu257Aspfs*3,Gly1961Glu])	Group I	CF	CF	Prognosis 3
80	M	53	10	c.859-9T>C	p.([=,Phe287_Arg452del])	c.885del	p.(Leu296Cysfs*4)	Group II	CF	CF	Prognosis 4
81	M	54	27	c.[2588G>C;5603A>T]	p.([Gly863Ala,Gly863del;Asn1868Ile])	c.4506C>A	p.(Cys1502*)	Group II	20/400	CF	Prognosis 4
82	M	50	10	c.3322C>T	p.(Arg1108Cys)	c.4139C>T	p.(Pro1380Leu)	Group III	20/400	20/400	Prognosis 4

83	M	83	20	c.3322C>T	p.(Arg1108Cys)	c.4139C>T	p.(Pro1380Leu)	Group III	HM	HM	Prognosis 4
84	F	57	17	c.3322C>T	p.(Arg1108Cys)	c.4139C>T	p.(Pro1380Leu)	Group III	20/400	20/400	Prognosis 4
85	F	76	19	c.4139C>T	p.(Pro1380Leu)	c.4139C>T	p.(Pro1380Leu)		HM	20/800	Prognosis 4
86	M	60	13	c.604del	p.(Asp202Thrfs*39)	c.5714+5G>A	p.([=,Glu1863Leufs*33])	Group III	20/400	20/300	Prognosis 4
87	M	55	17	c.4793C>A	p.(Ala1598Asp)	c.4793C>A	p.(Ala1598Asp)	Group III	HM	20/400	Prognosis 4
88	M	63	29	c.4139C>T	p.(Pro1380Leu)	c.5714+5G>A	p.([=,Glu1863Leufs*33])	Group III	20/400	20/400	Prognosis 4
89	F	55	11	c.4793C>A	p.(Ala1598Asp)	c.4919G>A	p.(Arg1640Gln)		20/400	20/400	Prognosis 4
90	M	74	14	c.[2588G>C;5603A>T]	p.([Gly863Ala, Gly863del;Asn1868Ile])	c.6088C>T	p.(Arg2030*)		HM	CF	Prognosis 4
91	F	69	8	c.161G>A	p.(Cys54Tyr)	c.2160+1G>C	p.(?)	Group III	HM	HM	Prognosis 4
92	F	52	13	c.2894A>G	p.(Asn965Ser)	c.6119G>A	p.(Arg2040Gln)	Group III	20/200	20/50	Prognosis 4
93	F	52	5	c.768G>T	p.(Leu257Valfs*17)	c.4539+2001G>A	p.([=,Arg1514Leufs*36])	Group III	20/400	20/400	Prognosis 4
94	F	55	27	c.5059del	p.(Ile1687Phefs*15)	c.5351T>G	p.(Leu1784Arg)	Group III	20/400	CF	Prognosis 4
95	M	75	25	c.3050+5G>A	p.(Leu973_His1017delinsPhe)	c.[4594G>A;5603A>T]	p.([Asp1532Asn;Asn1868Ile])	Group III	20/400	20/200	Prognosis 4
96	M	60	7	c.4139C>T	p.(Pro1380Leu)	c.4601del	p.(Leu1534Trpfs*1)	Group III	CF	CF	Prognosis 4
97	M	69	17	c.768G>T	p.(Leu257Valfs*17)	c.6320G>A	p.(Arg2107His)	Group III	20/400	20/400	Prognosis 4
98	M	62	11	c.3056C>T	p.(Thr1019Met)	c.3056C>T	p.(Thr1019Met)	Group III	CF	20/400	Prognosis 4
99	M	70	17	c.161G>A	p.(Cys54Tyr)	c.4773+3A>G	p.([Tyr1557Alafs*18,=])	Group III	HM	HM	Prognosis 4
100	F	56	21	c.2915C>A	p.(Thr972Asn)	c.5714+5G>A	p.([=,Glu1863Leufs*33])	Group III	CF	20/400	Prognosis 4
101	F	53	33	c.1037A>C	p.(Lys346Thr)	c.3350C>T	p.(Thr1117Ile)		CF	CF	Prognosis 4
102	M	70	10	c.[4594G>A;5603A>T]	p.([Asp1532Asn;Asn1868Ile])	c.[4594G>A;5603A>T]	p.([Asp1532Asn;Asn1868Ile])	Group III	20/400	20/200	Prognosis 4
103	F	68	18	c.4139C>T	p.(Pro1380Leu)	c.[4594G>A;5603A>T]	p.([Asp1532Asn;Asn1868Ile])		CF	HM	Prognosis 4
104	F	62	13	c.2894A>G	p.(Asn965Ser)	c.2894A>G	p.(Asn965Ser)	Group III	20/150	20/400	Prognosis 4
105	F	59	18	c.[5461-10T>C;5603A>T]	p.([Thr1821Valfs*13,Thr1821Aspfs*6])	c.6079C>T	p.(Leu2027Phe)		20/400	20/400	Prognosis 4
106	M	68	8	c.3056C>T	p.(Thr1019Met)	c.5687T>A	p.(Val1896Asp)	Group III	HM	HM	Prognosis 4
107	M	57	10	c.4540-2A>G	p.(?)	c.6320G>A	p.(Arg2107His)	Group III	CF	CF	Prognosis 4
108	F	69	18	c.1958G>A	p.(Arg653His)	c.6088C>T	p.(Arg2030*)	Group III	CF	CF	Prognosis 4
109	M	52	51	c.71G>A	p.(Arg24His)	c.1561del	p.(Val521Serfs*47)	Group III	CF	20/40	Prognosis 4
110	M	64	27	c.2531T>G	p.(Leu844Arg)	c.[2588G>C;5603A>T]	p.([Gly863Ala, Gly863del;Asn1868Ile])	Group III	20/400	20/400	Prognosis 4
111	M	66	5	c.[302+68C>T;4539+2028C>T]	p.(Arg1514Leufs*36)	c.6148-698_6670del	p.(?)	Group III	HM	HM	Prognosis 4
112	F	51	6	c.4139C>T	p.(Pro1380Leu)	c.4139C>T	p.(Pro1380Leu)	Group III	CF	CF	Prognosis 4

Supplemental Table 6 (Chapter 5): Summary of predicted pathogenicity and classification criteria of all ABCA4 variants identified in the study cohort.

cDNA	Protein	Location	Allele count	Coding effect	MAF (gnomAD)	Homozygotes (gnomAD)	M-CAP	REVEL	Eigen	CADD	SPIDEX z-score	SpliceAI	Study classification	Classification criteria	Reference
c.42_48del	p.(Trp15_Thr16delfs*23)	Exon 1	1	Frameshift	-	-	-	-	-	-	-		PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.52C>T	p.(Arg18Trp)	Exon 1	1	Missense	1.59E-05	0	0.328	0.771	0.006	24.6	-		Severe	Observed in trans to hypomorphic allele in patient	Zernant et al. (2017)
c.71G>A	p.(Arg24His)	Exon 2	1	Missense	0.000184	0	0.280	0.81	0.67	28.3	-0.765	-	Moderate		-
c.161G>A	p.(Cys54Tyr)	Exon 3	4	Missense	3.54E-05	0	0.823	0.98	0.87	33	-0.19	-	Severe	Observed in trans to hypomorphic allele in patient; Observed exon skipping (HEK293T)	Zernant et al. (2017); Fadaie et al. (2019)
c.247_250dup	p.(Ser84Thrfs*15)	Exon 3	1	Frameshift	-	-	-	-	-	-	-		PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.262G>A	p.(Gly88Arg)	Exon 3	1	Missense	3.98E-06	0	0.737	0.846	0.927	28.7	0.119	-	Severe	Observed in trans to hypomorphic allele in patient	Zernant et al. (2017)
c.604del	p.(Asp202Thrfs*39)	Exon 6	1	Frameshift	-	-	-	-	-	-	-		PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.641T>A	p.(Ile214Asn)	Exon 6	1	Missense	-	-	-	-	-	-	-	-	Moderate		-
c.768G>T	p.(Leu257Valfs*17)	Exon 6	2	Synonymous	8.49E-05	0	-	-	-	-	-2.024	Donor Loss at -0 bp (Δscore = 0.42); Donor Gain at -35 bp (Δscore = 0.71)	Severe	No residual RNA observed in midi-gene analyses (HEK293T)	Sangermano et al. (2018)
c.570+331_768+4523del	p.(?)	Intron 5	1	?	-	-	-	-	-	-	-	-	PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.770T>G	p.(Leu257Arg)	Exon 7	1	Missense	-	-	-	-	-	-	-	-	Moderate		-
c.859-9T>C	p.([=,Phe287_Arg452del])	Intron 7	1	?	7.83E-05	0	-	-	-	-	-0.785	-	Moderate	75% residual RNA observed in midi-gene analyses (HEK293T)	Sangermano et al. (2018)
c.885del	p.(Leu296Cysfs*4)	Exon 8	1	Frameshift	1.99E-05	0	-	-	-	-	-	-	PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.1015T>G	p.(Trp339Gly)	Exon 8	1	Missense	7.96E-06	0	0.866	0.85	0.74	26.8	-0.3	-	Moderate		-
c.1037A>C	p.(Lys346Thr)	Exon 8	1	Missense	7.96E-06	0	0.773	0.85	0.59	23.8	0.732	-	Moderate		-
c.1100-6T>A	p.(Thr367Serfs*6)	Intron 8	1	?	3.98E-06	0	-	-	-	-	-0.43	no mRNA	Severe	No residual RNA observed in midi-gene analyses (HEK293T)	Sangermano et al. (2018)

c.1522C>T	p.(Arg508Cys)	Exon 11	1	Missense	0.000166	0	0.171	0.656	0.457	24.3	-0.962	-	Moderate		-
c.1561del	p.(Val521Serfs*47)	Exon 12	1	Frameshift	-	0	-	-	-	-	-	-	PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.1804C>T	p.(Arg602Trp)	Exon 13	1	Missense	4.38E-05	0	0.498	0.93	0.59	25.6	-1.568	-	Severe	Observed in trans to hypomorphic allele in patient	Zernant et al. (2017)
c.1805G>A	p.(Arg602Gln)	Exon 13	1	Missense	2.13E-05	0	0.426	0.85	0.48	26.1	-1.689	-	Moderate		-
c.1868A>G	p.(Gln623Arg)	Exon 13	1	Missense	-	-	-	-	-	-	-	-	Moderate		-
c.1921T>C	p.(Cys641Arg)	Exon 13	1	Missense	-	-	-	-	-	-	-	-	Severe	Observed in trans to hypomorphic allele in patient	Zernant et al. (2017)
c.1957C>T	p.(Arg653Cys)	Exon 14	1	Missense	1.61E-05	0	0.774	0.85	0.93	32	0.508	-	Severe	Observed in trans to hypomorphic allele in patient	Zernant et al. (2017)
c.1958G>A	p.(Arg653His)	Exon 14	1	Missense	3.57E-05	0	0.779	0.889	0.581	25.7	-0.3	-	Moderate		-
c.1964T>G	p.(Phe655Cys)	Exon 14	1	Missense	0.000391	0	0.805	0.93	0.84	28.6	0.427	-	Moderate		-
c.1988G>A	p.(Trp663*)	Exon 14	2	Nonsense	-	-	-	-	-	-	-	Acceptor Gain at -2 bp (Δscore = 0.37)	PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.2023G>A	p.(Val675Ile)	Exon 14	1	Missense	7.57E-05	0	0.685	0.71	0.93	26.9	-0.41	-	Moderate		-
c.2041C>T	p.(Arg681*)	Exon 14	2	Nonsense	7.96E-06	0	-	-	0.43	35	-2.525	-	PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.2160+1G>C	p.(?)	Intron 14	1	Splice defect	3.98E-06	0	-	-	1.229	34	-1.976		PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.2382+179G>A	p.(?)	Intron 15	1	?	0.000892	0	-	-	-	-	-	Donor Gain at 179bp (Δscore = 0.01)	Severe	Observed in trans to hypomorphic allele in patient	Current study
c.2531T>G	p.(Leu844Arg)	Exon 16	1	Missense	-	-	-	-	-	-	-	-	Moderate		-
c.2543C>A	p.(Ala848Asp)	Exon 16	1	Missense	-	-	-	-	-	-	-	-	Severe	Observed in trans to hypomorphic allele in patient	-
c.2587+2T>C	p.(?)	Intron 16	1	Splice defect	-	-	-	-	-	-	-	-	PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.2894A>G	p.(Asn965Ser)	Exon 19	3	Missense	0.000124	0	0.551	0.78	0.61	25.7	-1.087	-	Severe	Observed in trans to hypomorphic allele in patient	Zernant et al. (2017)
c.2915C>A	p.(Thr972Asn)	Exon 19	1	Missense	3.58E-05	0	0.666	0.85	1.00	26.5	0.942	-	Moderate		-
c.2966T>C	p.(Val989Ala)	Exon 20	3	Missense	0.000297	0	0.288	0.84	0.27	23.5	-0.777	-	Moderate		-
c.3050+5G>A	p.(Leu973_His1017delinsPhe)	Intron 20	3	?	1.59E-05	0	-	-	-	-	-3.259	Donor Loss at 5 bp (Δscore = 0.89)	Severe	No residual RNA observed in midi-gene analyses (HEK293T)	Sangermano et al. (2018)

c.	p.	Exon/Intron		Type									Severity	Comment	Reference
c.3056C>T	p.(Thr1019 Met)	Exon 21	3	Missense	3.18E-05	0	0.611	0.96	1.10	28.8	-1.358	-	Severe	Associated with poor prognoses	Current study
c.3113C>T	p.(Ala1038 Val)	Exon 21	3	Missense	0.001755	2	0.651	0.53	-0.07	22.2	-0.825	-	Rare hypomorph		-
c.3211_3212insGT	p.(Ser1071 Cysfs*13)	Exon 22	1	Frameshift	-	-	-	-	-	-	-		PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.3287C>T	p.(Ser1096 Leu)	Exon 22	1	Missense	3.98E-06	0	0.839	0.88	0.79	29.2	-1.396	Acceptor Gain at -23 bp (Δscore = 0.26)	Moderate		-
c.3322C>T	p.(Arg1108 Cys)	Exon 22	3	Missense	0.000127	0	0.797	0.89	0.81	32	-1.982	-	Moderate		-
c.3323G>A	p.(Arg1108 His)	Exon 22	1	Missense	1.59E-05	0	0.868	0.84	0.64	31	-1.473	-	Moderate		-
c.3335C>A	p.(Thr1112 Asn)	Exon 23	1	Missense	-	-	-	-	-	-	-	Close to Exon 23 boundary; possible splice defect	Moderate		-
c.3350C>T	p.(Thr1117Ile)	Exon 23	1	Missense	-	-	-	-	-	-	-	-	Moderate		-
c.3607G>A	p.(Gly1203 Arg)	Exon 24	1	Missense	-	-	-	-	-	-	-	Donor Loss at 0 bp (Δscore = 0.35)	Severe	Observed in trans to hypomorphic allele in patient	Current study
c.3995_4018dup	p.(Gln1332_Cys1339dup)	Exon 27	1	In-frame duplication	-	-	-	-	-	-	-	-	Moderate		-
c.4139C>T	p.(Pro1380 Leu)	Exon 28	20	Missense	0.000234	0	0.391	0.87	0.70	25.2	-2.554	-	Severe	Observed in trans to hypomorphic allele in patient	Zernant et al. (2017)
c.4223G>A	p.(Trp1408 *)	Exon 28	1	Nonsense	-	-	-	-	-	-	-		PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.4249_4251del	p.(Phe1417 del)	Exon 28	1	In-frame deletion	-	-	-	-	-	-	-	-	Severe	Observed in trans to hypomorphic allele in patient	Zernant et al. (2017)
c.4253+5G>T	p.(Ile1377Hisfs*3)	Intron 28	1	?	-	0	-	-	-	-	-	Donor Loss at 5 bp (Δscore = 0.44)	Severe	No residual RNA observed in midi-gene analyses (HEK293T)	Sangermano et al. (2018)
c.4253+43G>A	p.([=,Ile1377Hisfs*3])	Intron 28	2	?	0.004694	5	-	-	-	-	-	Donor loss at 43 bp (Δscore = 0.6)	Rare hypomorph	Expresses milder, late-onset (foveal sparing) disease only when in trans to deleterious allele	Zernant et al. (2018)
c.4457C>T	p.(Pro1486 Leu)	Exon 30	2	Missense	0.000122	0	0.59	0.88	0.591	24.6	1.394	-	Rare hypomorph		-
c.4506C>A	p.(Cys1502 *)	Exon 30	1	Nonsense	5.15E-06	0	-	-	0.792	45	-3.602		PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.4538A>G	p.(Gln1513 Arg)	Exon 30	1	Missense	-	-	0.313	0.79	0.16	32	-1.381	Donor Loss at -1 bp (Δscore = 0.73)	Severe	Exon elongation observed in midi-gene analyses (HEK293T)	Sangermano et al. (2018)
c.4539+1G>T	p.(?)	Intron 30	1	Splice defect	-	-	-	-	-	-	-	Donor Loss at 1 bp (Δscore = 0.77)	PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.4539+2001G>A	p.([=,Arg1514Leufs*36])	Intron 30	1	?	3.18E-05	0	-	-	-	-	-		Moderate	Variant results in pseudoexon insertion (patient-derived	Albert et al. (2018)

cDNA	Protein	Location	N	Type	gnomAD	Hom						Splice prediction	Classification	Comment	Reference
														photoreceptor precursor cells) in ~15% of RNA (85% correct RNA)	
c.4540-2A>G	p.(?)	Intron 30	1	Splice defect	3.98E-06	0	-	-	1.01	33	-2.223	Acceptor Loss at -2 bp (Δscore = 0.94)	PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.4577C>T	p.(Thr1526Met)	Exon 31	3	Missense	6.36E-05	0	0.566	0.887	0.55	24.3	0.286	-	Rare hypomorph		-
c.4601del	p.(Leu1534Trpfs*1)	Exon 31	1	Frameshift	-	-	-	-	-	-	-		PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.4670A>G	p.(Tyr1557Cys)	Exon 33	1	Missense	3.98E-06	0	0.897	0.92	0.86	28.9	-1.007	-	Severe	Observed in trans to hypomorphic allele in patient	Zernant et al. (2017)
c.4685T>C	p.(Ile1562Thr)	Exon 33	4	Missense	0.001248	1	0.174	0.736	0.04	23.3	0.865	-	Rare hypomorph		-
c.4773+3A>G	p.([Tyr1557Alafs*18,=])	Intron 33	1	?	8.13E-05	0	-	-	-	-	-2.423	Donor Loss at 3 bp (Δscore = 0.23); Donor Gain at -22 bp (Δscore = 0.31)	Severe	25% residual RNA observed in midi-gene analyses (HEK293T)	Sangermano et al. (2018)
c.4793C>A	p.(Ala1598Asp)	Exon 34	4	Missense	2.4E-05	0	0.182	0.69	0.08	22	2.871	-	Severe	Associated with poor prognoses	Current study
c.4919G>A	p.(Arg1640Gln)	Exon 35	1	Missense	3.18E-05	0	0.343	0.77	0.82	32	-1.565	Acceptor Gain at -2 bp (Δscore = 0.56)	Moderate		-
c.5018+2T>C	p.(?)	Intron 35	1	Splice defect	-	-	-	-	-	-	-	Donor Loss at 2 bp (Δscore = 0.93)	PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.5059del	p.(Ile1687Phefs*15)	Exon 36	1	Frameshift	-	-	-	-	-	-	-		PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.5087G>A	p.(Ser1696Asn)	Exon 36	1	Missense	3.9E-05	0	0.756	0.69	0.79	28.7	0.779	-	Moderate		-
c.5113C>T	p.(Arg1705Trp)	Exon 36	1	Missense	1.07E-05	0	0.710	0.92	0.48	26.2	-1.004	-	Severe	Observed in trans to hypomorphic allele in patient	Current study
c.5116del_5117delinsTA	p.(Val1706*)	Exon 36	1	Nonsense	-	-	-	-	-	-	-		PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.5141T>C	p.(Phe1714Ser)	Exon 36	1	Missense	-	-	0.62	0.91	0.93	0.622	-	-	Severe	Observed in trans to hypomorphic allele in patient	Current study
c.5196+1056A>G	p.(Met1733Valfs*2)	Intron 36	1	?	-	-	-	-	-	-	-	Cryptic Donor Strongly activated	Severe	Intron elongation observed in RNA derived from patient keratinocytes	Braun et al. (2013)
c.5289del	p.(Val1764Trpfs*14)	Exon 37	1	Frameshift	-	-	-	-	-	-	-		PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.5351T>G	p.(Leu1784Arg)	Exon 38	1	Missense	7.95E-06	0	0.288	0.85	0.79	28.9	1.236	-	Moderate		-

c. variant	p. variant	Location	n	Type	Freq	Count	Score1	Score2	Score3	CADD	Score5	Splice	Classification	Comments	Reference
c.[5461-10T>C;5603A>T]	p.([Thr1821 Valfs*13,Thr1821Aspfs*6])	Intron 38	2	?	0.00022	0	-	-	-	-	-1.588	-	Severe	Variant results in exon skipping (patient-derived photoreceptor precursor cells)	Sangermano et al. (2016)
c.5603A>T	p.(Asn1868Ile)	Exon 40	18	Missense	0.042191	364	-	0.40	0.03	22.7	-1.38	-	Hypomorph	Expresses milder, late-onset (foveal sparing) disease only when in trans to deleterious allele	Zernant et al. (2017)
c.5646G>A	p.(Met1882Ile)	Exon 40	1	Missense	2.78E-05	-	0.127	0.84	0.64	24.7	-2.072	-	Moderate		-
c.5687T>A	p.(Val1896Asp)	Exon 40	1	Missense	3.98E-06	0	0.072	0.558	-0.489	22.4	2.729	-	Moderate		-
c.5714+5G>A	p.([=,Glu1863Leufs*33])	Intron 40	5	?	0.000297	0	-	-	-	-	-2.202	Donor Loss at 5 bp (Δscore = 0.31)	Severe	39% residual RNA observed in midi-gene analyses (HEK293T); associated with poor prognoses (current study)	Sangermano et al. (2018); Current study
c.5882G>A	p.(Gly1961Glu)	Exon 42	18	Missense	0.004564	10	-	0.76	0.35	28.4	2.133	-	p.(Gly1961Glu)		-
c.6079C>T	p.(Leu2027Phe)	Exon 44	3	Missense	0.000198	0	0.123	0.95	0.86	28.5	0.516	-	Severe	Observed in trans to hypomorphic allele in patient	Current study
c.6088C>T	p.(Arg2030*)	Exon 44	2	Nonsense	2.78E-05	0	-	-	0.54	45	-3.287	-	PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.6089G>A	p.(Arg2030Gln)	Exon 44	2	Missense	0.000354	0	0.122	0.86	0.62	29.5	-0.296	-	Rare hypomorph		-
c.6119G>A	p.(Arg2040Gln)	Exon 44	1	Missense	0.000311	1	0.305	0.91	0.52	26.2	-0.392	-	Moderate		-
c.6108_6130dup	p.(Ala2044Valfs*25)	Exon 44	1	Frameshift	-	-	-	-	-	-	-	-	PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.6229C>G	p.(Arg2077Gly)	Exon 45	1	Missense	3.98E-06	0	0.663	0.95	0.76	29.1	-1.808	-	Moderate		-
c.6229C>T	p.(Arg2077Trp)	Exon 45	1	Missense	3.98E-06	0	0.506	0.95	0.79	31	-0.786	-	Severe	Observed in trans to hypomorphic allele in patient	Zernant et al. (2017)
c.6286G>A	p.(Glu2096Lys)	Exon 46	1	Missense	7.98E-06	0	0.392	0.96	1.17	32	-0.371	-	Severe	Observed in trans to hypomorphic allele in patient	Current study
c.6316C>T	p.(Arg2106Cys)	Exon 46	2	Missense	0.000131	0	0.368	0.89	0.89	32	-0.273	-	Moderate		-
c.6320G>A	p.(Arg2107His)	Exon 46	8	Missense	0.002027	4	-	0.88	0.84	32	-0.319	-	Moderate		-
c.6342G>A	p.([Val2114_Ser2129delfs*5,=])	Exon 46	1	Synonymous	1.77E-05	0	-	-	-	-	-0.656	Donor Loss at -44 bp (Δscore = 0.6); Donor Gain at 3 bp (Δscore = 0.68)	Severe	Partial exon excision observed in RNA derived from patient keratinocytes	Braun et al. (2013)
c.6445C>T	p.(Arg2149*)	Exon 47	1	Nonsense	1.99E-05	0	-	-	0.95	54	-3.602	-	PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.6449G>A	p.(Cys2150Tyr)	Exon 47	1	Missense	2.83E-05	0	0.525	0.92	0.80	32	0.964	-	Moderate		-
c.6658C>T	p.(Gln2220*)	Exon 48	1	Nonsense	5.17E-05	0	-	-	1.056	49	-3.09	-	PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation	Richards et al. (2015)

															codon, single or multi-exon deletion)	
c.6148-698_6670del	p.(?)	Intron 40	1	?	-	-	-	-	-	-	-		PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion)	Richards et al. (2015)
c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	Exon12;Exon 21	5	(complex allele)	-	-	-	-	-	-	-	-	Severe	Observed in trans to hypomorphic allele in patient	Zernant et al. (2017)
c.[2588G>C;5603A>T]	p.([Gly863Ala, Gly863del; Asn1868Ile])	Exon 17;Exon 40	5	(complex allele)	-	-	-	-	-	-	-	-	Severe	Associated with poor prognoses	Current study
c.[4222T>C;4918C>T]	p.([Trp1408Arg;Arg1640Trp])	Exon 28;Exon 35	1	(complex allele)	-	-	-	-	-	-	-	-	Severe	Observed in trans to hypomorphic allele in patient	Zernant et al. (2017)
c.[4469G>A;5603A>T]	p.([Cys1490Tyr;Asn1868Ile])	Exon 30;Exon 40	1	(complex allele)	-	-	-	-	-	-	-	-	Moderate		-
c.[302+68C>T;4539+2028C>T]	p.(Arg1514Leufs*36)	Intron 30;Intron 3	1	(complex allele)	-	-	-	-	-	-	-	-	Moderate	c.4539+2028C>T results in pseudoexon insertion (patient-derived photoreceptor precursor cells) in 30% RNA (70% correct RNA)	Albert et al. (2018)
c.[4594G>A;5603A>T]	p.([Asp1532Asn;Asn1868Ile])	Exon 31;Exon 40	8	(complex allele)	-	-	-	-	-	-	-	-	Severe	Associated with poor prognoses	Current study
c.[769-784C>T;5882G>A]	p.([Leu257Aspfs*3,Gly1961Glu])	Intron 6,Exon 42	8	(complex allele)	-	-	-	-	-	-	-	-	Moderate		Lee et al. (2021)
c.[3758C>T;5882G>A]	p.([Thr1253Met;Gly1961Glu])	Exon25;Exon 42	2	(complex allele)	-	-	-	-	-	-	-	-	Moderate		

192

Supplemental Table 7 (*Chapter 6*). Genotype associations associated with selection filters.

ABCA4 genotypes	Allele-specific total (n)	Presence of DDAF (Allele-specific %)	Comparison relative to other genotypes		Age (years)					
			OR [95% CI]	p-value	Mean	Median	Min	Max	SD	IQR
p.(Gly1961Glu)	85	27 (31.8%)	0.35 [0.21; 0.58]	< 0.0001	54.0	50	26	80	14	22
p.(Asn1868Ile)	43	23 (53.5%)	1.07 [0.57; 2.01]	0.8418	57.8	56	29	74	12.5	18
Rare hypomorphs	32	21 (65.6%)	1.84 [0.86; 3.92]	0.114	52.6	52	27	75	12.5	18
Other alleles	164	85 (51.8%)	0.99 [0.66; 1.46]	0.9425	47.6	48	13	85	16.1	25
1 PVS1 allele	74	46 (65.2%)	1.65 [0.99; 2.77]	0.0562	45.0	47	10	80	17.7	24
2 PVS1 allele	17	14 (82.4%)	4.53 [1.28; 16.00]	0.019	31.4	28	19	49	10.1	14

ABCA4 genotypes	Allele-specific total (n)	DDAF <30° SW-AF (Allele-specific %)	Comparison relative to other genotypes		Age (years)					
			OR [95% CI]	p-value	Mean	Median	Min	Max	SD	IQR
p.(Gly1961Glu)	27	26 (96.3%)	25.19 [3.35; 189.41]	0.0017	53.1	50	26	80	13.5	22
p.(Asn1868Ile)	23	16 (69.6%)	1.88 [0.74; 4.77]	0.163	52.4	54	29	72	10.9	11
Rare hypomorphs	21	17 (81.0%)	3.64 [1.18; 11.22]	0.001	51	47	27	75	12.9	20.5
Other alleles	85	44 (51.8%)	0.73 [0.42; 1.26]	0.2606	40.6	41	13	78	14.4	22.5
1 PVS1 allele	46	17 (37.0%)	0.36 [0.18; 0.71]	0.0032	39.8	47	10	66	16.9	25
2 PVS1 allele	14	2 (14.3%)	0.11 [0.02; 0.52]	0.0052	24	24	23	25	1.4	-

ABCA4 genotypes	Allele-specific total (n)	DDAF well-delineated (Allele-specific %)	Comparison relative to other genotypes		Age (years)					
			OR [95% CI]	p-value	Mean	Median	Min	Max	SD	IQR
p.(Gly1961Glu)	26	22 (85.6%)	3.76 [1.20; 11.77]	0.023	52.7	50	26	80	13.4	17
p.(Asn1868Ile)	16	14 (87.5%)	4.42 [0.95; 20.44]	0.058	51.6	53.5	29	72	11.6	13
Rare hypomorphs	17	13 (76.5%)	1.92 [0.59; 6.30]	0.282	55.5	52	41	75	10.6	17.5
Other alleles	44	20 (45.5%)	0.27 [0.12; 0.59]	0.001	49.9	53	24	78	12.7	15
1 PVS1 allele	17	9 (52.9%)	0.56 [0.20; 1.58]	0.276	52	52	43	66	6.7	7.5
2 PVS1 allele	2	1 (50.5%)	0.54 [0.03; 8.83]	0.665	25*	-	-	-	-	-

Supplemental Table 8 (*Chapter 7*): Summary of molecular and clinical symptoms of 27 individuals with CERT1 mutations.[qq]

Patient	Sex	Current age (ys)	Age of onset (months)	cDNA	Protein	Inheritance	Infantile feeding difficulties	Failure to thrive	Hypotonia	Growth delay	Seizures	Motor delay	Intellectual disability	Speech delay	Regression	ASD	ADHD	Stereotypy	Aggression	Self-injury	Pain insensitivity	Sleep disturbances	Craniofacial/ Skeletal dysmorphology	GI symptoms	Immune dysregulation	Heart defects
1	f	8	Congenital	279T>G	S93R		+	−	−	+	−	+	+	+++	−	+	+	+	+	+	−	+	+	−		
2	m	24	2 mos	395C>T	S132L	*de novo*	−		+	−	+	++++	++++	++++	−	possible	−	−	−	+	+	+	+		+	
3	m	18	Congenital	395C>T	S132L	*de novo*	+	+	+		+	++++	++++	++++	−	−	−	−	−	−	−	+	+	+	+	+
4	m	1	Congenital	395C>T	S132L	*de novo*	+	+	hypertonic	+	−	++++	+++	++++								−	−	+		+
5	m	20	Congenital	395C>T	S132L	*de novo*			+	−	+	++++	+++	++++	−	−						+		+		
6	f	23	Congenital	403T>C	S135P	*de novo*			+	+	+	++++	+++	+++								+		+		
7	f	2.6	Congenital	403T>C	S135P	*de novo*	+	+	+	+	−	++++	+++	+++	−	possible	−	+	−	−	−	−	+		+	
8	m	50	48 mos	413C>G	S138C	*de novo*	+		−	−	−	+	++	−	+	+				+	+	+	−	+		
9	m	3.6	6 mos	413C>G	S138C	*de novo*	+	−	+	−	−	+	++	+	−	+	+	+	−	−	+	+	+	+	+	
10	m	9		496A>G	T166A	*de novo*				−	+		+			+										
11	f	3	12 mos	496A>G	T166A	*de novo*	−		+			+	+	+		+	+							−		
12	m	5	10 mos	496A>G	T166A	*de novo*	−	−	−	+	+	+	+	+	−	−	−	−	−	−	−	+	−	+	−	+

No	Sex	Age	Onset	Nucleotide	Protein	Inheritance																				
13	f	9	Congenital	496 A>G	T166A	*de novo*	+	+	+	+	+	+	+++	+++	+	+		+	–	+	+	+	+	+	+	–
14	f	8	4 mos	727 G>A	G243R	*de novo*	–	–	+	+	+	+++	+++	+++	+	+		+	+	+	+	–	+	+	+	
15	m	17	36 mos	727 G>A	G243R	*de novo*	+				+	++++	+++	+++		–				+			+		+	
16	m	5	12 mos	728 G>A	G243E	*de novo*	+	+			+	++	+++	+++	–	+	+		–			+	+	+	+	
17	f	4	<12 mos	740C>T	T247I	*de novo*	+	–	+	+	–	++++	+++	+++	–	–	–	+	–	–	+	–	+			
18	m	1	<12 mos	751 A>G	T251A	*de novo*	–		+	+	–	–	+++	+++	–	+	–				–				
19				872 A>G	Y291C						+		++		+											
20	m	7	24 mos	887C>G	T296R		–	–	–	+	+	–	+	+	–	–	+	–	–	–	–	–	+		+	
21	f	12	Congenital	976 G>T	V326F	*inherited*	+	–	–	+	–	+	++	++	+	+	+	+	+	+	+	–	+			
22	f	8	Congenital	985 G>C	A329P	*de novo*	+	+	+	+	+	+	+++	++++	–	+	+	+	–	+	–	+	+	+	+	+
23	f	5	Congenital	989T>C	L330P	*de novo*	+	–	–	+	+	+	+++	+++	–	+	–		+			–	+			
24				1097 G>C	R366T						–		+	–			+						+			
25	m	3	36 mos	1141 A>G	I381V	*de novo*							+	+		+										
26	f	2.6	Congenital	1346 C>T	A449V	*inherited*		–	+	+			+	++		possible		+			–	+	+	+		+
27	f	3.6	Congenital	1499 C>T	P500L	*de novo*	+	+	+	+	–	+++	+++	+++	–	–		+	–	–	+	–	+			+
28	m	12.8	36 mos	934 G>C	G312R	*de novo*	–	–	+	–	–	+	+	+	–	–	+	–	–	–	–	–	–	+	–	–
29	f	3.6	Congenital	998C>G	L330V	*de novo*	–	–	+	+	–	+	+++	++++	+	+	–	+	–	–	–	–	+	+		–
30	m	14	6 mos	421 A>C	S141R	*de novo*	–	–	–	–	+	+	++++	++++	–	+		+	–	+			+	+		–
31	m	6	Congenital	1271 A>G	E424G	*de novo*	+	+	+	–	+	+	++	++	–	+	+	+	+	+	–	+	+	+	–	–

Supplemental Table 9 (*Chapter 7*): Genetic and phenotypic summary of other subjects with *CERT1* variants abstracted from various public access databases and not included in this project.

Subject ID	Sex	cDNA (NM_005713.3)	protein (NP_005704)	Inheritance	Zygosity	Source	Clinical phenotype
285350	-	c.-303C>T	n/a	mosaic in mother	Het	DECIPHER v9.31	abnormality of the cerebral ventricles, global developmental delay, macrocephaly, overgrowth, preauricular pit, spotty hyperpigmentation, ventriculomegaly
VCV000521299.1	-	c.-75C>T	5' UTR	de novo	Het	ClinVAR	hereditary/inborn disease
VCV000548566.1	-	c.-114G>T	5' UTR	n/a	Het	ClinVAR	Mental retardation, autosomal dominant 34 (MIM# 616351)
VCV000689370.1	-	c.-162_-152del	5' UTR	de novo	Het	ClinVAR	neurodevelopmental abnormality, microcephaly and spastic paraparesis
JASD_Fam0207	-	c.177T>G	D59E	de novo	Het	Takata, A, et al., Cell Rep., 2018	autism spectrum disorder
VCV000801357.1	-	c.349-12dup	n/a	n/a	Het	ClinVAR	Mental retardation, autosomal dominant 34 (MIM# 616351)
		c.392G>A	G131D	de novo	Het	Homsy et al., Science, 2015	Congenital heart disease
257812	M	c.395C>T	S132L	de novo	Het	de Ligt, J., et al., N. Engl. J. Med., 2012 (7)	Mental retardation, autosomal dominant 34 (MIM# 616351)
258712	M	c.395C>T	S132L	de novo	Het	de Ligt, J., et al., N. Engl. J. Med., 2012 (7)	Mental retardation, autosomal dominant 34 (MIM# 616351)
261651	M	c.395C>T	S132L	de novo	Het	de Ligt, J., et al., N. Engl. J. Med., 2012 (7)	Mental retardation, autosomal dominant 34 (MIM# 616351)
265277	F	c.404C>G	S135C	de novo	Het	DECIPHER v9.31	abnormal corpus callosum morphology, agenesis of corpus callosum, gastroesophageal reflux, global developmental delay, recurrent urinary tract infections
Fam_86872	-	c.404C>T	S135F	de novo	Het	Heldbig, K., et al., Genet. Med. 2016 (18)	epileptic encephalopathy
265676	F	c.413C>G	S138C	n/a	Het	DECIPHER v9.31	constipation, delayed speech and language development, drooling, intellectual disability, profound, Intermittent hyperventilation, postnatal microcephaly
SSC12779	F	c.496A>G	T166A	de novo	Het	De Rubeis, S., et al., Nature 2014; Kosmicki, JA., et al., Nat. Genet. 2017; Lim, ET., et al., Nat. Neurosci. 2017; Satterstrom, FK., et al., Cell 2020 (96-98)	autism spectrum disorder
DEASD_0027_001	F	c.544T>C	F182L	de novo	Het	De Rubeis, S., et al., Nature 2014 (96)	autism spectrum disorder
COL4A3BP_000005	-	c.718G>A	D240N	n/a	Het	VKGL-NL_Rotterdam	Mental retardation, autosomal dominant 34 (MIM# 616351)
264615	M	c.761G>T	G254V	de novo	Het	DECIPHER v9.31	depressed nasal bridge, epicanthus, moderate global developmental delay, single transverse palmar crease, sleep disturbance, upslanted palpebral fissure
VCV000634569.1	-	c.1301T>G	V434G	n/a	Het	ClinVAR	Mental retardation, autosomal dominant 34 (MIM# 616351)
VCV000592133.1	-	c.1335del	A446Lfs*30	n/a	Het	ClinVAR	Mental retardation, autosomal dominant 34 (MIM# 616351)

Supplemental Table 10 (*Chapter 7*): Summary of sequencing methodology and results of other genetics analyses.

Subject	Recruitment/ Source	*CERT1* variant	Sequencing approach	Capture reagent	Sequencer	Karyotype	Fragile X	aCGH	Targeted gene/panel screening/ SNP array	Metabolic screening
Subject 1	GeneMatcher	CERT1:NM_005713:c.279T>G:p.S93R	singleton-based WES (research laboratory)	Agilent SureSelect XT2 All Exon V4 (XomeAnalyzer)	HiSeq2000 (Illumina)	–	–	–	–	
Subject 2	GeneMatcher	CERT1:NM_005713:c.395C>T:p.S132L	trio-based WES (clinical diagnostic laboratory)	Agilent SureSelect XT2 All Exon V4	HiSeq2000 (Illumina)	Normal (46, XY)	–	Normal	Negative (SLC9A6, UBE3A). Subtelomeric MLPA (normal).	Negative
Subject 3	GeneMatcher	CERT1:NM_005713:c.395C>T:p.S132L	–	–	–	Normal (46, XY)	Normal (31 CGG repeats)	Normal	–	–
Subject 4	GeneMatcher	CERT1:NM_005713:c.395C>T:p.S132L	trio-based WES (clinical diagnostic laboratory)	Agilent SureSelect XT capture library	Illumina	–	–	arr[GRCh37] 7q21.12(87668378_87739123)x1, mat	ID gene panel (1000 genes): negative	–
Subject 5	GeneMatcher	CERT1:NM_005713:c.395C>T:p.S132L	trio-based WES (research laboratory)	Nextera Rapid Capture Exomes - Illumina		Normal (46, XY)	Normal (31 CGG repeats)	Normal	–	–
Subject 6	Murakami, H., et al., *Plos One* (2020)	CERT1:NM_005713:c.403T>C:p.S135P	trio-based WES (clinical diagnostic/research laboratory)	–	–	–	–	–	–	–
Subject 7	GeneMatcher	CERT1:NM_005713:c.403T>C:p.S135P	trio-based WGS (research laboratory)	-	NovaSeq 6000 (Illumina)	-	Normal (31 CGG repeats)	Normal	locus SNRPN normal	
Subject 8	de Ligt, J, et al., N. Engl. J. Med. (2012)	CERT1:NM_005713:c.413C>G:p.S138C	trio-based WES (research laboratory)	Agilent SOLiD-Optimized SureSelect Human Exome Kit (2.50Mb)	SOLiD 4 System (Life Technologies)	–	Normal (31 CGG repeats)	–	*NEK5* c.1377C>T p.(=) (benign)	–
Subject 9	GeneMatcher	CERT1:NM_005713:c.413C>G:p.S138C	trio-based WES (clinical diagnostic laboratory)	Agilent SureSelect Human All Exon V6 kit	HiSeq X® (Illumina)	–	–	–	–	–
Subject 10	Iossifov, I., et al., Nature (2014)	CERT1:NM_005713:c.496A>G:p.T166A	WES	SeqCap EZ Human Exome Library v2.0 (Roche NimbleGen)	HiSeq2000 (Illumina)	–	–	–	–	–
Subject 11	GeneMatcher	CERT1:NM_005713:c.496A>G:p.T166A	MIP	–	–	–	–	–	–	–
Subject 12	GeneMatcher	CERT1:NM_005713:c.496A>G:p.T166A	trio-based WES (clinical diagnostic/research laboratory)	Agilent SureSelect Target Enrichment v6	HiSeq2500 (Illumina)	Normal (46, XY)	Normal (31 CGG repeats)	maternally inherited duplication 10q24	Negative (SCN1A)	Negative
Subject 13	GeneMatcher	CERT1:NM_005713:c.496A>G:p.T166A	trio-based WES/WGS (clinical diagnostic/research laboratory)		TruSEQ (Illumina); NovaSeq-6000 (Illumina)	Normal (46, XX)	Normal (31 CGG repeats)	Normal	Negative (MECP2), normal methylation pattern (at 15q11-q13)	Negative
Subject 14	GeneMatcher	CERT1:NM_005713:c.727G>A:p.G243R	trio-based WES (clinical diagnostic laboratory)	Agilent SureSelect XT Human All Exon 50Mb Kit	Illumina HiSeq2000TM at BGI-Europa	–	Normal (31 CGG repeats)	Normal	Negative (MEF2C, MECP2, FOXG1, CDKL5, CLN1,CLN2)	Negative
Subject 15	Hamdan, F.F., et al., Plos Genet. (2014)	CERT1:NM_005713:c.727G>A:p.G243R	trio-based WES (research laboratory)	Agilent SureSelect v4 exome capture kit	HiSeq2000 (Illumina)	–	–	–	–	

197

Subject		Variant	Method	Capture	Sequencer					
Subject 16	GeneMatcher	CERT1:NM_005713:c.728 G>A:p.G243E	singleton-based WES (research laboratory)	Agilent SureSelect XT2 All Exon V4 (XomeAnalyzer)	HiSeq2000 (Illumina)	–	–	Normal	–	–
Subject 17	GeneMatcher	CERT1:NM_005713:c.740 C>T:p.T247I	trio-based WES (research laboratory)	NovaSeq6000 (Illumina)	SureSelect Human All Exon v7 (Agilent)	Normal (46, XX)	-	Normal	Panel for Rett/Rett-like syndrome negative (MECP2, MEF2C, CDKL5, FOXG1, SMC1A, TCF4, SHANK3, SATB2, STXBP1, SCN1A, SCN2A, SCN8A,GRIN2A, GRIN2B, HCN1, SLC6A1, WDR45, ST3GAL5, MFSD8). MLPA 15q11.2 (normal)	-
Subject 18	GeneMatcher	CERT1:NM_005713:c.751 A>G:p.T251A	MIP	–	–	–	–	–	–	
Subject 19	GeneMatcher	CERT1:NM_005713:c.872 A>G:p.Y291C	singleton-based WES (research laboratory)	Agilent SureSelect XT2 All Exon V4 (XomeAnalyzer)	HiSeq2000 (Illumina)	–	–	–	–	–
Subject 20	GeneMatcher	CERT1:NM_005713:c.887 C>G:p.T296R	singleton-based WES (clinical diagnostic laboratory)	Roche-Nimblegen SeqCap EZ MedExome	NextSeq500 (Illumina)	Normal (46, XY)	Normal (31 CGG repeats)	Normal	N/A	Negative
Subject 21	GeneMatcher	CERT1:NM_005713:c.976 G>T:p.V326F	trio-based WES (research laboratory)	Agilent Sure Select All exon V6	HiSeq4000 (Illumina)	–	–	arrXp21.2(29,584,944x2,29,594,080-20,708,100x3,29,717,001x2)dn	MECP2 negative, UBE3A negative, ID/ASD panel (74 genes) negative. Normal (chr15q methylation).	–
Subject 22	GeneMatcher	CERT1:NM_005713:c.985 G>C:p.A329P	trio-based WES (clinical diagnostic laboratory)	Truseq_v1.2_45 Mb (Illumina)	NextSeq500 (Illumina)	Normal (46, XX)	–	Normal	*MECP2* c.603G>A, p.A201A (likely benign). 2 variants in *MOCS2* (compund heterozygous). Subtelomeric MLPA (normal); SAICAR test (normal); glycosylation defects and biotinidase (normal).	–
Subject 23	GeneMatcher	CERT1:NM_005713:c.989 T>C:p.L330P	–	–	–	Normal (46, XX)	Normal (31 CGG repeats)	15q11q11.2(20262224-22751244)*3 (likely benign)	–	Negative
Subject 24	GeneMatcher	CERT1:NM_005713:c.1097 G>C:p.R366T	singleton-based WES (research laboratory)	Agilent SureSelect XT2 All Exon V4 (XomeAnalyzer)	HiSeq2000 (Illumina)	–	–	–	–	–

Subject 25	GeneMatcher	CERT1:NM_005713:c.1141 A>G:p.I381V	MIP	–	–	–	–	–	–	–
Subject 26	GeneMatcher	CERT1:NM_005713:c.1346 C>T:p.A449V	singleton-based WES (clnical diagnostic laboratory)	Kit xGen Exome Panel v1.0(IDT)	NextSeq 500 (illumina)	–	–	Normal	–	–
Subject 27	GeneMatcher	CERT1:NM_005713:c.1499 C>T:p.P500L	singleton-based WES (diagnostic laboratory)	Human Core Exome (Twist Bioscience)	HiSeq2500 (Illumina)	-	-	Normal	–	-
Subject 28	GeneMatcher	CERT1:NM_005713:c.934 G>C:p.G312R	-	-	-	-	-	-	-	-
Subject 29	GeneMatcher	CERT1:NM_005713:c.998 C>G:p.L330V	-	-	-	-	-	-	-	-
Subject 30	GeneMatcher	CERT1:NM_005713:c.421 A>C:p.S141R	-	-	-	-	-	-	-	-
Subject 31	GeneMatcher	CERT1:NM_005713:c.1271 A>G:p.E424G	-	-	-	-	-	-	-	-

Supplemental Table 11 (*Chapter 8*): Demographic, genetic and clinical characteristics of 62 patients with *PUM1*-associated disease in the study cohort.

Subject	Sex	Current age (years)	Age of onset (years)	PUM1 variant	Inheritance	Intellectual disability	Speech delay	Seizures	MRI abnormalities	EEG abnormalities	Motor delay	Ataxia	Gait abnormalities	ASD	Ocular issues	Ptosis	Short stature	Hypotonia	Cryptorchidism	Hearing issues	Cardiac issues	Gastrointestinal disturbance	Facial dysmorphia
Subject 1				Deletion (4.84 Mb)		yes	yes	yes			yes	yes											
Subject 2				Deletion (2.00 Mb)			yes				yes												
Subject 3	Male	16		Deletion (3.91 Mb)	*de novo*	yes	yes	no	yes		yes	no		no	yes		yes		yes				yes
Subject 4				Deletion (0.28 Mb)		yes	yes	yes	yes		yes						yes						yes
Subject 5				Deletion (0.59 Mb)		yes	yes																
Subject 6				Deletion (2.59 Mb)		yes	yes				yes												
Subject 7	Male	2		Deletion (1.78 Mb)				yes			yes	yes					yes	yes	yes				yes
Subject 8	Female	0.5		Deletion (5.60 Mb)	*de novo*	yes	yes	no	yes		yes	yes					no	yes					yes
Subject 9	Female	7		Deletion (2.05 Mb)	*de novo*	yes	yes		WNL		yes	yes					no	yes			yes		yes
Subject 10	Female	8	2.5	p.(Arg1139Trp)	*de novo*	no	no		WNL		yes	yes			no		yes	yes					no
Subject 11	Female	9	0.5	p.(Arg1147Trp)	*de novo*	yes	yes	yes	yes		yes	yes			yes	yes	yes	yes					yes
Subject 12	Male	2.5	0.25	p.(Arg1147Trp)	*de novo*	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	no	yes	yes
Subject 13	Female	10		p.(Arg1147Trp)			yes	yes									yes	yes					yes
Subject 14	Male	17	3	p.(Arg1147Trp)	*de novo*	yes	yes	yes	yes		yes				no	yes	yes	yes	yes				yes
Subject 15	Male	1.5	0.3	Deletion (3.53 Mb)	*de novo*	yes	yes	yes		yes	yes	no			no		no						yes
Subject 16	Female	5	1.5	Deletion (1.95 Mb)		yes	yes	no	yes		yes	no	yes	no	no		no	yes	no	no	no	no	no
Subject 17	Male	14.5	0.1	Deletion (4.35 Mb)	de novo	yes	yes	yes	yes		yes	yes	yes				no		no				yes
Subject 18	Female	9		c.1158+1_1158+2dup	(not maternal)	yes	yes	yes			no	no	no		no		no		no				yes
Subject 19	Male	4.5	0	c.363+1G>A	Affected mother (S20)	yes	yes	no	yes		yes	yes	yes	no	no	yes	yes	yes	yes	no	yes		yes
Subject 20	Female	31	0	c.363+1G>A	*de novo*	Moderate	yes	yes			yes	no	yes	yes	no	yes	yes	yes	no	no	yes	no	yes

Subject	Sex			Variant	Inheritance																		
Subject 21	Female	9		p.(Arg1147Trp)	de novo	yes		yes				yes											
Subject 22	Male	49	35	p.(Arg1139Trp)	de novo	no			yes			yes	yes		yes								
Subject 23	Female	51	12	p.(Lys151Arg)	de novo	no			yes			yes	yes		yes								
Subject 24	Female	.		p.(Asn973Asp)		Moderate		no				no											
Subject 25	Male	5	5	p.(Arg818Gln)		yes		yes			yes		yes										no
Subject 34	Female	74	52	p.(Cys302Tyr)		no	no	no	yes		no	yes			yes								no
Subject 35				p.(Glu1036Gln)		no																	
Subject 36		0.1		p.(Arg580*)	de novo																yes		
Subject 37	Male	15	0.1	p.(Arg1147Trp)	de novo	yes		yes	yes	yes		no			yes		yes	yes	yes				yes
Subject 38	Male	18	1	p.(Arg837*)	de novo	yes	yes	yes	WNL			yes		yes				no		no				yes
Subject 40	Male	3	1.5	p.(Glu792Lys)	de novo	yes	yes	no			yes	no	no	yes	no		no	No	no	yes	no	no		yes
Subject 41	Female	6	0.1	Deletion (12.51 kb)	Affected mother (S41.1)	yes	yes	yes	yes		yes	yes		no	yes			no	yes		yes			yes
Subject 41.1	Female	34	1.5	Deletion (12.51 kb)	de novo	Mild	no	yes	yes		no	yes		no	no			no			no	no	no	no
Subject 44	Female	4		p.(Asp16Tyr)	de novo	yes	yes	no	yes		yes	yes		yes	yes			no	yes		no	no	no	yes
Subject 45	Male	37	33	p.(His1050Tyr)	Unaffected father			no	yes			yes		no	yes									no
Subject 46	Female	62	42	p.(Met746Val)	Unaffected mother	No	no	no	yes		No	yes	yes	no	no			no	no		no	no	yes	no
Subject 47	Female	4		p.(Ala52Gly)	Unaffected mother	Mild						yes										yes		
Subject 48	Male	6	0.1	p.(His1090Profs*16)	Affected father (S50)	Mild	yes	yes	WNL	yes	yes	no		no				no	yes		no	no	yes	yes
Subject 49	Male	6	2	p.(His1090Profs*16)	Affected father (S50)	no	yes	no			no	no		no				no	mild		no	no	no	yes
Subject 50	Male	41	8	p.(His1090Profs*16)		Mild	no	no			no	no		no	no	yes		no	no		no			yes
Subject 51	Male	4	0.1	p.(Pro66Leu)	de novo	yes	yes	no	yes			yes		no	no				yes	yes		no	no	yes
Subject 52	Female	13.2	0	p.(Arg840Trp)	de novo	yes					yes	yes	yes	no	yes			no			no	yes	no	yes

Subject 53	Female	59	30	p.(Thr1035Ser)	Affected father (S56)	No			yes			yes			yes								
Subject 54	Female	58	30	p.(Thr1035Ser)	Affected father (S56)	No			yes			yes			yes								
Subject 55	Female	52	40	p.(Thr1035Ser)	Affected father (S56)	No			yes			yes			yes								
Subject 56	Male	81	50	p.(Thr1035Ser)	Affected father (S56)	No						yes											
Subject 58	Male	2.5	0.3	Deletion (2.43 Mb)	de novo			yes		WNL				yes									
Subject 59	Male	67	61	p.(Ile417Val)			yes		yes			yes		no	no			yes	no	no	no	yes	no
Subject 60	Male	2.5	1.5	p.(Asp78Asn)	de novo						yes	no		no	yes		no		no	no	no	no	no
Subject 61	Female	43	10	p.(Gly95Arg)	de novo				yes			yes	yes										
Subject 62	Male	69	48	p.(Gly647=)	Affected father				yes			yes	yes										
Subject 63	Female	5.5	1.5	p.(His1090Profs*16)	de novo	yes	Yes	yes	yes	yes	yes	yes	yes	no	yes		no	yes	no	no	no	no	no
Subject 64	Male	3		p.(Val396Ile)	Affected mother	yes					yes	yes			yes								
Subject 66		11		p.(Pro299His)	Unaffected father	yes		yes	yes			yes	yes										
Subject 67	Female	3	0.1	Duplication (9.05 kb)	de novo	yes	yes	yes	yes			yes						yes					
Subject 68	Male	20	8	Deletion (2.59 Mb)	de novo	yes	yes	yes	yes	yes											yes	yes	yes
Subject 69	Female	1.5	0	Deletion (0.49 Mb)	Affected father (S70)	yes		yes	WNL		yes	no		no	yes		no	yes			yes	yes	yes
Subject 70	Male	24	8	Deletion (0.49 Mb)		no		yes				yes	yes	no			no	yes					yes
Subject 71	Male	4	3.5	p.(Leu387Cysfs*13)		no	yes	yes	WNL	yes	no	no	no	no	no		no		no	no	yes	yes	no
Subject 72	Female	3	0.3	p.(Arg1147Trp)	de novo	yes	yes	yes	yes	yes	yes					yes		yes					yes
Subject 73	Female	7	1	p.(Gln879Arg)		no	yes	no	WNL		no	no		no	yes		no		no	yes	no	no	no
Subject 74	Female	6.5	0.5	p.(Met955Ile)	de novo	no	mild	no	WNL		yes	yes	yes	no	yes					no		no	no

Supplemental Table 12 (*Chapter 8*): The genomic coordinates, size and encompassed genes of the copy-number variants (CNV) identified in patients with PUM1-associated disease.

Subject	Inheritance	Genomic coordinates	Size (bp)	STRVCTVRE score	Genes included in CNV
Subject 70	*(unknown)*	chr1:31394262:31883885:DEL	489624	0.820	*FABP3, NKAIN1, **PUM1**, SERINC2, SNRNP40, ZCCHC17*
Subject 69	Paternal	chr1:31394262:31883885:DEL	489624	0.820	*FABP3, NKAIN1, **PUM1**, SERINC2, SNRNP40, ZCCHC17*
Subject 68	*de novo*	chr1:31241410:33828389:DEL	2586980	0.945	*A3GALT2, ADGRB2, **AK2**, AZIN2, BSDC1, CCDC28B, COL16A1, DCDC2B, EIF3I, ENSG00000212673, ENSG00000267885, ENSG00000268950, FABP3, FAM167B, FAM229A, FNDC5, HCRTR1, HDAC1, **HPCA**, IQCC, KHDRBS1, KIAA1522, KPNA6, **LCK**, MARCKSL1, NKAIN1, PEF1, PHC2, PTP4A2, **PUM1**, RBBP4, RNF19B, S100PBP, SDC3, SERINC2, SNRNP40, SPOCD1, SYNC, TINAGL1, TMEM39B, TMEM54, TMEM234, TRIM62, TSSK3, TXLNA, YARS1, ZBTB8A, ZBTB8B,ZBTB8OS, ZCCHC17, ZNF362*
Subject 67	*de novo*	chr1:31409393:31418428:DUP	9036	0.836	***PUM1***
Subject 58	*de novo*	chr1:29265023:31691213:DEL	2426191	0.813	***EPB41**, LAPTM5, MATN1, **MECR**, NKAIN1, PTPRU, **PUM1**, SDC3, SRSF4, TMEM200B*
Subject 41.1	*de novo*	chr1:31409179:31421689:DEL	12511	0.839	***PUM1***
Subject 41	Maternal	chr1:31409179:31421689:DEL	12511	0.839	***PUM1***
Subject 17	*de novo*	chr1:28666871:33012401:DEL	4345531	1.000	*ADGRB2, BSDC1, CCDC28B, COL16A1, DCDC2B, EIF3I, ENSG00000212673, ENSG00000268950, **EPB41**, FABP3, FAM167B, FAM229A, GMEB1, HCRTR1, HDAC1, IQCC, KHDRBS1, KPNA6, LAPTM5, **LCK**, MARCKSL1, **MECR**, NKAIN1, OPRD1, PEF1, PHACTR4, PTP4A2, PTPRU, **PUM1**, RAB42, RCC1, SDC3, SERINC2, SNRNP40, SPOCD1, SRSF4, TAF12, TINAGL1, TMEM39B, TMEM200B, TMEM234, TRNAU1AP, TSSK3, TXLNA, YTHDF2, ZBTB8A, ZBTB8B, ZCCHC17*
Subject 16	*(unknown)*	chr1:30199705:32145611:DEL	1945907	0.882	*COL16A1, ENSG00000268950, FABP3, HCRTR1, LAPTM5, MATN1, NKAIN1, PEF1, **PUM1**, SDC3, SERINC2, SNRNP40, TINAGL1, ZCCHC17*
Subject 15	*de novo*	chr1:28835332:32362980:DEL	3527649	1.000	*ADGRB2, COL16A1, ENSG00000268950, **EPB41**, FABP3, GMEB1, HCRTR1, LAPTM5, MATN1, **MECR**, NKAIN1, OPRD1, PEF1, PTPRU, **PUM1**, RAB42, RCC1, SDC3, SERINC2, SNRNP40, SPOCD1, SRSF4, TAF12, TINAGL1, TMEM200B, TRNAU1AP, YTHDF2, ZCCHC17*
Subject 9	*de novo*	chr1:31091243:33142346:DEL	2051104	0.947	*ADGRB2, BSDC1, CCDC28B, COL16A1, DCDC2B, EIF3I, ENSG00000212673, ENSG00000268950, FABP3, FAM167B, FAM229A, HCRTR1, HDAC1, IQCC, KHDRBS1, KPNA6, LAPTM5, **LCK**, MARCKSL1, MATN1, NKAIN1, PEF1, PTP4A2, **PUM1**, RBBP4, SDC3, SERINC2, SNRNP40, SPOCD1, TINAGL1, TMEM39B, TMEM234, TSSK3, TXLNA, ZBTB8A, ZBTB8B, ZBTB8OS, ZCCHC17*
Subject 8	*de novo*	chr1:28743173:34340430:DEL	5597258	1.000	*A3GALT2, ADGRB2, **AK2**, AZIN2, BSDC1, CCDC28B, COL16A1, CSMD2, DCDC2B, EIF3I, ENSG00000212673, ENSG00000267885, ENSG00000268950, **EPB41**, FABP3, FAM167B, FAM229A, FNDC5, GMEB1, HCRTR1, HDAC1, HMGB4, **HPCA**, IQCC, KHDRBS1, KIAA1522, KPNA6, LAPTM5, **LCK**, MARCKSL1, MATN1, **MECR**, NKAIN1, OPRD1, PEF1, PHACTR4, PHC2, PTP4A2, PTPRU, **PUM1**, RAB42, RBBP4, RCC1, RNF19B, S100PBP, SDC3, SERINC2, SNRNP40, SPOCD1, SRSF4, SYNC, TAF12, TINAGL1, TMEM39B, TMEM54, TMEM200B, TMEM234, TRIM62, TRNAU1AP, TSSK3, TXLNA, YARS1, YTHDF2, ZBTB8A, ZBTB8B, ZBTB8OS, ZCCHC17, ZNF362, ZSCAN20*
Subject 7	*(unknown)*	chr1:31113947:32897001:DEL	1783055	0.949	*ADGRB2, BSDC1, CCDC28B, COL16A1, DCDC2B, EIF3I, ENSG00000212673, ENSG00000268950, FABP3, FAM167B, FAM229A, HCRTR1, HDAC1, IQCC, KHDRBS1, KPNA6, LAPTM5, **LCK**, MARCKSL1, MATN1, NKAIN1, PEF1, PTP4A2, **PUM1**, SDC3, SERINC2, SNRNP40, SPOCD1, TINAGL1, TMEM39B, TMEM234, TSSK3,TXLNA, ZCCHC17*
Subject 6	*(unknown)*	chr1:31239605:33825029:DEL	2585425	0.945	*A3GALT2, ADGRB2, **AK2**, AZIN2, BSDC1, CCDC28B, COL16A1, DCDC2B, EIF3I, ENSG00000212673, ENSG00000267885, ENSG00000268950, FABP3, FAM167B, FAM229A, FNDC5, HCRTR1, HDAC1, **HPCA**, IQCC, KHDRBS1, KIAA1522, KPNA6, **LCK**, MARCKSL1, NKAIN1, PEF1, PHC2, PTP4A2, **PUM1**, RBBP4, RNF19B, S100PBP, SDC3, SERINC2, SNRNP40, SPOCD1, SYNC, TINAGL1, TMEM39B, TMEM54, TMEM234, TRIM62, TSSK3, TXLNA, YARS1, ZBTB8A, ZBTB8B, ZBTB8OS, ZCCHC17, ZNF362*
Subject 5	*(unknown)*	chr1:31284806:31872758:DEL	587953	0.824	*FABP3, NKAIN1, **PUM1**, SDC3, SNRNP40, ZCCHC17*
Subject 4	*(unknown)*	chr1:31442430:31720099:DEL	277670	0.779	*NKAIN1, **PUM1***
Subject 3	*de novo*	chr1:28716929:32629424:DEL	3912496	1.000	*ADGRB2, COL16A1, ENSG00000212673, ENSG00000268950, **EPB41**, FABP3, GMEB1, HCRTR1, KHDRBS1, KPNA6, LAPTM5, MATN1, **MECR**, NKAIN1, OPRD1, PEF1, PHACTR4, PTP4A2, PTPRU, **PUM1**, RAB42, RCC1, SDC3, SERINC2, SNRNP40, SPOCD1, SRSF4, TAF12, TINAGL1, TMEM39B, TMEM200B, TRNAU1AP, YTHDF2, ZCCHC17*

Subject 2	*(unknown)*	chr1:30073835:32064524:DEL	1990690	0.837	*ENSG00000268950, FABP3, LAPTM5, MATN1, NKAIN1,* **PUM1,** *SDC3, SERINC2, SNRNP40, TINAGL1, ZCCHC17*
Subject 1	*(unknown)*	chr1:28751378:33588455:DEL	4837078	1.000	*ADGRB2,* **AK2,** *AZIN2, BSDC1, CCDC28B, COL16A1, DCDC2B, EIF3I, ENSG00000212673, ENSG00000267885, ENSG00000268950,* **EPB41,** *FABP3, FAM167B, FAM229A, FNDC5, GMEB1, HCRTR1, HDAC1,* **HPC***A, IQCC, KHDRBS1, KIAA1522, KPNA6, LAPTM5,* **LCK,** *MARCKSL1, MATN1,* **MECR,** *NKAIN1, OPRD1, PEF1, PHACTR4, PTP4A2, PTPRU,* **PUM1,** *RAB42, RBBP4, RCC1, RNF19B, S100PBP, SDC3, SERINC2, SNRNP40, SPOCD1, SRSF4, SYNC, TAF12, TINAGL1, TMEM39B, TMEM54, TMEM200B, TMEM234, TRNAU1AP, TSSK3, TXLNA, YARS1, YTHDF2, ZBTB8A, ZBTB8B, ZBTB8OS, ZCCHC17*

Supplemental Table 13 (*Chapter 8*): Pathogenicity analysis of single nucleotide variants (SNV) and small insertion-deletions identified within the *PUM1* locus.

cDNA (NM_001020658.2)	Protein	Location	Allele count (study cohort)	Allele frequency (gnomAD exomes)	Allele count (gnomAD exomes)	phyloP100	M-CAP	REVEL	Eigen	CADD-PHRED	SpliceAI
c.3415C>T	p.(Arg1139Trp)	Exon 21	2	*(not found)*	-	3.296	0.067	0.552	0.385	25.1	-
c.3439C>T	p.(Arg1147Trp)	Exon 22	7	-	-	7.766	0.072	0.593	0.943	32	-
c.363+1G>A	p.(?)	Intron 2	1 (2)	*(not found)*	-	6.803	-	-	1.173	34	SpliceAI Δscore = 99%, donor loss at +1bp
c.452A>G	p.(Lys151Arg)	Exon 4	1	0.00003254	8	7.803	0.02	0.265	0.753	27.7	-
c.2917A>G	p.(Asn973Asp)	Exon 18	1	*(not found)*	-	7.871	0.041	0.498	0.603	25.6	-
c.2453G>A	p.(Arg818Gln)	Exon 15	1	*(not found)*	-	7.820	0.017	0.251	0.488	26.6	-
c.905G>A	p.(Cys302Tyr)	Exon 7	1	0.000004096	1	5.558	0.016	0.103	0.576	28.7	-
c.3106G>C	p.(Glu1036Gln)	Exon 19	1	0.000004063	1	7.778	0.01	0.223	0.477	24.7	-
c.2509C>T	p.(Arg837*)	Exon 15	1	0.00000398	1	2.422	-	-	0.957	38	-
c.2374G>A	p.(Glu792Lys)	Exon 15	1	*(not found)*	-	7.808	0.03	0.261	0.812	32	-
c.46G>T	p.(Asp16Tyr)	Exon 2	1	*(not found)*	-	7.487	0.091	0.457	0.456	28.4	-
c.3148C>T	p.(His1050Tyr)	Exon 20	1	*(not found)*	-	9.797	0.014	0.54	0.585	28.7	-
c.2236A>G	p.(Met746Val)	Exon 14	1	0.00000398	1	8.042	0.007	0.245	0.315	23.3	-
c.155C>G	p.(Ala52Gly)	Exon 2	1	0.00002437	6	4.052	0.003	0.063	0.139	22.5	-
c.197C>T	p.(Pro66Leu)	Exon 2	1	*(not found)*	-	5.615	0.004	0.05	0.016	22.7	-
c.2518C>T	p.(Arg840Trp)	Exon 15	1	*(not found)*	-	10.003	0.018	0.557	1.051	32	-
c.3103A>T	p.(Thr1035Ser)	Exon 19	1 (4)	0.000008127	1	6.174	0.011	0.142	0.095	23	-
c.1249A>G	p.(Ile417Val)	Exon 8	1	0.0000239	6	3.971	0.007	0.094	0.019	22.8	-
c.232G>A	p.(Asp78Asn)	Exon 2	1	*(not found)*	-	7.320	0.019	0.15	0.491	28.3	-
c.283G>A	p.(Gly95Arg)	Exon 2	1	*(not found)*	-	5.190	0.012	0.23	0.647	27.6	-
c.1941C>T	p.(Gly647=)	Exon 13	1	*(not found)*	-	2.662	-	-	-	16.41	SpliceAI Δscore = 91%, donor gain at +2bp
c.1186G>A	p.(Val396Ile)	Exon 8	1	0.00000813	2	5.542	0.015	0.139	0.573	24.6	-
c.896C>A	p.(Pro299His)	Exon 7	1	*(not found)*	-	9.133	0.016	0.393	0.901	28.3	-
c.1158+1_1158+2dup	p.(?)	Intron 7	1	*(not found)*	-	8.871	-	-	-	-	SpliceAI Δscore = 100%, donor loss at +2bp
c.1738C>T	p.(Arg580*)	Exon 12	1	*(not found)*	-	4.086	-	-	0.996	39	-
c.3267_3270del	p.(His1090Profs*16)	Exon 21	2 (4)	*(not found)*	-	9.873	-	-	-	-	*(no predicted effect)*
c.1159del	p.(Leu387Cysfs*13)	Exon 8	1	*(not found)*	-	2.126	-	-	-	-	SpliceAI Δscore = 100%, acceptor loss at +1bp SpliceAI Δscore = 100%, acceptor gain at 0bp
c.2636A>G	p.(Gln879Arg)	Exon 16	1	*(not found)*	-	7.871	0.0108	-	0.552	24.4	-
c.2865G>A	p.(Met955Ile)	Exon 18	1	*(not found)*	-	7.653	0.01012	-	-0.273	23.1	-

Supplemental Table 14 (*Chapter 8*): Analysis of probability the synonymous variant, c.1941C>T, p.(Gly647=) causes skipping of exon 13 based on the Ex-SKIP algorithm.

Sequence	PESS	FAS-ESS	FAS-ESS	IIE	IIE	NI-ESS	NI-ESS	PESE	RESCUE	EIE	EIE	NI-ESE	NI-ESE	ESS	ESE	ESS/ESE
		hex2	hex3			trusted	all		-ESE			trusted	all			
	(count)	(count)	(count)	(count)	(sum)	(count)	(sum)	(count)	(count)	(count)	(sum)	(count)	(sum)	(total)	(total)	(ratio)
WT	1	2	1	5	93.867	6	-7.511	3	3	22	284.869	21	27.794	15	49	**0.31**
MUT	1	2	1	5	93.867	6	-8.583	3	3	17	235.476	20	25.544	15	43	**0.35**

Supplemental Table 15 (*Chapter 8*): Residual Variation Intolerance Score (RVIS) analysis of all genes encompassed within large deletions identified in patients in the study cohort.

Gene	RVIS	%ExAC RVIS	ExAC LoF FDR	%ExAC v2 RVIS	Edge Case (%OE-ratio)
A3GALT2	NA (NA)	NA	0.99935	NA (NA)	NA (NA)
ADGRB2	-1.52 (3.47%)	5.68%	0.000113	-0.9966 (13.2942%)	N (12.2793%)
AK2	0.22 (68.13%)	93.11%	0.438909	0.8556 (83.0890%)	N (52.9728%)
AZIN2	7.61E-05 (53.98%)	62.57%	0.133098	0.2449 (61.0655%)	N (69.3973%)
BSDC1	-0.47 (23.25%)	30.61%	0.025544	-0.2590 (37.4976%)	N (72.6452%)
CCDC28B	0.26 (70.06%)	52.89%	1	0.2688 (62.0821%)	N (83.1297%)
COL16A1	0.58 (82.18%)	69.13%	0.001914	0.7012 (78.7977%)	N (63.0637%)
DCDC2B	0.26 (70.44%)	83.80%	57.63%	0.3209 (64.3695%)	N (46.2977%)
EIF3I	-0.16 (41.25%)	20.22%	28.48%	-0.5819 (24.6530%)	N (2.0205%)
FABP3	0.19 (66.57%)	56.36%	40.07%	-0.0800 (45.6598%)	Y (61.9232%)
FAM167B	NA (NA)	75.91%	91.88%	0.5015 (71.7986%)	Y (63.0695%)
FAM229A	NA (NA)	NA	NA	NA (NA)	NA (NA)
FNDC5	-0.25 (35.42%)	NA	41.66%	0.0923 (53.9883%)	Y (51.0855%)
HCRTR1	0.15 (64.74%)	62.88%	13.09%	0.4289 (68.8759%)	N (24.7496%)
HDAC1	-0.69 (14.97%)	17.40%	5.52%	-0.7644 (19.0909%)	N (8.6957%)
HPCA	-0.1 (46.2%)	32.05%	55.43%	-0.2537 (37.7419%)	Y (16.7661%)
IQCC	7.61E-05 (53.98%)	80.71%	26.14%	0.7761 (81.0166%)	N (74.8162%)
KHDRBS1	-0.25 (35.42%)	18.44%	0.138107	-0.6188 (23.3822%)	N (12.2214%)
KIAA1522	0.43 (77.33%)	47.06%	0.000496	0.7071 (79.0029%)	N (33.6826%)
KPNA6	-0.45 (24%)	20.50%	0.63%	-0.6762 (21.6813%)	N (5.0426%)
LCK	0.44 (77.7%)	53.32%	0.08%	-0.2404 (38.3284%)	N (5.7662%)
MARCKSL1	-0.27 (33.97%)	53.48%	24.91%	0.1079 (54.6823%)	Y (95.8895%)
NKAIN1	0.06 (58.26%)	57.27%	20.16%	0.1695 (57.6931%)	Y (17.3450%)
PEF1	-0.27 (34.32%)	42.86%	13.02%	0.0163 (50.2444%)	N (39.0146%)
PHC2	-0.48 (22.78%)	14.84%	3.23%	-0.8928 (15.9042%)	N (30.9500%)
PTP4A2	-0.01 (52.85%)	48.02%	22.54%	-0.1283 (43.3236%)	Y (5.8878%)
PUM1	-1.02 (8.1%)	2.45%	1.64E-07	-1.7581 (4.2815%)	N (8.5683%)
RBBP4	-0.23 (36.86%)	28.85%	8.58%	-0.3963 (31.6813%)	Y (0.5095%)
RNF19B	-0.78 (12.77%)	11.75%	23.82%	-0.8275 (17.4585%)	N (31.2279%)
S100PBP	-0.29 (33.2%)	28.43%	0.04%	-0.5240 (26.6471%)	N (46.0082%)
SDC3	-0.02 (52.25%)	81.99%	15.77%	0.7621 (80.6256%)	N (39.1015%)
SERINC2	0.18 (66.17%)	87.57%	61.42%	1.0325 (86.8817%)	N (94.9227%)
SNRNP40	-0.36 (28.63%)	21.53%	4.75%	-0.8355 (17.2239%)	N (18.6765%)
SPOCD1	1.45 (95.15%)	49.29%	4.85%	0.3199 (64.3109%)	N (58.4612%)
SYNC	1.48 (95.31%)	NA	32.93%	NA (NA)	NA (NA)
TINAGL1	-0.69 (15.12%)	23.85%	6.41%	-0.3887 (31.9453%)	N (60.6669%)
TMEM39B	-0.82 (11.68%)	41.10%	12.61%	-0.1848 (40.6647%)	N (55.5260%)
TMEM54	0.06 (58.74%)	70.43%	42.42%	0.6692 (77.6735%)	N (36.8321%)

TMEM234	0.55 (81.22%)	53.77%	64.01%	0.1151 (55.0733%)	Y (48.3877%)
TRIM62	-0.31 (32.06%)	37.64%	74.67%	-0.1765 (41.0459%)	N (52.2318%)
TSSK3	-0.27 (34.6%)	53.39%	0.057689	-0.2472 (38.0059%)	N (77.2361%)
TXLNA	-0.29 (33.34%)	59.33%	0.02%	-0.2776 (36.6764%)	N (18.0918%)
ZBTB8A	-0.49 (22.36%)	56.41%	5.49%	0.0418 (51.5054%)	N (21.4149%)
ZBTB8B	0.13 (62.74%)	NA	75.05%	NA (NA)	NA (NA)
ZCCHC17	0.06 (58.26%)	33.42%	5.47%	-0.3111 (35.2395%)	N (60.9564%)
EPB41	-0.6 (18.19%)	49.05%	0.09%	0.1115 (54.8778%)	N (37.6136%)
LAPTM5	0.28 (71.41%)	78.26%	6.29%	0.4239 (68.6510%)	N (48.7582%)
MATN1	-0.75 (13.58%)	64.27%	48.68%	0.4943 (71.5445%)	N (52.0292%)
MECR	0.71 (85.63%)	56.41%	10.23%	-0.0453 (47.2630%)	N (46.1240%)
PTPRU	-1.78 (2.28%)	1.86%	2.39E-05	-1.4210 (6.7840%)	N (28.9874%)
SRSF4	0.06 (58.85%)	34.43%	0.23%	0.0315 (51.0068%)	N (7.7288%)
TMEM200B	NA (NA)	NA	60.57%	NA (NA)	NA (NA)
GMEB1	-0.47 (23.25%)	12.19%	0.96%	-0.9443 (14.4282%)	N (10.7393%)
OPRD1	-0.14 (43.29%)	25.06%	0.767051	-0.0947 (44.9853%)	N (66.8500%)
PHACTR4	-0.44 (24.53%)	29.33%	0.015068	-0.1943 (40.2933%)	N (40.5083%)
RAB42	0.72 (85.92%)	57.58%	0.994676	0.1296 (55.7674%)	Y (95.6001%)
RCC1	-0.56 (19.54%)	14.17%	0.098523	-0.7533 (19.4721%)	N (16.0250%)
TAF12	0.17 (65.33%)	64.03%	0.085929	0.1509 (56.8035%)	Y (8.1920%)
TRNAU1AP	-0.23 (37.11%)	25.47%	0.009062	-0.3781 (32.4047%)	N (11.3298%)
YTHDF2	-0.01 (53.51%)	12.07%	0.000627	-0.9131 (15.2981%)	N (11.6714%)
ZBTB8OS	0.13 (62.74%)	60.09%	0.008074	0.1109 (54.8387%)	Y (83.1355%)
CSMD2	-1.92 (1.92%)	0.94%	NA	-1.7400 (4.4086%)	N (NA)
HMGB4	0.68 (85.04%)	78.10%	0.697905	0.7247 (79.6285%)	N (92.6533%)
ZNF362	0.04 (56.92%)	25.65%	0.076676	-0.5450 (25.8944%)	N (5.4247%)
ZSCAN20	-0.53 (20.91%)	14.51%	0.001005	-0.4277 (30.4203%)	N (31.0369%)

Appendix C: Statistical Calculations

Prognosis categories in *ABCA4*: age, age of onset and best-corrected visual acuity difference comparisons (Chapter 5)

Summary statistics and comparison of age (years) at examination between patients in Prognosis categories of ABCA4 disease:

Prognosis	Ntrial	Mean	SD	Median	MAD	IQR	Trimean
1	28	57.9	7.6	55.0	6.9	14.2	56.6
2	31	61.4	9.7	60.3	12.0	16.3	60.5
3	20	59.1	8.4	56.1	8.3	13.6	57.9
4	33	61.8	8.6	60.0	10.7	13.9	61.0

Abbreviations: SD, standard deviation; MAD, median absolute deviation; IQR, interquartile range.

One-Way ANOVA Test between Prognosis 1, 2, 3 and 4

F-ratio	P-value
1.37478	0.254411

Kruskal-Wallis Test between Prognosis 1, 2, 3 and 4

H-statistic	P-value
4.4582	0.21605

Post Hoc Tukey HSD

Pair-wise comparisons	P-value
Prognosis 1 : Prognosis 2	0.43374
Prognosis 1 : Prognosis 3	0.95585
Prognosis 1 : Prognosis 4	0.34532
Prognosis 2 : Prognosis 3	0.75007
Prognosis 2 : Prognosis 4	0.99863
Prognosis 3 : Prognosis 4	0.65598

Summary statistics and comparison of age of symptomatic onset between patients in Prognosis categories of ABCA4 disease:

Prognosis	Ntrial	Mean	SD	Median	MAD	IQR	Trimean
1	24	41.2	11.4	44.5	8.2	14.2	43.3
2	29	40.9	12.1	45.0	10.4	17.0	42.2
3	17	31.5	17.6	30.0	23.7	33.0	30.8
4	32	17.1	9.6	17.0	10.4	10.2	16.1

Abbreviations: SD, standard deviation; MAD, median absolute deviation; IQR, interquartile range.

One-Way ANOVA Test between Prognosis 1, 2, 3 and 4

F-ratio	P-value
25.0399	< 0.00001

Kruskal-Wallis Test between Prognosis 1, 2, 3 and 4

H-statistic	P-value
43.1082	< 0.00001

Post Hoc Tukey HSD

Pair-wise comparisons	P-value
Prognosis 1 : Prognosis 2	0.99989
Prognosis 1 : Prognosis 3	**0.03759**
Prognosis 1 : Prognosis 4	**<0.0001**
Prognosis 2 : Prognosis 3	**0.04461**
Prognosis 2 : Prognosis 4	**<0.0001**
Prognosis 3 : Prognosis 4	**0.00061**

*Statistically significant (P<0.05) values in **bold***

Summary statistics and comparison of best-corrected visual acuity (logMAR equivalent) of the "better eye" between patients in Prognosis categories of ABCA4 disease.

Prognosis	Ntrial	Mean	SD	Median	MAD	IQR	Trimean
1	26	0.4	0.4	0.5	0.6	0.8	0.4
2	31	0.6	0.5	0.6	0.6	0.9	0.6
3	18	0.7	0.7	0.5	0.6	1.0	0.6
4	33	1.6	0.6	1.3	0.4	0.7	1.5

Abbreviations: SD, standard deviation; MAD, median absolute deviation; IQR, interquartile range.

One-Way ANOVA Test between Prognosis 1, 2, 3 and 4

F-ratio	P-value
26.94054	< 0.00001

Kruskal-Wallis Test between Prognosis 1, 2, 3 and 4

H-statistic	P-value
46.847	< 0.00001

Post Hoc Tukey HSD

Pair-wise comparisons	P-value
Prognosis 1 : Prognosis 2	0.55288
Prognosis 1 : Prognosis 3	0.40099
Prognosis 1 : Prognosis 4	**<0.0001**
Prognosis 2 : Prognosis 3	0.99452
Prognosis 2 : Prognosis 4	**<0.0001**
Prognosis 3 : Prognosis 4	**<0.0001**

*Statistically significant (P<0.05) values in **bold***

Distribution of allele combinations across ABCA4 prognosis categories (Chapter 5)

2x3 Fisher Exact Test for count data comparing distributions of mild, moderate and severe allele combinations across Prognosis categories:

Mild allele combination cases

Allele 1	Mild	Prognosis 1	Prognosis 2	Prognosis 3	Prognosis 4
	Mild	0	0	0	0
Allele 2	Moderate	5	1	0	0
	Severe	17	23	6	0

Two-sided P-value	Two-sided P-value (Monte Carlo simulation based on 200 replicates)
0.1177	0.1164

Moderate allele combination cases

Allele 1	Moderate	Prognosis 1	Prognosis 2	Prognosis 3	Prognosis 4
	Mild	5	1	0	0
Allele 2	Moderate	3	1	2	1
	Severe	3	6	11	17

Two-sided P-value	Two-sided P-value (Monte Carlo simulation based on 200 replicates)
0.0007353	**0.001499**

Severe allele combination cases

Allele 1	Severe	Prognosis 1	Prognosis 2	Prognosis 3	Prognosis 4
	Mild	17	23	6	0
Allele 2	Moderate	3	6	11	17
	Severe	0	0	1	15

Two-sided P-value	Two-sided P-value (Monte Carlo simulation based on 200 replicates)
-	**0.0004998**

Comparison of atrophy classifications in *ABCA4*: DDAF, h-DDAF and others (Chapter 6)

Summary statistics and comparison of age (years) at examination between h-DDAF, DDAF and Other (QDAF or none).

Type	Ntrial	Mean	SD	Median	MAD	IQR	Trimean
DDAF	186	51.7	14.4	52.1	14.6	19.7	52.2
h-DDAF	30	27.6	11.3	25.8	11.8	16.6	26.7
Other	199	28.6	13.3	26.0	12.5	16.9	26.5

Abbreviations: SD, standard deviation; MAD, median absolute deviation; IQR, interquartile range.
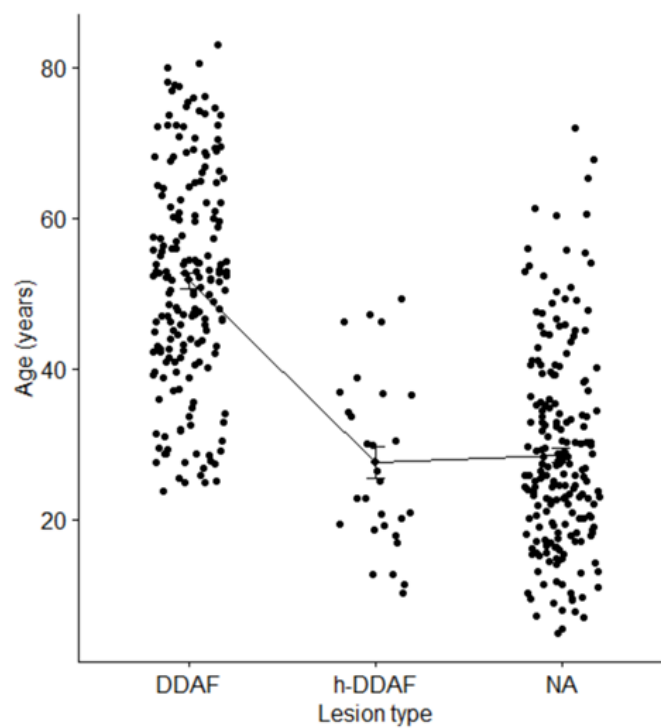
One-Way ANOVA Test between h-DDAF, DDAF and Other (QDAF or none)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lesion type	2	55632	27816	152.2	148.9 <2e-16 ***
Residuals	412	76963	187		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Post Hoc Tukey HSD

	diff	lwr	upr	p adj
h-DDAF - DDAF	-24.1100086	-30.435280	-17.784737	0.000000
Other - DDAF	-23.1509551	-26.429769	-19.872141	0.000000
Other - h-DDAF	0.9590535	-5.337457	7.255564	0.9317052

diff: difference between means of the two groups; lwr, upr: the lower and the upper end point of the confidence interval at 95% (default); p adj: p-value after adjustment for the multiple comparisons.

Comparison of linear growth rates between ABCA4 genotypes (Chapter 6)

Comparison of linear (square-root transformed) DDAF growth rates (mm)between *ABCA4* genotypes:

One-Way ANOVA

	p.(Gly1961Glu)	p.(Asn1868Ile)	Rare hypomorphs	Other *ABCA4* alleles	Total
n	9	9	11	17	46
∑X	4.994	7.111	11.067	21.071	44.243
Mean	0.5549	0.7901	1.0061	1.2395	0.962
∑X²	3.1957	6.5389	11.5659	27.5081	48.8086
Standard deviation	0.2304	0.3392	0.2077	0.2949	0.3728

Result details:

Source	SS	df	MS
Between-genotypes	3.0878	3	1.0293
Within-genotype	3.1677	42	0.0754
Total	6.2555	45	

Interpretation: The f-ratio = 13.64678 and p-value <0.00001 indicating a significant difference in the linear DDAF growth rates between ABCA4 genotypes.

Post Hoc Tukey HSD

Pairwise comparisons		HSD.05 = 0.3168 HSD.01 = 0.3920	Q.05 = 3.7830 Q.01 = 4.6805
p.(Gly1961Glu) vs. p.(Asn1868Ile)	M1 = 0.55	0.24	Q = 2.81 (p = .20952)
	M2 = 0.79		
p.(Gly1961Glu) vs. Rare hypomorphs	M1 = 0.55	0.45	**Q = 5.39 (p = .00244)**
	M3 = 1.01		
p.(Gly1961Glu) vs. Other *ABCA4* alleles	M1 = 0.55	0.68	**Q = 8.17 (p = .00000)**
	M4 = 1.24		
p.(Asn1868Ile) vs. Rare hypomorphs	M2 = 0.79	0.22	Q = 2.58 (p = .27674)
	M3 = 1.01		
p.(Asn1868Ile) vs. Other ABCA4 alleles	M2 = 0.79	0.45	**Q = 5.37 (p = .00255)**
	M4 = 1.24		
Rare hypomorphs vs. Other ABCA4 alleles	M3 = 1.01	0.23	Q = 2.79 (p = .21528)
	M4 = 1.24		

Interpretation: Significant pairwise differences exist between p.(Gly1961Glu) and Rare hypomorphs (p = .00244) and Other ABCA4 alleles (p < 0.00001), as well as p.(Asn1868Ile) and Other ABCA4 alleles (p = .00255) in the linear DDAF growth rates between ABCA4 genotypes.

Kruskal-Wallis Test

H statistic = $(12/(N(N+1))) * (\sum T^2/n) - 3(N+1)$ = 0.006 * 26065.325 - 138 = **13.1033**

P-value = **.00442**

Correlation between DDAF growth rates and circularity index in ABCA4 (Chapter 6)

Pearson correlation coefficient (r)

$$r = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\Sigma(x_i - \bar{x})^2 \Sigma(y_i - \bar{y})^2)}}$$

Pearson correlation between untransformed DDAF growth rate (mm^2/year) and circularity index:

$$r = \frac{-2.5497}{\sqrt{(1.69*25.5092)}} = \textbf{-0.3883}$$

P-value	0.007655
Covariance	-0.05666
Sample size (n)	46
Statistic	-2.7953

Interpretation: Results of the Pearson correlation indicated that there is a significant very small negative relationship between X and Y, (r(44) = .388, p = .008).

Pearson correlation between transformed DDAF growth rate (mm/year) and circularity index:

$$r = \frac{-0.3404}{\sqrt{(1.69*0.4688)}} = \textbf{-0.3824}$$

P-value	0.008715
Covariance	-0.007565
Sample size (n)	46
Statistic	-2.7456

Interpretation: Results of the Pearson correlation indicated that there is a significant very small negative relationship between X and Y, (r(44) = .382, p = .009).

Spearman rank correlation between untransformed DDAF growth rate (mm^2/year) and circularity index:

Spearman's rho (r_s)	-0.44481
P-value (2-tailed)	0.00195
S	23427.67
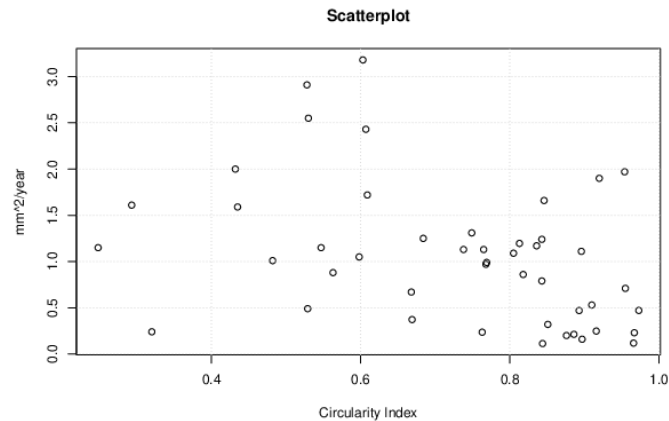
Scatterplot

Interpretation: By normal standards, the association between the two variables would be considered statistically significant

Spearman rank correlation between untransformed DDAF growth rate (mm^2/year) and circularity index:

Spearman's rho (r_s)	-0.45443
P-value (2-tailed)	0.00151
S	23583.68
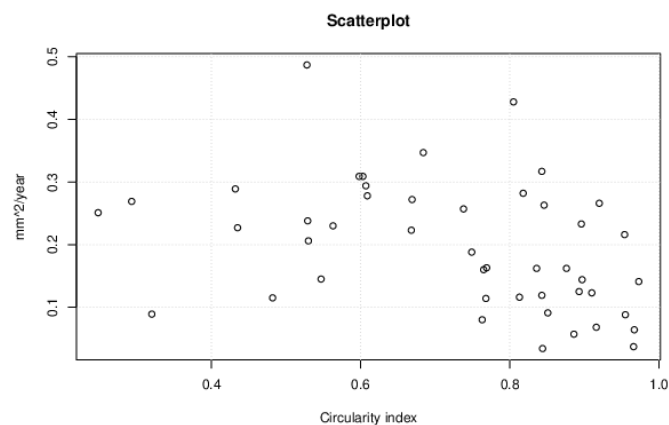
Scatterplot

Interpretation: By normal standards, the association between the two variables would be considered statistically significant

Association between mild alleles and outer retinal tubulation

Redundant familial alleles (siblings) were removed from the original cohort resulting in a total count of 413 unrelated cases for statistical analyses. Significance level (α) was defined as P ≤ 0.05.

2x2 Fisher's Exact Test for count data comparing genotype enrichment among Stargardt (ABCA4) disease patients with outer retinal tubulation (ORT).

Genotypes	p.(Asn1868Ile) + p.(Ile1562Thr) + p.(Arg2030Gln)	Other *ABCA4*
ORT	10	6
no ORT	49	348

P-value	Odds ratio [95% confidence interval]
0.000006973	11.7 [3.7 - 41.0]

Interpretation → ORT are significantly associated with all hypomorphic *ABCA4* alleles: p.(Asn1868Ile), p.(Ile1562Thr) and p.(Arg2030Gln).

Genotypes	p.(Asn1868Ile)	Other *ABCA4*
ORT	5	11
no ORT	47	350

P-value	Odds ratio [95% confidence interval]
0.03856	3.4 [0.8 - 11.1]

Interpretation → ORT are significantly associated with the hypomorphic ABCA4 allele p.(Asn1868Ile).

Genotypes	p.(Ile1562Thr) + p.(Arg2030Gln)	Other *ABCA4*
ORT	5	11
no ORT	2	395

P-value	Odds ratio [95% confidence interval]
8.969E-07	84.7 [12.3 - 969.5]

Interpretation → ORT are significantly associated with the hypomorphic ABCA4 alleles p.(Ile1562Thr) and p.(Arg2030Gln).

Mann-Whitney U test comparing age at ORT detection between patients with hypomorphic and other ABCA4 genotypes. Density plot: blue = patients with hypomorphic alleles ABCA4 alleles, p.(Asn1868Ile), p.(Ile1562Thr) and p.(Arg2030Gln); red = all other ABCA4 alleles; dotted lines = mean age.

Test statistic W	P-value
69	0.000754

Interpretation → ORT detection in patients with non-hypomorphic *ABCA4* alleles occurs at a significant younger age.
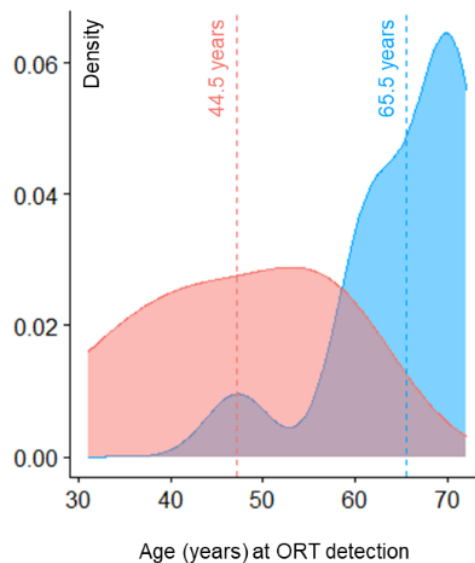
Appendix D: Clinical case descriptions (Chapter 8)

Subject 1: Subject 1 is an 8-year-old Caucasian girl of Dutch descent born after normal pregnancy and delivery at term (40 weeks + 3 days). She was small at birth, with weight and length of 2.4kg and 45cm, respectively. Infantile feeding difficulties were noted. Sleep disturbances were reported during early childhood. Motor, speech, and mild intellectual delays were evident upon examination. She began walking at 16 months with a broad-based gait; she was able to run at 5 years of age. Behaviorally, she exhibits physical and emotional aggression, head-banging, and features suggestive of ADHD and autism spectrum disorder (ASD). Her behavioral issues require continuous supervision. The parents report short sleep cycles (she falls asleep around 8pm and awakens at around 3am). At the most recent examination (age 8 years), her IQ was estimated to be 55. Speech was limited to 1-2- word sentences. She remained smaller than average: weight 22.2kg (-2.1 SD), height 118.8cm (0.4 SD). Microcephaly was suspected by her most recent occipitofrontal circumference (OFC) of 47.5cm (-2.5 SD). A brain MRI showed mild encephalopathy although this was attributed to an environmental cause. Variant 4-finger crease was noted in her right hand. Otherwise, the patient was considered to be in good physical health. Whole exome sequencing (WES) of the patient identified the missense variant, c.279T>G (p.S93R), in *CERT1* (NM_005713.3). Parental samples were not available for testing. The inheritance is presumed de novo as there is no family history of disease. Despite the fact that this variant is predicted to be weakly pathogenic by in silico algorithms and occurs in a weakly conserved position along the protein (PolyPhen2 (HVAR = 0.872)), it is not found in the general population according to gnomad.broadinstitute.org and bravo.sph.umich.edu (assessed September 2021).

Subject 2: Subject 2 is a 24-year-old Caucasian man (Danish ancestry) with a history of developmental delay and infantile-onset febrile seizures. He was born through normal vaginal delivery at term (birthweight: -1 SD). Developmental delay was noted at 8 weeks. Febrile seizures began at 10 months (twice convulsive status epilepticus), and he continued to have seizures until approximately 22 years of age. Seizures were typically focal to bilateral tonic-clonic and impaired consciousness. Due to bouts of severe epilepsy, constant surveillance was needed especially at night. Attacks generally require immediate hospitalization and intensive care. Gross motor delay was apparent from the record of late sitting (42 months) and support-dependent walking at 6 years of age. There is no growth delay but the patient is generally hypotonic. An IQ score was never formally calculated but is estimated to be below 20 according to the DSM-V scale. A history of behavioral issues include self-injury (arm-biting), which became less frequent at age 15- 17 years. His parents report excessive involuntary movements, persistent shaking and a propensity for object mouthing. An absence of eye contact and good visual memory are noted but a formal diagnosis of ASD has never been made. His parents report the patient as having a high pain threshold. An interpretation of periventricular leukomalacia (PVL) was made from an MRI at age 7 years, but computed tomography (CT) scanning at age 19 years did not show demyelination or other signs of leukodystrophy; rather, it showed moderate central to frontal as well as diffuse cerebral atrophy. Brain EEG waveform patterns showed diffuse slowing, variable localization, no epileptic discharges nor photosensitivity. Together, all tests are consistent with global cerebral atrophy. At the most recent examination (age 23 years), the patient showed relatively normal growth parameters, but remains non-verbal, still requires support for walking, and show stereotypic features. On a functional level, the patient requires assistance with routine activities such as eating and using the bathroom. Craniofacial/skeletal

features: high forehead, slightly down-slanted palpebral fissures, midface hypoplasia, medial/central flaring of eyebrows, brushy and broad eyebrows, synophrys, downslanting palpebral fissures, swelling lower eyelids, groove below lower eyelids, short nose, anteverted nares, broad/bulbous nasal tip, short columella, long philtrum, cupid's bow upper lip, large ears, hypoplastic antitragus, large ear lobe, pes valgus, hallux varus (sandal gap), broad/short and hammer toes, overriding toe(s), hypoplastic/short nails on toes, broad distal phalanges (toes), and hypermobile DIP-joints in both hands. Direct sequencing of the *SLC9A6* and *UBE3A* genes was negative for disease-causing mutations. Karyotyping, aCGH, metabolic screening and subtelomeric MLPA were also normal. Whole exome sequencing (WES) of the patient and both parents identified a de novo missense variant, c.395C>T (p.S132L) (rs1064794019), in *CERT1* (NM_005713.3) in the patient. This variant is absent in the general population according to gnomAD (gnomad.broadinstitute.org) and BRAVO (bravo.sph.umich.edu), last accessed September 2021. This variant has been reported in at least 5 other individuals with similar clinical features: three individuals in two large case-control studies, one individual submitted by geneDX (accession: SCV000567600.3) (https://www.ncbi.nlm.nih.gov/clinvar/variation/419654/) and another included in the present study (Subject 3).

Subject 3: Subject 3 is an 18-year-old Caucasian man of British ancestry born by normal vaginal delivery at 39 weeks (birthweight: 3kg). As an infant, he had feeding difficulties and failed to thrive, sleep disturbance, requiring percutaneous endoscopic gastrostomy (PEG) and fundoplication. He was diagnosed with neonatal global developmental delay and had febrile seizures at 18 months of age, and began having focal and myoclonic seizures from 21 months of age. Fine motor dysfunction includes possible sensory issues, as the patient was unable to grab

objects until 6 years of age. Patient was unable to sit without support until 4 years and 6 months and learned to walk with the support at age 8, before becoming immobile at age 12 years. Although an IQ score was never formally calculated, the intellectual disability is clinically estimated to be profound. Systemically, the patient has a history of gastrointestinal disturbances characterized by episodes of severe constipation and of diarrhea. The cause is uncertain although the patient previously had malrotation for which he underwent Ladd's procedure. There have also been recurrent urinary tract infections. He has exotropia of the right eye but no other significant visual impairment aside from myopia and astigmatism. The patient has mild scoliosis and osteoporosis. Occipital frontal circumference (OFC) measurement taken at age 13 years was 52.2 cm (2nd percentile) and his weight was 40 kg (25th centile). A slight prominence of the temporal horn of the right lateral ventricle and a rather generously-sized fourth ventricle with widely open foramina were noted on MRI at 5 years. At age 17 years, the patient continued to have tonic-clonic and absence seizures, with the former requiring hospitalization at times. A brain EEG recording at this age showed diffuse slowing of background rhythm, although no epileptiform activity was noted. His repeat brain MRI performed at 18 years revealed subtle asymmetry in the size of cerebellar hemispheres. Frontal lobes appear small relative to the remainder of the brain, and there is thinning of the corpus callosum, slightly enlarged ventricles, ex vacuao changes and an apparent absence of subcortical white matter throughout (see Supplemental Figure 1B). Currently, at age 18 years, growth parameters are within normal limits but the patient remains immobile and requires complete supervision in daily living. Brachycephaly with occipital flattening and mild turricephaly was also noted. Craniofacial/skeletal dysmorphia include: sloping forehead, synophrys, short nose, anteverted nares, cupid's bow upper lip, large ears and lobes, prominent ears, uplifted ear lobe, hypoplastic

223

antitragus, temporal narrowing/ narrow front, brachycephaly, occipital flattening, turricephaly, micrognathia, palpebral ptosis, pronounced philtrum, diastema of upper central incisors, lateral thickenings under the lower lip, 3rd- 4th fingers syndactyly, gothic palate/ marked alveolar ridges, gum hypertrophy. Karyotyping, Fragile X repeat expansion testing and aCGH were normal. Whole exome sequencing (WES) of the patient and both parents identified a de novo missense variant, c.395C>T (p.S132L) (rs1064794019), in *CERT1* (NM_005713.3). This variant is absent in the general population according to gnomAD (gnomad.broadinstitute.org) and BRAVO (bravo.sph.umich.edu), last accessed September 2021. This variant has been detected in at least 5 other individuals with similar clinical features: three individuals in two large case-control studies, one individual submitted by geneDX (Accession: SCV000567600.3) (https://www.ncbi.nlm.nih.gov/clinvar/variation/419654/) and another included in the present study.

Subject 4: Subject 4 is a 13-month-old boy of Dutch descent diagnosed with severe global developmental delay and noted clinical abnormalities at birth. He was delivered at 37 weeks + 6/7 days by uncomplicated cesarean section at a birth weight of 2.97kg (10th-50th percentile). He exhibited feeding difficulties and failure-to-thrive (weight-to-height = -2.1 SD). At 10 months, was evaluated according to the Bayley Scales of Infant Development-III and developmental age was found to be equivalent to 1 month. He exhibited delays in motor milestones (head control=hold at 45 degrees, gross=0.2 months and fine=0.4 months) and as such can neither sit nor walk. His verbal ability is currently limited to making noises. He is generally a quiet and passive infant. Behaviorally, he makes no eye contact or social smile and adopts a typical posture gazing at his right hand (see photograph on the right). Brain EEG was normal. He is hypertonic and stiff after waking but after a while becomes more active and

normotonic. Currently, weight-to-height falls at -2.13 SD, height 0.46 SD and OFC -0.73 SD. Metabolic screening has been normal. There are no apparent systemic issues although there is suspected sensorineural hearing loss and suspected cerebral visual impairment. At the most recent examination, he was noted to have a lock of hypopigmented (white) hair and very dry skin. Dysmorphic features: prominent forehead, short nose, anteverted nares, midface hypoplasia, cupid's bow upper lip, broad/bulbous nasal tip, hypoplastic alae nasi, palpebral ptosis, swelling lower eyelids, long philtrum, full upper eyelids, full cheeks, protruding columella, everted lower lip, medial/central flaring of eyebrows, epicanthal folds, short palpebral fissures, hammer toes, 2nd-3rd toes proximal syndactyly, temporal narrowing/ narrow front, pronounced philtrum. Genetic work up included analysis of copy number variation by aCGH which identified a deletion at arr[GRCh37] 7q21.12(87668378_87739123)x1, mat. A trio-exome-based intellectual disability panel (1000 genes) identified de novo variant, c.779C>T (p.S132L), in *CERT1* (NM_005713.3). This variant is also pathogenic in Subjects 2, 3, and 5.

Subject 5: Subject 5 is a 20-year-old man of German descent with noted developmental delay at 0 months. He was delivered at 36 weeks + 4 days at a birth weight of 3.12kg grams, length of 48 cm (0.71 SD) with no other reported neonatal issues. He was diagnosed with severe intellectual disability and epilepsy at early but unknown age. Infantile cerebral palsy was noted. At the most recent examination, he appears to be hypotonic and short in stature. His current weight is 46 kg (-2.7kg), height 165 cm (-1.8SD) and OFC 55.5 cm (-0.66 SD). His intellectual disability is severe (estimated IQ range 20-35 according to the DSM-V scale) and he is non-verbal. There was profound motor delay, as he only began to walk, with assistance, at 7 years of age. Currently, his gait is noted to be ataxic. Hypomyelination was noted in neuro-imaging. A cMRI at 4 years of age showed increased subcortical T2- weighted intensity throughout the cerebrum,

with occipital dominance (see panel below, and Supplemental Figure 1B). Prominent Virchow-Robin spaces were also noted, some of which were microcystic in the occipital region in both hemispheres. EEG recordings showed seizure susceptibility; he has had one documented seizure. No progression was found in follow-up imaging within the last 10 years. He has high tolerance for pain and multiple allergies. Dysmorphic/skeletal features: high forehead, anteverted nares, cupid's bow upper lip, hypoplastic alae nasi, protruding columella, everted lower lip, hammer toes, flared nares, micrognathia, long face, hallux varus (sandal gap), deep-set eyes, epicanthus, hypoplastic ear lobe, broad/short toes, hypoplastic/short nails on toes, broad eyebrows, short distal phalanx in toes, brachycephaly, occipital flattening, scoliosis. Trio-based exome sequencing performed in both Subject 5 and his parents identified the de novo variant c.779C>T (p.S132L) in *CERT1* (NM_005713.3). This variant is also pathogenic in Subjects 2, 3, and 4.

Subject 6[ʈʈ]: Subject 6 is a 23-year-old woman of Japanese descent born to unaffected non-consanguineous parents. She was born at full term (gestational age = 37 weeks, 5 days) following an uneventful pregnancy but was small (birth weight = 2.46 kg, height = 44.0 cm, OFC = 33.2 cm). Severe delay in developmental milestones were noted, along with hypotonia and failure to thrive. Although she attained head control at 6 months of age, she was unable to sit without support until age 4 years. She was treated for febrile seizures at age 5, which were believed to have been triggered by a bout of influenza. MRI of the brain at this age showed delayed myelination in the cerebrum and corpus callosum hypoplasia. At 6 years old she was unable to stand or speak meaningful words and her IQ was estimated to be below 35. Follow-up brain MRI at age 23 confirmed leukodystrophy, which was especially apparent in the frontal

---

[ʈʈ] This patient was reported by Murakami et al., (26), and the following clinical details were confirmed by the authors.

lobe. Physical examination at this time identified other systemic complications related to growth, central nervous system, and muscular tonus. Craniofacial/skeletal dysmorphia include: short nose, anteverted nares, diastema of upper central incisors, uplifted ear lobe, midface hypoplasia, groove below lower eyelids, upslanting palpebral fissures, hypertrichosis, bushy eyebrows, hypotelorism, long palpebral fissures, flat nasal bridge, flared nares, short philtrum. She also develop scoliosis and torticollis. Trio exome sequencing identified a de novo variant, c.403T>C, (p.S135P), in *CERT1* (NM_005713.3). This variant is also pathogenic in Subjects 7.

Subject 7: Subject 7 is a 2-year, 3-month-old Caucasian girl with developmental delay conceived by intracytoplasmic sperm injection (ICSI) and born from an otherwise healthy bichorial biamniotic pregnancy. She was delivered vaginally at 34 weeks at a birth weight of 1.98 kg (-1.7 SD), length 43 cm (-1.7 SD) and OFC 31 cm (-1.2 SD). She exhibited infantile feeding difficulties due to gastro-esophageal reflux and failure-to-thrive. Examination at 6 months revealed global delay. Her developmental milestones were thus far are head-holding (3 months) and sitting (19 months). She has not yet begun to walk. There is also apparent speech delay. She is considered to have severe intellectual disability, though she is still too young to be evaluated for IQ. Behaviorally, she exhibits autistic behaviors such as stereotypic and repetitive behavior and anxiety in response to changes in environment but, unlike some other CerTra patients, shows no sign of aggression or automutilation. At 9 months of age, MRI of the brain showed delay in myelination and dilated ventricles, without abnormalities in head size. EEG of the brain was normal. Neurologically, the patient has not yet had any documented seizures and does not suffer from sleep disturbances. Examination of the fundus was normal although she is astigmatic. There are no apparent neurosensory issues such as hearing loss or abnormal sensitivity to pain. At her current age she remains short in stature and appears hypotonic. Her

most recent growth parameters are weight 9.,4kg (1%), 82 cm (-2 SD) and 47 cm (-1 SD). With

the exception of developmental delay and gastro-esophageal reflux, she is otherwise systemically

healthy. Blood count, ionogram was normal. TSH and immunoglobulin tests were also normal.

Craniofacial/skeletal dysmorphia include: epicanthus, anteverted nares, low set ears, thin

upperlip, micrognathia, buccal hypotonia. Single transverse palmar crease in the hands was also

noted. Family history is negative for significant disease. Fragile X (CGG expansion), array-CGH

testing and analysis of the SNRPN locus were negative. Trio genome sequencing of the Subject 7

and both parents identified a de novo c.403T>C, (p.S135P) in *CERT1* (NM_005713.3) in the

proband. This variant is absent from the general population (gnomAD) and is predicted

deleterious by in silico algorithms: M-CAP = 0.108, REVEL = 0.516, CADDv1.6 = 27.9 and

Eigen = 0.853. This variant was also reported in Subject 6.

Subject 8[ss]: Subject 8 is a 50-year-old Caucasian man of Dutch descent. He was born by vaginal

delivery at full term; birthweight was reportedly normal but subsequent infantile feeding

difficulties were noted. Development was normal (walking age ~1 year old) until 4 years of age,

after which there was a reported regression in speech (few-words) and motor development.

Immediate family history was negative except for a nephew (sister's son) who reportedly had

pervasive developmental delay-not otherwise specified (PDD-NOS). At his most recent

examination (age 50 years), the patient appears to have moderate intellectual disability (formal

IQ score was not calculated but is estimated to be between 36-49 according to DSM-V) and

behavioral issues such as general anxiety, aggression, hypersensitivity to touch, and self-

injurious behavior, which are consistent with autism spectrum disorder (ASD). His weight is

---

[ss] Subject 8 is a case that was included in de Ligt et al. from a large screening for intellectual disability. The following information was summarized from the article along with further clinical details we obtained from the clinicians.

79.4kg; height is 177.7cm and OFC 55.8cm (20th percentile). Systemically, the patient is generally healthy. Brain MRI or EEG were performed. There is no history of epilepsy or seizures. Basic metabolic screening, including carbohydrate-deficient transferrin for congenital disorders of glycosylation (CDG), was normal. Mild facial dysmorphia were noted: deeply set eyes, a long face and a prominent forehead. Fragile X and aCGH testing were normal. Whole exome sequencing identified a *de novo* missense variant, c.413C>G (p.S138C), in *CERT1* (NM_005713.3). This variant is absent in the general population from both gnomAD (gnomad.broadinstitute.org) and BRAVO (bravo.sph.umich.edu), assessed September 2021. This variant has been detected in another subject in our cohort (Subject 9), who is currently only 21 months old but already exhibits anxiety and repetitive behaviors that might develop into an ASD phenotype. Unlike Subject 8, however, Subject 9 experienced developmental delay early in life.

Subject 9: Subject 9 is a 3.5-year-old Caucasian boy of Polish descent. He was born at term (40 weeks) by vaginal delivery with no perinatal complications. He was slightly smaller than average size: birth weight was 3.35 kg (10th – 50th percentile), length 51cm (10th – 50th percentile) and OFC 34 cm (10th – 50th percentile). Microcephaly was noted in infancy. Infantile feeding difficulties were also noted, but this was likely due to an allergy to cow's milk leading to precipitous drop in weight at 3 months. Neurodevelopmental issues and moderate intellectual disability were noted at 6 months. Brain MRI and EEG testing at 13 months were normal. Self-supported sitting did not occur until 10 months. At 21 months, he shows speech delay and he is hypotonic. General growth parameters are now somewhat above average—weight 12.5kg (90th) and height 87cm (75th percentile)—but OFC at 50cm (90th – 97th percentile) suggests acquired macrocephaly. Repetitive movements were observed and he reportedly suffers from anxiety and timidity in unfamiliar places and need daily support. He also experiences sleep disturbances.

Apparent craniofacial/skeletal dysmorphia include: high and prominent forehead, temporal narrowing/ narrow front, high frontal hairline, hair growth in the lateral parts of the front, retrognathia/micrognathia, cupid's bow upper lip, hypoplastic ear lobe, broad/bulbous nasal tip, full cheeks, lateral thickenings under the lower lip, everted upper and lower lips, medial/central flaring of eyebrows, swelling lower eyelids, groove below lower eyelids, lateral flaring of eyebrows, short columella, prominent ears, uplifted ear lobe, down-slanting palpebral fissures, anteverted nares, long philtrum, hammer toes, hypoplastic/short nails on toes, broad distal phalanges (toes), hallux varus (sandal gap), clubfeet, overriding toe(s), broad/short toes, 2nd-3rd toes proximal syndactyly, short distal phalanx in toes, clubfeet Whole exome sequencing (WES) of the patient and both parents, performed at a diagnostic laboratory (MEDGEN), identified a de novo missense variant, c.413C>G (p.S138C), in *CERT1* (NM_005713.3). This variant is absent in the general population according to gnomAD (gnomad.broadinstitute.org and bravo.sph.umich.edu; assessed September 2021) and is also reported in Subject 8 in this study.

Subject 10[tt]: He is a 9-year-old boy with behavioral issues; intellectual disability (VIQ = 54; NVIQ = 47); and seizures. Short stature and growth delays were not reported. Trio WES identified de novo variant c.496A>G (p.T166A) in *CERT1* (NM_005713.3). This variant has been detected in three more individuals enrolled in this study (Subjects 11, 12, and 13).

Subject 11: Subject 11 is a 3-year old girl. At the age of one year she presented with motor delay; autism spectrum disorder (ASD); ADHD; mild to moderate intellectual disability (PPVT IQ score = 49). She did not have speech delay, seizures, feeding difficulties, hypotonia, short stature, or gastrointestinal problems. Her brain MRI was normal. Apparent craniofacial

---

[tt] Subject 10 is a case that was included in Iossifov et al. (28), after screening for autism spectrum disorder.

dysmorphia include: cupid's bow upper lip, hypoplastic ear lobe, full cheeks, lateral thickenings under the lower lip, medial/central flaring of eyebrows, swelling lower eyelids, groove below lower eyelids, short columella, upslanting palpebral fissures, prominent chin, broad eyebrows. Molecular Inversion Probe (MIP) screening in the patient and parents identified a de novo variant, c.496A>G (p.T166A), in *CERT1* (NM_005713.3), which is also in Subjects 10, 12, and 13.

Subject 12: Subject 12 is a 5-year-old Caucasian boy of German descent delivered by C-section at the 37th week (birth weight 2.7 kg; length 51cm; OFC 33cm) with an otherwise unremarkable pre- and perinatal history. He presented at 10 months with mild motor delay and recurring febrile seizures. The patient learned to sit and walk at 10 and 18 months, respectively. An evaluation at 37 months revealed his developmental age to be 22 months according to the Bayley Scales of Infant Development. At 5 years of age, he continues to exhibit mild speech delay, although he is able to speak in complete sentences. There are no overt behavioral abnormalities, but he was reported to have difficulty concentrating. In total, the patient suffered from seven febrile seizures (three of which were complicated). Brain MRI and EEG recordings were normal. The most recent physical exam found weight and height in the lower centiles (16 kg, 3rd-10th percentile; 105cm, 3rd-10th percentile) and OFC between the 10th and 25th centile (50 cm). Aside from mild recurrent bronchopulmonal infections, the patient is otherwise healthy. Mild craniofacial/skeletal dysmorphic features: prominent forehead, hypoplastic ear lobe, full cheeks, medial/central flaring of eyebrows, swelling lower eyelids, upslanting palpebral fissures, prominent chin, skull/facial asymmetry, prominent ears, protruding columella, flared nares, arched eyebrows, overriding toe(s), temporal narrowing/ narrow front. Metabolic screening, karyotyping and Fragile X repeat expansion testing were normal. A maternally-inherited

231

duplication of 10q24 was found in chromosomal microarray analysis. Direct sequencing of the sodium voltage-gated channel alpha subunit 1 (*SCN1A*) gene, which is associated with febrile seizures, was negative. Whole exome sequencing (WES) of the patient and both parents identified a de novo missense variant, c.496A>G (p.T166A) in *CERT1* (NM_005713.3). This variant is absent in the general population according to both gnomAD (gnomad.broadinstitute.org) and BRAVO (bravo.sph.umich.edu), but occurs at a weakly conserved position in the protein and is predicted to be moderately pathogenic by *in silico* pathogenicity algorithms: CADD-PHRED = 23.7, Eigen = 0.3585. This variant was found to be pathogenic in three other subjects in the present study.

Subject 13: Subject 13 is a 9-year-old Caucasian girl of Irish descent who was born with cleft lip and palate at term (39 weeks + 5 days) via normal vaginal delivery. She had several neonatal complications: feeding difficulties, pharyngeal incompetence, and episodes of stiffening at day 2 of life. Feeding difficulties continued even after repair of the cleft lip. At 6 months of age she had abnormal movements with eye-rolling episodes that were self-resolving (likely seizures). By 9 months of age she exhibited cognitive and developmental delays, and by 18 months developed autistic features. She was able to sit without support at 9 months, crawl at 15 months and walk at 19 months. She did not babble, point, or clap, and further development stalled after 19 months of age. She has some gait ataxia, falls frequently if not supported, and requires a special chair. At age 4, she was formally diagnosed with autism spectrum disorder (ASD) likely because of repetitive movements (head-banging) and obsessive behaviors; she involuntarily clenches at times while alert. She also experiences sleep disturbance. At the same age she was diagnosed with epileptic encephalopathy and later on with severe intellectual disability. She appears agitated at times but has a very friendly disposition and social smile; there is no aggressive

behavior. She communicates without clear words or gestures by smiling, crying, or staring through the corners of her eyes. She has healthy skin generally with freckles, but has dermatographia (e.g., red marks noticeable when body is gently lifted up, sore after bandages etc.). However, she has a high pain threshold. Hearing assessment was found to be normal although she had grommets inserted to relieve fluid build-up in the ear. Brain MRI at age 2 years of age was normal although her EEG showed bursts of generalized spike and slow wave activity. She experiences various types of seizures, multiple times throughout the day. Two episodes required admission to the ICU for intubation/ventilation. The range of seizures over the years include jerky movements, myoclonus, absence, tonic seizures, cluster seizures, and status epilepticus. Antiepileptic drugs have been tried (Phenobarbital, Levetiracepam, Valproate, Phenytoin, Ethosuccimide, Topiramate, Clobazam, Clonazepam, Midazolam, Lacosamide, Zonisamide, Rufinamide, Brivaracetam, Topiramate, Gabapentin), as well as a trial of a ketogenic diet, all without much success. Extensive metabolic, genetic and endocrine work-up was unrevealing. A follow-up MRI brain at 8 years of age revealed diffuse supra- and infratentorial brain atrophy and thinning of the corpus callosum, with an appropriate myelination pattern for her age. Observed craniofacial/skeletal dysmorphia included: prominent chin, high and prominent forehead, temporal narrowing/ narrow front, cupid's bow upper lip, hair growth in the lateral parts of the front, tapering fingers, brachydactyly, prominent ears, lateral flaring of eyebrows, pronounced philtrum, short philtrum, tubular nose, long face, thickening of all parts of the ear (except the earlobe), hammer toes, hypoplastic/short nails on toes, broad distal phalanges (toes), broad/short toes, 2nd-3rd toes proximal syndactyly, short distal phalanx in toes. Whole genome sequencing (WGS) of Subject 13 and both parents identified a de novo variant in the patient, c.496A>G (p.T166A) in *CERT1* (NM_005713.3), which was subsequently confirmed by

233

Sanger sequencing. This variant was found to be pathogenic in three other subjects in the present study (Subject 10, 11, and 12).

Subject 14: Subject 14 is an 8-year-old girl of Dutch and Moroccan parentage. The maternal uncle was reported to have had trisomy 13. Perinatal history was unremarkable (gestation: 40+1, Apgar score 7-8-9) although the newborn was on continuous positive airway pressure (CPAP) for a short period of time; birth weight was 3.4 kg (0 SD). She did not exhibit feeding difficulties or failure-to-thrive. Developmental delay was noted from 4 months of age with emerging behavioral abnormalities including head banging, aggression, wrist-biting, hand-wringing, hand-flapping, severe bruxism in subsequent years. Delays in motor development were apparent in late sitting and walking ages at 15 months and 2 years 8 months, respectively. Developmental regression caused loss of her first words. A formal diagnosis of autism spectrum disorder (ASD) was made along with intellectual disability (estimated IQ range was 20-35 according to the DSM-V scale). She had a febrile seizure at 7.2 years of age (tonic-clonic insult during onset of illness) for which she was prescribed paracetamol (to suppress fevers); episodes of staring have been noted for several years. Patient underwent a 24-hour EEG one month later, which showed asymmetry and frequent epileptic activity with a maximum in the right parietal area Brain MRI/MRS in 2012 and in 2018 after her first epileptic insult did not reveal structural abnormalities but interpretation might have been hindered by movement artefacts. The results were reviewed by a pediatric neuroradiologist who noted no cerebellar abnormalities, normal myelination for her age, normal white matter. Symmetric hyperintense areas at the ventral mesencephalon were noted on Diffusion-Weighted Imaging (DWI), but their significance is unclear and may be to imaging artifacts. A second evaluation of the MRI by a radiologist at Columbia University Irving Medical Center (CUIMC) confirmed normal myelination patterns

with the added observation of normal basal ganglia, questionable subtle asymmetry in size of cerebellar hemispheres (right<left), small frontal lobes relative to the remainder of the brain and mild thinning of the corpus callosum. At the most recent examination at age 8 years, the patient exhibited severe intellectual disability and cortical visual impairment (~20% vision estimated). The patient is currently non-verbal. Her cognitive level was estimated to be at 9-11 months of age. The patient has a high pain threshold and skin xerosis. She appears to be hypotonic and petite (-2.18 SD) with growth delay. Her Occipital Frontal Circumference (OFC) is -0.8 SD. Craniofacial/skeletal dysmorphic features: prominent and high forehead, widow's peak, temporal narrowing front, dolichocephaly, midface hypoplasia, medial/central flaring of eyebrows, synophrys, epicanthal folds, full cheeks, short nose, anteverted and flared nares, short columella, full lips, everted lower and upper lip, cupid's bow upper lip, diastema of upper central incisors, lateral thickenings under the lower lip, large ears, hypoplastic tragus, hypoplastic ear lobe, broad/short and hammer toes, 2nd-3rd toes proximal syndactyly, hypoplastic/short nails on toes. Genetic analyses were performed at age 2 years beginning with direct sequencing of the following candidate genes: *MEF2C*, *MECP2*, *FOXG1*, *CDKL5*, *CLN1* and *CLN2*; no known disease-causing mutations were found. Fragile X testing, aCGH, extensive metabolic and lysosomal tests were normal. Whole exome sequencing (WES) of the patient and both parents identified a de novo missense variant, c.727G>A (p.G243R), in *CERT1* (NM_005713.3) (See Sanger sequencing traces). This variant was also detected in Subject 15 in our study but is absent in the general population according to gnomAD (gnomad.broadinstitute.org) and BRAVO (bravo.sph.umich.edu), last accessed September 2021. In 2016, the whole-exome sequencing (WES) result of the patient was published as part of a large study of 2,104 trios that sought to identify genes underlying intellectual disability.

Subject 15[uu]: Subject 15 is a 17-year-old boy diagnosed with severe global developmental delay, severe intellectual disability, self-mutilation, and aggression. Perinatal history was unremarkable except for feeding difficulties. He reportedly began experiencing febrile seizures at the age of 3 years; unprovoked focal motor seizures occurred at age 5 years, which were treated with clobazam. He did not walk until age 5 years. He has episodes of agitation and self-injurious behavior alternating with periods of apathy, which are treated with risperidone, paroxetine and valproate. He has a left renal agenesis and mild right vesicoureteral reflux. Brain computed tomography (CT) showed mild ventriculomegaly and mild periventricular leukomalacia. At the most recent examination at the age of 17, he had begun using utensils for eating although he could not get dressed by himself. He has now been off anticonvulsants and seizure-free for several years. His speech was limited to a few words; he understands simple commands. His weight was 53.7 kg (10th percentile), height 163.7 cm (3rd -10th percentile) and head circumference 57.3 cm (98th percentile). He had poor eye contact. Craniofacial/Skeletal features noted during examination included: large/broad distal phalanges in fingers, and broad distal phalanges (toes). Chromosomal microarray revealed a maternally-inherited 0.069 Mb duplication on chromosome 15q26.3 although his mother was examined and was phenotypically normal. STXBP1 mutation analysis and metabolic studies (plasma lactate and ammonia measurements, plasma amino acids chromatography) were normal. Whole exome sequencing (WES) of the patient and both parents was performed after which a de novo missense variant, c.727G>A, (p.G243R) in the *CERT1* (NM_005713.3) gene was identified. This variant is absent in the

---

[uu] The genetic analysis and clinical characteristics of this patients were previously reported in a large case-control study of moderate to severe intellectual disability.

236

general population according to gnomAD (gnomad.broadinstitute.org) and BRAVO (bravo.sph.umich.edu), last accessed September 2021.

Subject 16: Subject 16 is a 5-year-old Caucasian boy of Dutch descent. He was born prematurely (36 weeks) by vaginal delivery. His birth weight was 2.5kg, length 45cm and occipital frontal circumference (OFC) 33cm. Neonatal feeding difficulties and failure-to-thrive were noted and he was placed on a nasal tube for 1 week. He had early-onset (age 1 year) moderate to severe developmental delay and seizures. Independent (unaided) walking began at 30 months. His IQ is estimated to be below 55; at 5 years of age, his developmental age was estimated to be approximately 12 months, giving a developmental quotient score of 20 (DQ = CA/DA). Severe speech delay and cerebral visual impairment (CVI) were noted along with myopia. The patient exhibits features of autism spectrum disorder and ADHD. He suffers from insomnia, has difficulty falling asleep and often awakes in the middle of the night (around 3:00AM). At the most recent examination, weight, height and OFC were +0.5 SD, -2 SD, and -1 SD, respectively. The patient has severe constipation which requires colonic lavage. Metabolic screening showed low amino acid levels. The patient is also IgA-deficient and suffers from frequent infections. His skin has a yellow hue. Focal motor seizures are controlled with oxcarbazepine; EEG recordings detected a normal background rhythm and no epileptic discharges. Relevant findings from brain MRI include delayed myelination. Craniofacial/Skeletal features noted during examination included: high forehead, micrognathia, medial/central flaring of eyebrows, broad eyebrow, upslanting palpebral fissures, swelling lower eyelids, groove below lower eyelids, broad nasal root, short nose, anteverted and flared nares, broad/bulbous nasal tip, short and protruding columella, thin upper lip, prominent ears, hypoplastic antitragus, uplifted ear lobe, hypoplastic/ short nails on fingers, pes planovalgus. Array-CGH testing was normal.

Whole exome sequencing of the patient and his parents identified a de novo variant, c.728G>A (p.G243E), in the *CERT1* gene (NM_005713.3). This variant occurs at a highly conserved position along the protein (PolyPhen2 (HVAR = 0.999). It is absent in the general population according to gnomAD (gnomad.broadinstitute.org) and BRAVO (bravo.sph.umich.edu), last accessed September 2021, and is predicted to be deleterious by various in silico pathogenicity algorithms: CADD=27.3, M-CAP=0.023, and Eigen=0.7437.

Subject 17: Subject 17 is a 4-year-old Caucasian girl of Italian descent. She was born at term by vaginal delivery and was on the smaller side: birth weight of 3.18 kg, length of 50 cm and OFC of 33 cm. Her neonatal history noted hypotonia and feeding difficulties that included poor suction and later chewing difficulties, though there was no failure to thrive. Gross motor milestones were severely delayed. Head-holding was achieved at 18 months and sitting without support at 4 years of age. Cognitive delays were not quantified but she appeared to have severe intellectual disability (estimated IQ range 20-35 according to DSM-V scale) and no verbal ability. She exhibits repetitive hand-washing behavior. There is no history of seizures. Neurosensory issues include increased tolerance to pain, problems with smell and taste and possible unspecified mild hearing impairment. The most recent physical exam noted general hypotonia and low weight for her age (10.1 kg, <3 percentile) but not short stature (90 cm, 25-50th centile). She still has feeding difficulties. Brain MRI at age 3 years showed hypoplasia of the corpus callosum and abnormality of the pons, diffuse paucity of supratentorial white matter with abnormal signal seen most marked in the periatrial white matter (Supplemental Figure 1B). Craniofacial/skeletal dysmorphias included: high and prominent forehead, cupid's bow upper lip, protruding columella, micrognathia, long face, hypoplastic ear lobe, broad eyebrows, broad/bulbous nasal tip, full cheeks, epicanthal folds, temporal narrowing/ narrow front, high

frontal hairline, bushy eyebrows, lateral thickening under the lower lip, malar hypoplasia, everted upper lip, hypoplastic antitragus, hair growth in the lateral parts of the front, dolichocephaly. The patient also show marked dermatoglyphs on the soles and lumbar kyphoscoliosis. Initial genetic screening consisted of a diagnostic gene panel for Rett or Rett-like syndromes *MECP2*, *MEF2C*, *CDKL5*, *FOXG1*, *SMC1A*, *TCF4*, *SHANK3*, *SATB2*, *STXBP1*, *SCN1A*, *SCN2A*, *SCN8A*, *GRIN2A*, *GRIN2B*, *HCN1*, *SLC6A1*, *WDR45*, *ST3GAL5* and *MFSD* which was negative. Multiplex ligation-dependent probe amplification (MLPA) of 15q11.2 was normal. Trio exome sequencing of Subject 12 and both unaffected parents identified a de novo missense variant, c.740C>T (p.T247I), in *CERT1* (NM_005713.3). The variant is predicted to be deleterious by several algorithms and is absent from the general population on both gnomAD (gnomad.broadinstitute.org) and BRAVO (bravo.sph.umich.edu), last searched September 2021.

Subject 18: Subject 18 is a one-year-old boy with severe intellectual disability and severe speech delay (per Gesell Development Schedules), growth delay with musculoskeletal issues, and hypotonia. At this young age, there is as yet no history of seizures, visual impairment or hearing loss. No dysmorphic features have been reported. Brain MRI and EEG were normal. Molecular Inversion Probe (MIP) screening in the patient and parents identified a de novo variant, c.751A>G, (p.T251A), in the *CERT1* gene (NM_005713.3).

Subject 19: Subject 19 was identified as having moderate intellectual disability; epileptic seizures; developmental regression. Singleton screening identified c.872A>G (p.Y291C) in *CERT1* (NM_005713.3).

Subject 20: Subject 20 is a 7-year-old African-American boy of Nigerian descent delivered by C-section at 41 weeks + 5 days (birthweight was 3.95 kg, 80th centile). Perinatal history was unremarkable. The patient exhibited normal motor development: head control at 4 months,
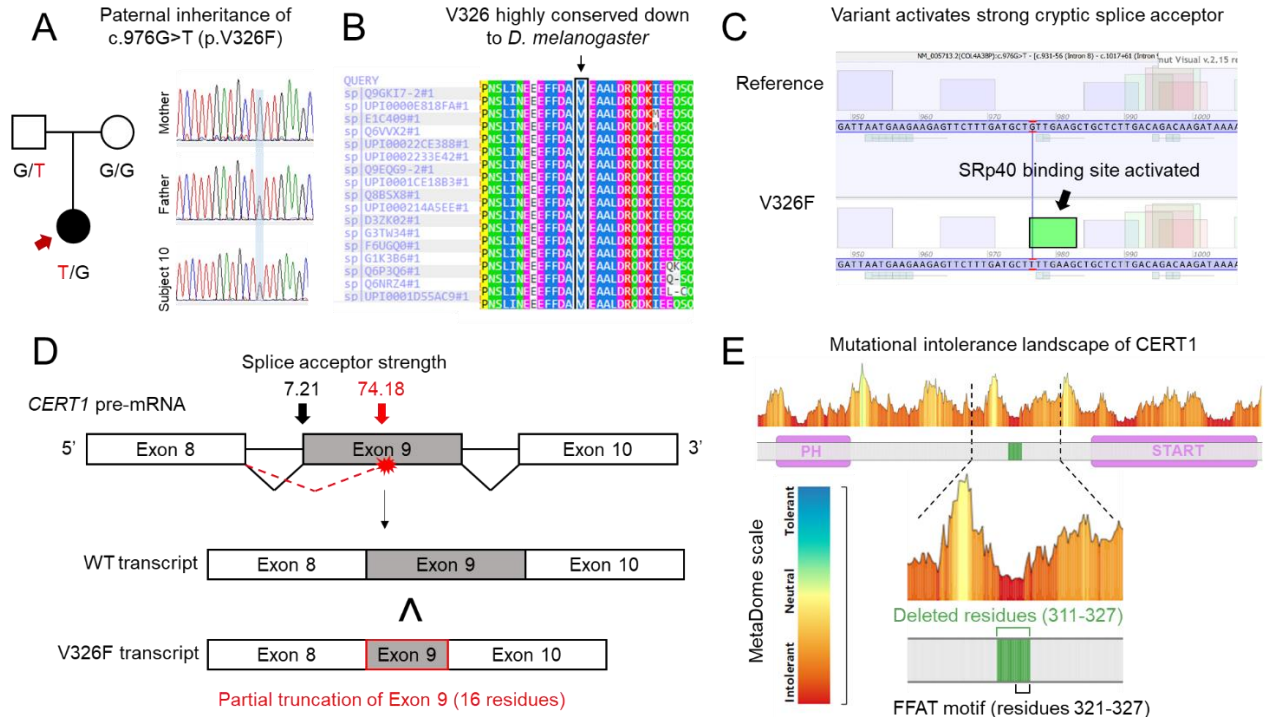
sitting without support at 6 months, and independent walking at 1 year. The presenting symptom, at two years, was myclonic astatic epilepsy (Doose syndrome). Brain MRI was normal but EEG recordings showed mild irritative activity in temporal areas of both hemispheres. The patient was treated with valproic acid, zonisamide and risperidone and stopped having seizures after 5 years of age. Other neurological issues included non-epileptic paroxysmal events related to emotional frustration. At the time of examination (7 years and 8 months), the patient exhibited mild speech delay and mild to moderate intellectual disability, with features consistent with ADHD (restlessness and impulsivity). There were no musculoskeletal abnormalities. Growth parameters were normal: weight 34.7kg (91st percentile), height 129cm (61st percentile) and OFC 55cm (96th percentile). Dry skin (xerosis) was noted. Observed craniofacial/skeletal dysmorphia included: prominent forehead, high frontal hairline, hair growth in the lateral parts of the front, temporal narrowing/ narrow front, retrognathia/ micrognathia, medial/central flaring of eyebrows, bushy eyebrows, broad eyebrows, arched eyebrows, upslating palpebral fissures, full cheeks, flat nasal bridge, flared nares, broad/bulbous nasal tip, short columella, long philtrum, smooth philtrum, full lips, cupid's bow upper lip, diastema of upper central incisors, prominent chin low-set ears, hypoplastic ear lobe, tapering fingers, few palmar creases, brachydactyly Serology and metabolic screening were normal. Karyotyping (46, XY) and Fragile X (31 CGG repeats) testing were normal. Array-CGH analysis was also normal. Whole exome sequencing of the patient identified presence of a de novo missense variant, c.887C>G (p.T296R), in *CERT1* (NM_005713.3) gene. A maternal sample was not available and the father is deceased. The inheritance is presumed de novo as there is no family history of disease. Although this variant occurs at a moderately conserved position along the protein (PolyPhen2 (HVAR = 0.872) and is predicted to be only weakly pathogenic by in silico algorithms, it is not found in the general

population according to gnomad.broadinstitute.org and bravo.sph.umich.edu (assessed September 2021).

Subject 21: Subject 21 is a 12-year-old Caucasian girl of Italian and Albanian descent with a history of developmental delay and ataxia. She was born at term (40 weeks and 3 days) by vaginal delivery; perinatal history is unremarkable. Her birth measurements were at the low end of normal: weight was 3.2kg, length was 50cm and OFC 34.5cm. She exhibited infantile feeding difficulties and regression of weight, growth, and OFC at around 8-9 months of age. She had mild motor delay, learning to sit without support at 9 months and walk at 18 months. She has an ataxic gait. At her most recent examination (12 years of age), the patient showed limited bi-syllabic speech and, based on this, was estimated to have moderate intellectual disability. She exhibits repetitive/obsessive hand-washing, self-injurious hand-biting, aggression, ADHD (hyperactivity) and other features consistent with autism spectrum disorder (ASD). Behavioral issues were noted at infancy but grew less frequent by 10 years of age. There is no history of seizures. She is underweight (21 kg, 3rd percentile) and has microcephaly (51 cm), also observed in infancy. Brain MRI was normal (last MRI performed in 2016). Apparent craniofacial/skeletal dysmorphia include: micrognathia, cupid's bow upper lip, hypoplastic ear lobe, broad/bulbous nasal tip, full cheeks, epicanthal folds, lateral thickenings under the lower lip, everted upper lip, hypoplastic antitragus, dolichocephaly, hypoplastic *alae nasi*, everted lower lip, deep-set eyes, medial/central flaring of eyebrows, pronounced philtrum, uplifted ear lobe, short philtrum, downslanting palpebral fissures, tubular nose, broad nasal root, full lips, large ears, hammer toes, hypoplastic/short nails on toes, prominent upper central incisors, broad distal phalanges (toes). Array-CGH testing identified a 114kb duplication at Xp21.2 (arr Xp21.2(29,584,944x2, 29,594,080 - 20,708,100x3, 29,717,001x2). The duplication includes part of the *IL1RAPL1* gene

(exon 6). Quantitative PCR confirmed the duplication to be de novo, but aCGH and qPCR are not able to establish if this anomaly is a tandem duplication or an insertional unbalanced translocation. A tandem duplication would cause an effective alteration of the *IL1RAPL1* gene. Unfortunately, the family moved to another country and are not available for further study. Direct sequencing of the *MECP2* and *UBE3A* genes were negative. A gene panel consisting of 74 intellectual disability-autism spectrum disorder genes was performed (full citation below), with negative results. Whole exome sequencing (WES) of the patient and both her parents identified a paternally inherited missense variant, c.976G>T (p.V326F), in *CERT1* (NM_005713.3) (**Appendix Figure D1A**). This variant is highly deleterious according to several *in silico* algorithms (PolyPhen2 HVAR = 0.964, CADD-PHRED = 34, REVEL = 0.37 and Eigen = 0.7073) and is predicted to have putative effects on splicing (**Appendix Figure D1C-Appendix Figure D1E**). The variant results in substitution of a highly conserved nucleotide (phyloP: 7.52 [-20.0;10.0]) (**Appendix Figure D1B**) and is not found in the general population according to gnomad.broadinstitute.org and bravo.sph.umich.edu (assessed September 2021).

Appendix Figure D1.[vv]

A — Paternal inheritance of c.976G>T (p.V326F)

B — V326 highly conserved down to *D. melanogaster*

C — Variant activates strong cryptic splice acceptor

Reference

SRp40 binding site activated

V326F

D

Splice acceptor strength

7.21   74.18

CERT1 pre-mRNA

5' Exon 8 | Exon 9 | Exon 10 3'

WT transcript: Exon 8 | Exon 9 | Exon 10

V326F transcript: Exon 8 | Exon 9 | Exon 10

Partial truncation of Exon 9 (16 residues)

E — Mutational intolerance landscape of CERT1

PH   START

MetaDome scale — Tolerant / Neutral / Intolerant

Deleted residues (311-327)

FFAT motif (residues 321-327)

[vv] *In silico* analysis of the c.976G>T (p.V326F) variant in *CERT1* (NM_005713.2) in Subject 10 who exhibited an intermediate CerTra1 disease phenotype. (A) Sanger sequencing confirmed inheritance of the variant by Subject 10 (red arrow) from her phenotypically unaffected mother. (B) The variant results in substitution of a highly conserved nucleotide (phyloP: 7.52 [-20.0;10.0]) and amino acid residue (valine-236, black arrow), up to *D. melanogaster*. (C) The variant effect is predicted to introduce a binding site for the pre-mRNA-splicing factor SRp40 putatively resulting in aberrant splicing of exon 9 (Alamut v2.3.1). (D) Model illustrating alternatively spliced mRNA products resulting from the c.976G>T (p.V326F) variant. Preference for the significantly stronger downstream exonic acceptor site (SSF-like score = 74.2), compared to the weaker canonical acceptor site (SSF-like = 7.21) may result in excision of 16 amino acid residues (311-327) from the 5' end of exon 9. (E) The predicted excision (green bracket) spatially corresponds to a region of the CERT protein that is highly intolerant to genetic variation, and at the tertiary level, completely abolishes the FFAT motif spanning residues 321-327 (black bracket).

Subject 22: Subject 22 is an 8-year-old Caucasian (Spanish) girl of a consanguineous marriage (the father's maternal grandmother and the mother's maternal grandfather were first cousins). It was noted by the 5th month of pregnancy that the fetus was small-for-gestational-age (SGA). Normal vaginal delivery occurred at the 38th week (induced 3 days prior). Birth weight was 2.5 kg (10th percentile), length 48 cm (10th -25th percentile) and OFC 32 cm (10th percentile); neonatal jaundice, facial dysmorphisms, failure-to-thrive and hypoglycemia were noted. Neurological symptoms were first noted at 4 months of age (scarce movement, no smile, inability to fix and follow). The patient was able to sit at 12 months and began walking between 15 months and 2 years of age, but her seizures, despite being controlled by levetiracetam, appeared to interfere with her continued development. She has severe intellectual disability (does not speak), sleep disturbance, and exhibits various behavioral issues including ADHD/hyperactivity, self-injury (anxiety-induced face-hitting), stereotypy (body-rocking) and other characteristics consistent with autism spectrum disorder (ASD). Frequent vomiting occurred from infancy but stopped after switching to a gluten- and casein-free diet. Due to feeding difficulties continuing from infancy, the patient was placed on an extended regimen of hypercaloric supplementation that seems to have enabled gains in height and head circumference (i.e., skeletal growth) while weight has lagged behind (see table below). At the last examination (8 years, 3 months), the patient remained hypotonic. Neurosensory testing revealed mild right hypoacusia and mild-moderate left hypoacusia. Hearing loss was also noted in her twin sister, mother (partial) and maternal-grandmother. She also shows delayed visual maturation that is likely linked to her overall developmental delay. Brain MRI revealed delayed myelinization, especially in the frontal and temporal lobes. Brain EEG recordings were largely unremarkable, but a video-EEG during sleep showed scarce epileptiform anomalies in left frontotemporal and

temporal regions. Lastly, she has mild tricuspid insufficiency and scalp nevi. The most recent

assessment of craniofacial/skeletal dysmorphia include: high forehead, synophrys, short nose,

anteverted nares, cupid's bow upper lip, hypoplastic antitragus, micrognathia, diastema of upper

central incisors, uplifted ear lobe, midface hypoplasia, medial/central flaring of eyebrows,

groove below lower eyelids, broad/bulbous nasal tip, *pes valgus*, hair growth in the lateral parts

of the front, lateral flaring of eyebrows, arched eyebrows, epicanthal folds, short palpebral

fissures, upslanting palpebral fissures, deep-set eyes, hypoplastic alae nasi, smooth philtrum,

prominent upper central incisors, low-set ears, hypoplastic helix, hypoplastic ear lobe, short

upper limbs (<3rd centile), lower limbs are normal but appear to have scant muscular mass

distally, hypoplastic/ short nails on fingers, hypertrichosis, distal scant muscular mass in the

lower limbs, broad/short and hammer toes, hypoplastic/short nails on toes, 2nd-3rd toes proximal

syndactyly, hypothenar hypoplasia, few palmar creases, low-set and widely-spaced nipples.

Karyotyping was normal. No significant CNV's were detected from aCGH testing except a few

polymorphic CNV. Subtelomeric multiplex ligation-dependent probe amplification (MLPA)

testing was normal. SAICAR, glycosylation and biotinidase screening were also normal. Direct

sequencing of the MECP2 gene detected a synonymous variant, c.603G>A, p.A201=

(rs267608504). The variant is likely benign due to its conservative change and occurrence in the

general population at a minor allele frequency of 0.00005. Whole exome sequencing (WES) of

the patient and both parents identified variation in two candidate genes. Compound heterozygous

variants were detected in the MOCS2 gene. A de novo missense variant, c.985G>C (p.A329P),

was identified in *CERT1* (NM_005713.3). This variant is absent in the general population

according to both gnomAD (gnomad.broadinstitute.org) and BRAVO (bravo.sph.umich.edu),

last assessed September 2021. This variant occurs at a highly conserved position in the protein

and is predicted to be highly pathogenic by several in silico pathogenicity algorithms: CADD = 32, Eigen = 0.68.

Subject 23: Subject 23 is a 5-year-old girl of Chinese descent with a family history of mild intellectual disability in a cousin. She was born at term (36 weeks and 6 days) following an uncomplicated pregnancy, with birth weight of 2.3kg (10th - 50th percentile), length 50cm (50th percentile) and OFC 32.8cm (10th percentile). Maternal Thyroid Stimulating Hormone (TSH) was elevated before conception, but thyroxine restored normal thyroid lab results during pregnancy. First trimester screening was reported to be low risk. The fetal morphology scan at 20 weeks was normal. Oligohydramnios prior to delivery led to a C-section. The newborn was admitted to the NICU for neonatal pneumonia. Neonatal echo and hearing screening were normal. Follow up hearing test revealed concerns about hearing in her right ear, the details of which are uncertain. At 6 months of age she was diagnosed with asymmetry and developmental delay. At age 2 years, her developmental quotient (DQ) was measured to be 37 (25<DQ<39), which falls in the severe range of the evaluation scale developed by the Chinese National Health Commission. She exhibits severe autistic behaviors and aggression. Walking age was 20 months. At 3 years of age, she has a vocabulary of about 5 words. There has been no regression. There are no unusual hand movements. She grabs food but does not feed herself. Sleep is normal. At 3 years of age her weight was measured to be 12.1 kg (10th percentile), height 88.5cm (10th percentile). At this point, microcephaly was noted, with a measured OFC of 46 cm (2nd percentile). Episodic seizures were noted at the most recent examination. Cardiovascular system and skin were normal. Brain MRI and EEG were normal. She had a unilateral inguinal hernia which was repaired. Urine metabolic screening was normal. At her last examination (5 years) her weight was 19kg (50th-70th percentile), height 115cm (50th- 70th percentile), and OFC 46cm

(2nd percentile). Craniofacial/Skeletal features noted during examination: sloping forehead, brachycephaly, occipital flattening, micrognathia, bushy and broad eyebrows, epicanthal folds (ethnic), groove below lower eyelids, anteverted nares, short columella, cupid's bow upper lip, diastema of upper central incisors, gothic palate/ marked alveolar ridges, large ears, hypoplastic and uplifted ear lobe. Metabolic screening, karyotyping/chromosomal microarray and Fragile X testing were normal. A triplication at chromosome 15, 15q11.2 (20262224-22751244)*3, was identified by aCGH. Whole exome sequencing of the patient and parents identified a de novo missense variant, c.989T>C, (p.L330P) in *CERT1* (NM_005713.3) gene. This variant is absent in the general population according to gnomAD (gnomad.broadinstitute.org) and BRAVO (bravo.sph.umich.edu), last accessed September 2021. The variant occurs at a highly conserved position along the protein and is predicted to be deleterious across several in silico pathogenicity algorithms: M-CAP = 0.084, CADD = 30, REVEL = 0.674 and Eigen = 0.7185.

Subject 24: The little clinical information available on Subject 24 notes him or her to have ADHD; impulsivity; hyperactivity; mild intellectual disability; myopia; high forehead; low-set ears. Singleton screening identified c.1097G>C (p.R366T) in *CERT1* (NM_005713.3).

Subject 25: Subject 25 is a 3-year old boy who recently presented with autism spectrum disorder (ASD); mild intellectual disability; behavioral issues; and mild speech delay. Molecular Inversion Probe (MIP) screening in the patient and parents identified a de novo variant, c.1141A>G, (p.I381V), in the *CERT1* (NM_005713.3) gene. This variant was found in an individual (female) of European (non-Finnish) descent in the general population at a minor allele frequency of 0.00000398 in gnomAD (https://gnomad.broadinstitute.org/variant/5-74695182-T-C?dataset=gnomad_r2_1 (assessed April 2020).

Subject 26: Subject 26 is a 2-year, 5-month old girl of Spanish (Mediterranean) descent whose parents also have developmental delay. Clinical abnormalities were noted prenatally. She was born by vaginal delivery at term (40 weeks and 5 days) following an uncomplicated pregnancy but was smaller than average: birth weight was 2.8kg (-1.62 SD), length was 47 cm (-2.15 SD) and head circumference was 33 cm (-1.62 SD). She was hypotonic during infancy but did not exhibit feeding difficulties or failure to thrive. She was diagnosed with ventricular septal defect postnatally. She has achieved major motor milestones: head-holding (3 months), sitting (10 months) and independent walking (24 months). Speech was limited to approximately 5 discernible words at 2 year and 5 months of age. Behaviorally, she exhibits occasional outbursts and stereotyped/ repetitive movements that are consistent with mild autism spectrum disorder (ASD). She has difficulty falling asleep and precocious awakenings, both of which are partially responsive to melatonin. Gastrointestinal issues include constipation. Recent lab tests revealed anemia. Neuro-imaging was not performed. As of the most recent examination at 2 years, 5 months, she has made some gains in growth parameters: weight is 13 kg (-0.7 SD), height 88 cm (-1.37 SD) and OFC 48.5 cm (-1.08 SD). Both fine and gross motor skills remain impaired. She has mild craniofacial/skeletal dysmorphisms: cupid's bow upper lip, brachydactyly, tubular nose, flared nares, full cheeks, medial/central flaring of eyebrows, upslanting palpebral fissures, broad/bulbous nasal tip, swelling lower eyelids, protruding columella, lateral thickening under the lower lip, groove below lower eyelids, hypoplastic alae nasi, hypoplastic/short nails on fingers. aCGH testing was negative for copy number variants. Singleton whole exome screening identified the variant, c.1346C>T (p.A449V), in *CERT1* (NM_005713.3) which was subsequently found to be inherited from her developmentally delayed mother by Sanger sequencing. A history of schizophrenia was noted in her maternal grandmother in whom genetic

248

screening for this CERT1 variant is currently being arranged. The c.1346C>T (A449V) variant is predicted to be deleterious by Mutation Taster, LRT and Fathmm and is reported in one healthy male of Latino/American admixture in gnomAD (MAF = 0.000004121).

Subject 27: Subject 27 is a 3.5-year old girl of Caucasian and Filipino descent who was born at term (39 wks+4 days) by vaginal delivery at a birth weight of 3.28kg, length 52cm and OFC of 33cm (3rd-10th centile) with an otherwise unremarkable neonatal history. During early infancy, she suffered from transitory feeding difficulties and exhibited failure to thrive, resulting in a body mass index (BMI) in the 3rd percentile. She was noted to have global developmental delay at 5 months of age. She was delayed in meeting gross motor milestones: head-holding (10 months) and sitting (14 months). Although she learned to walk at 36 months, she exhibited a wide, possibly ataxic gait. At a recent examination (3 years and 6 months old), her height was normal for her age (95.4 cm), but she was moderately underweight (11.2 kg, <5th centile); OFC measurement (45.5 cm, <3rd percentile) at this time indicated microcephaly. She had severe intellectual disability (54%, Griffiths scale), was non-verbal and showed stereotypic behavior suggestive of autism spectrum disorder (ASD). There was no history of seizures or sleep disturbances, although she appeared to have a high pain tolerance. A functional heart murmur was also noted although ultrasound tests were normal. She is otherwise healthy. Trio exome sequencing of Subject 13 and both unaffected parents identified a de novo missense variant, c.1499C>T (p.P500L), in the *CERT1* (NM_005713.3). The variant occurs within the functional START domain of *CERT1* and is predicted to be deleterious by several prediction algorithms. It is absent from both gnomAD (gnomad.broadinstitute.org) and BRAVO (bravo.sph.umich.edu), last checked September 2021.

Subject 28: Subject 28 is a 12-year and 10-month old Caucasian boy whose early history was unremarkable. He was born at 40 weeks by vaginal delivery at a birth weight of 3.6 kg and normal birth length and OFC. Sitting and independent walking occurred at 8 and 18 months, respectively. He presented with mild motor and speech delays noted at approximately 2 years of age. His first words were delayed and his ability to use multiple words together occurred between ages 3-4 years. Intellectual disability became apparent over time but is relatively mild. A Wechsler Preschool & Primary Scale of Intelligence-IV (WPPSI-IV) full scale intelligence quotient fell <1 percentile at age 4 years and 6 years. He exhibits no significant behavioral anomalies with the exception of ADHD. As of his most recent examination at the age of 12 years, there has been no apparent regression. He is hypotonic and has constipation and headaches. His weight at age 11- years, 9-months was 29.5kg (z = -1.48), height at age 10-years, 10-months was 139.6cm (z = -0.4) and OFC at 9-years, 2-months was 53.5cm (z = -0.76). A hyper-pigmented macule (2.5 x 0.7 cm) was noted on skin on the abdomen; otherwise, there is no unusual hue. Mild facial dysmorphia include large ears, deep set eyes, long tubular nose with hypoplastic nasal alae, a thin upper lip (see accompanying frontal and lateral facial photographs). There is an accessory nipple. Family history includes a sibling with autism spectrum disorder (level I) and normal cognition, and other family members with dyslexia. Fragile X and array-CGH screening were negative. A gene panel test of >2,600 genes identified the de novo occurrence the variant, c.934G>C (p.G312R) variant in *CERT1* (NM_005713.3). The variant is absent from gnomAD (https://gnomad.broadinstitute.org/gene/ENSG00000113163?dataset=gnomad_r2_1) and is predicted deleterious by multiple pathogenicity algorithms (CADDv1.6 = 28.9; M-CAP = 0.153).

Subject 29: Subject 29 is a 16-year old girl of Indian descent who was born 36 weeks+5 days by spontaneous vaginal delivery at a birth weight of 2.79 kg (75th – 97th percentile), length 47.5 cm (10th percentile) and OFC of 31.5 cm (3rd percentile). She was noted to be hypotonic. She exhibited motor delays at around 1 year of age but acquired walking by 16 months. At 2 years of age, she was diagnosed with autism spectrum disorder by a psychologist primarily on the basis of stereotyped behaviors, and she had her first seizure, reportedly tonic-clonic. Despite her relatively mild early delays, she has developed severe global developmental delay and severe intellectual disability (she is non-verbal), possibly due to seizures. Her seizures were treated with levetiracetam (Keppra). At age 11, an EEG study showed epileptiform discharges in the right frontal–temporal region (during sleep), but as of her most recent examination at 15 years of age, she had been seizure-free for the past year. She continues to exhibit stereotypical behaviors such as hand flapping. She has profound visual field defects in the right eye and is suspected to have cortical blindness, since her anterior and posterior segments appear healthy. External ocular exam revealed right gave vs left field preference, full extraocular movements and visual field full to confrontation. She has maintained her small stature: at age 15, her height was 139.5 cm (<3rd percentile), weight 32.7 kg, (<3rd percentile), OFC 51 cm (3rd percentile). An MRI of the brain at this time showed mild thinning of the anterior part of the corpus callosum with slightly reduced frontal parietal cerebral volume. MR spectroscopy of the ventricles showed a medium-sized lactate peak, of indeterminate significance. A small, incidental arachnoid cyst over the left frontal convexity was noted but not thought to exert any mass effect. The patient was diagnosed with congenital adrenal hyperplasia (CAH, 21-hydroxylase deficiency); she has occasional abdominal pains and headaches, ambiguous genitalia, and salt wasting. The CAH could also contribute to her small stature, which is proportional (whereas some CerTra patients are

underweight for their height). Metabolic testing including plasma acylcarnitine and plasma amino acids were normal. Hgb (low MCV) and K levels are occasionally low. Family history is reportedly negative for ID or developmental delay syndromes. Genetic screening by BluePrint Genetics identified a de novo heterozygous missense variant c.1372C>G, p.(Leu458Val) in *CERT1*, a hemizygous c.955C>T, p.(Gln319*) variant in *CYP21A2*, inherited from her father and a *CYP21A2* deletion from her mother. These latter mutations of the gene that encodes the steroid 21- hydroxylase, which is involved in the synthesis of aldosterone and cortisol, is consistent with the congenital adrenal hypoplasia.

Subject 30: Subject 30 is a 14-year-old boy of Caucasian descent who was born at term (38.5 weeks) by vaginal delivery at a birth weight of 7 lbs and 19 inches in length. He presented with global developmental delays between 6 and 8 months of age. He sat at 10 months and walked with an unsteady gait at 17 months, at which time developed seizures. At most recent examination his intellectual disability is estimated to be severe to profound. He is non-verbal but makes frequent repetitive vocalizations. He exhibits repetitive behavior consistent with ASD, and he is treated with lorazepam (Ativan) for generalized anxiety disorder. He has episodes of emesis when anxious or distressed and hits his chin when irritated. He has generalized convulsive epilepsy, but seizures are infrequent and controlled with lacosamide (Vimpat) at adjusted doses over the years. The EEG performed in 2008 was normal; repeat EEG (performed a year later in 2009) revealed excessive slowing in the right hemisphere without epileptic activity; a subsequent EEG (performed in 2013) revealed significant dysfunction on the left hemisphere with epileptic tendencies arising from both cerebral hemispheres and an asymmetry on the background with slower frequencies noted over the left hemisphere. The most recently EEG (in 2018) did not reveal any epileptic, focal or diffuse abnormality while the patient was

awake, even while having abnormal movements. His most recent weight was 43.1 kg (11th percentile, Z= -1.25, based on CDC data), height 166 cm (45th percentile, Z= -0.14, based on CDC data), and OFC 54.2 cm (32nd percentile, Z= -0.46, based on Nellhaus data). There is a history of constipation, urinary tract infections with pain and microscopic hematuria. Renal ultrasound and brain MRI have been unrevealing. Metabolic testing showed normal CBC with differential, comprehensive metabolic panel, Vitamin D, thyroid function studies, pyruvic acid, lactic acid, creatine kinase, total and free carnitine, plasma amino acids, acylcarnitine profile, oligosaccharides, mucopolysaccharidosis urine electrophoresis, urine organic acids, plasma and urine GAA levels, and spinal fluid amino acids, lactic acid, and neurotransmitters; plasma and urine creatine levels were elevated but repeat urine testing on a morning void revealed normal urine creatine levels. Global Metabolomic Assisted Pathway Screening (Global MAPs) revealed no significant perturbations of metabolic pathways. Facial dysmorphisms include palpebral fissures and almond-shaped eyes with mild epicanthal folds bilaterally. Infraorbital creases are noted bilaterally. Flat nasal bridge with short nasal tip. The mouth is typically held open. Pes planus were noted in the feet and hair exhibits normal texture and distribution. Karyotyping, Fragile X screening and array-CGH screening were negative. A Syndromic Autism panel was normal. Epilepsy panel testing revealed one maternally inherited variant in the SCN11A gene that is likely not pathogenic nor clinically significant. Direct sequencing of *ARX*, *MECP2*, *UBE3A* and *SHANK3* were negative. Prader-Willi/Angelman syndrome methylation studies were also negative. Trio exome sequencing identified a de novo missense variant, p.S141R, in *CERT1* (NM_005713.3).

Subject 31: Subject 31 is a 6-year-old boy of Indian descent diagnosed with epileptic encephalopathy and developmental delay beginning at 9 months of age. He was born at term (38

weeks) by vaginal delivery at a birth weight of 2.5 kg. Feeding difficulties and failure to thrive were noted during the neonatal and infantile periods. Head-holding occurred at 9 months, sitting at 12 months and walking at 18 months. He exhibits moderate intellectual disability, speech delay (some verbal ability) and motor delays. No regression has been noted. Behaviorally, he exhibits ADHD, aggressive behavior, self-injurious behavior and repetitive movements consistent with autism spectrum disorder. Most recent OFC measurement (47 cm) is consistent with microcephaly, although we have no information on his height and weight. Brain MRI was normal. Brain EEG showed generalized epileptiform discharges. Seizures were reported to be generalized tonic-clonic (1-2 episodes/week) and myoclonic jerks 5-7 times/day. The patient has been treated with valproate sodium, levetriacetam, and clobazam and has sleep disturbances. He has exotropia. There are no apparent deviations in pain tolerance. Other than constipation, there are no other significant systemic issues. Metabolic, blood/serology, hormone (TSH) and urinalysis testing was normal. The face was noted to be long but otherwise not dysmorphic; there are no observable abnormalities of the hands and feet. Karyotyping, Fragile X screening, array-CGH and gene panel testing were negative. Trio exome sequencing identified a de novo missense variant, p.E424G, in *CERT1* (NM_005713.3).

Appendix E: Longitudinal Analysis of a Resolving Foveomacular Vitelliform Lesion in ABCA4 Disease[ww]

Background

Pathogenic variation in the gene encoding the ATP-binding cassette, sub-family A, member 4 (ABCA4) protein is the underlying cause of a retinal degenerative disorder characterized by progressive deterioration of central vision over time.[1] To date, >1,500 disease-causing variants have been identified in patients who exhibit a wide range of clinical phenotypes depending on the severity of the causal variants.[2,30] Mild, hypomorphic variants result in a slow progressing, late-onset maculopathies in which the preservation of central vision persists into 5th and 6th decades of life while variants resulting in complete loss-of-function (nonsense, frameshift, canonical splice site) are associated with severe chorioretinopathies that begin at early adolescence and rapidly progress to widespread degeneration across the posterior pole in a short period of time.[2,284]

ABCA4 disease exhibits many characteristic disease features that have become valuable diagnostic aids for ophthalmologists in the clinic. In addition to the expanding degenerative lesion in the central macula, other recognizable features include heterogeneous patterns of yellow, pisciform flecks across the fundus, peripapillary sparing,[299] and specific retinal imaging findings such as a dark choroid on fundus angiography[300] and increased levels of fundus autofluorescence.[22,24] Although widely regarded as pathognomonic characteristics of ABCA4 disease, some of these features may also be found to varying degrees in other genetically distinct disorders. For instance, both yellowish flecks and peripapillary sparing are present in certain

---

[ww] Text and figures from this chapter have been recently published (Lee et al. *Ophthalmology Retina*. 2022 Sep;6(9):847-860).

variants of *PRPH2*-associated pattern dystrophy (MIM# 169150).[301] Likewise, rare

complications typically associated with other non-inherited conditions such as choroidal

neovascularization,[302,303] pigment epithelial detachments[304] and macular holes[305,306] have also

been documented in ABCA4 disease. Consolidating the clinical landscape of both common and

infrequent sub-phenotypes for a disease elevates patient care in the clinic and advances research

efforts at the forefront of medicine.

This report details the longitudinal examination of a 25-year-old man who presented to the clinic

with bilateral vitelliform lesions and recent history of central vision loss. The patient was

followed over a 35-month period during which electrophysiological testing and serial retinal

imaging was performed to monitor the clinical course of the lesions in both eyes. Exome

sequencing analysis was performed and known pathogenic *ABCA4* variants were identified as the

underlying cause of disease. Retinal imaging data were quantitatively analyzed to identify

features that distinguish this rare manifestation of ABCA4 disease from other vitelliform

associated diseases.

Clinical phenotype

A 25-year-old Caucasian man (Patient 1, **Appendix Table 1**) presented to the clinic with a

referring diagnosis of autosomal dominant Best vitelliform macular dystrophy (VMD2, MIM#

153700). Progressive deterioration of central vision reportedly began at 20 years of age.

Systemic history was unremarkable and no family history of retinal disease was noted. Best-

corrected visual acuity (BCVA) was Snellen 20/60 in both eyes after correcting for mild

refractive error (OD: plano; OS: +0.50 −0.25 x 142). Intraocular pressures (IOP) were 16 mmHg

in the right eye and 13 mmHg in the left eye. Slit-lamp examination of the anterior segment was

unremarkable.

Appendix Table E1.[xx]

Patient	Sex	Age	AO	Snellen BCVA		F/U duration (years)	Optical gap cavitation	ABCA4 variant 1		ABCA4 variant 2	
				OD	OS			cDNA change (NM_000350.2)	Protein change (NP_000341.2)	cDNA change (NM_000350.2)	Protein change (NP_000341.2)
1	M	25	20	20/60	20/60	2.9	✓	c.5882G>A	p.(Gly1961Glu)	c.4139C>T	p.(Pro1380Leu)
2	M	24	20	20/40	20/40	-		c.5882G>A	p.(Gly1961Glu)	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])
3‡	M	34	25	20/30	20/30	-		c.5882G>A	p.(Gly1961Glu)	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])
4‡	M	34	25	20/30	20/30	-		c.5882G>A	p.(Gly1961Glu)	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])
5	M	18	16	20/50	20/60	-		c.[2470A>G;3113C>T]	p.([Ile824Val;Ala1038Val])	c.247_250dup	p.(Ser84Thrfs*15)
6	F	14	14	20/60	20/40	5.3		c.5882G>A	p.(Gly1961Glu)	c.1622T>C	p.(Leu541Pro)
7‡	F	25	24	20/80	20/50	2.7		c.[769-784C>T;5882G>A]	p.([Leu257Aspfs*3,Gly1961Glu])	c.286A>G	p.(Asn96Asp)
8‡	F	22	21	20/40	20/50	-		c.[769-784C>T;5882G>A]	p.([Leu257Aspfs*3,Gly1961Glu])	c.286A>G	p.(Asn96Asp)
9‡	F	24	14	20/150	20/150	1.4		c.5882G>A	p.(Gly1961Glu)	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])
10‡	F	20	14	20/80	20/70	2.5		c.5882G>A	p.(Gly1961Glu)	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])
11	M	24	20	20/40	20/80	-	✓	c.5882G>A	p.(Gly1961Glu)	c.4234C>T	p.(Gln1412*)
12	M	28	23	20/30-1	20/30-1	1.3		c.5882G>A	p.(Gly1961Glu)	c.3065A>G	p.(Glu1022Gly)
13	F	27	15	20/100	20/100	3.1	✓	c.[769-784C>T;5882G>A]	p.([Leu257Aspfs*3,Gly1961Glu])	c.6448T>C	p.(Cys2150Arg)
14	F	23	18	20/40	20/30	-		c.5882G>A	p.(Gly1961Glu)	c.4139C>T	p.(Pro1380Leu)
15	M	12	10	20/150	20/100	1.4	✓	c.6449G>A	p.(Cys2150Tyr)	c.2461T>A	p.(Trp821Arg)
16	F	24	21	20/50	20/100	1.7		c.5882G>A	p.(Gly1961Glu)	c.5196+1056A>G	p.(Met1733Valfs*2)
17‡	F	27	17	20/100	20/80	5.0	✓	c.5882G>A	p.(Gly1961Glu)	c.1622T>C	p.(Leu541Pro)
18‡	F	24	15	20/80	20/80	5.0	✓	c.5882G>A	p.(Gly1961Glu)	c.1622T>C	p.(Leu541Pro)
19	M	23	22	20/40	20/30	-		c.5882G>A	p.(Gly1961Glu)	c.5318C>T	p.(Ala1773Val)

[xx] Demographic, clinical and genetic characteristics of the index case and other patients exhibiting the optical gap phenotype of ABCA4 disease. M, male; F, female; AO, age of onset; BCVA, best-corrected visual acuity; OD, right eye; OS, left eye; F/U, follow-up; ‡, sibling pairs (P3 and P4; P7 and P8; P9 and P10; P17 and P18); Patient 1 is the proband described in the study. Gametic phase (i.e., biallelic compound heterozygosity) of the ABCA4 variants was confirmed by parental screening in: P2, P5, P6, P7, P8, P9, P10, P11, P13, P14, P15, P16, P17 and P18. The following patients have been previously described in Noupuu et al (Invest Ophthalmol Vis Sci. 2014 Oct 9;55(11):7217-26): P2, P3, P7, P8, P9, P10, P13, P14, P15, P16, P17, P18 and P19.

A dilated funduscopic exam showed normal optic nerves, healthy retinal vessels and no disease

changes in the periphery. Examination of the macula revealed round sub-retinal lesions

approximately 1 mm in diameter filled at the base with a dense, egg-yolk-like deposited

substance in both eyes (**Appendix Figure E1A, inset**). The deposit exhibited an intense

autofluorescence signal and was consistent in appearance with vitelliform fluid (**Appendix**

Figure E1B, inset) at the vitelliruptive (scrambled egg) stage in the right eye and

pseudohypopyon stage in the left eye. Horizontal SD-OCT scans through the fovea showed a

focal disruption of the photoreceptor-attributable inner segment ellipsoid (ISe) band resulting in

a cavitated, hyporeflective subretinal space between the RPE and neurosensory retina (**Appendix**

Figure E1C). In a more inferiorly positioned SD-OCT scan, the vitelliform fluid was reflective

displaced the overlying external limiting membrane, outer nuclei and inner retinal layers

(**Appendix Figure E1D**).
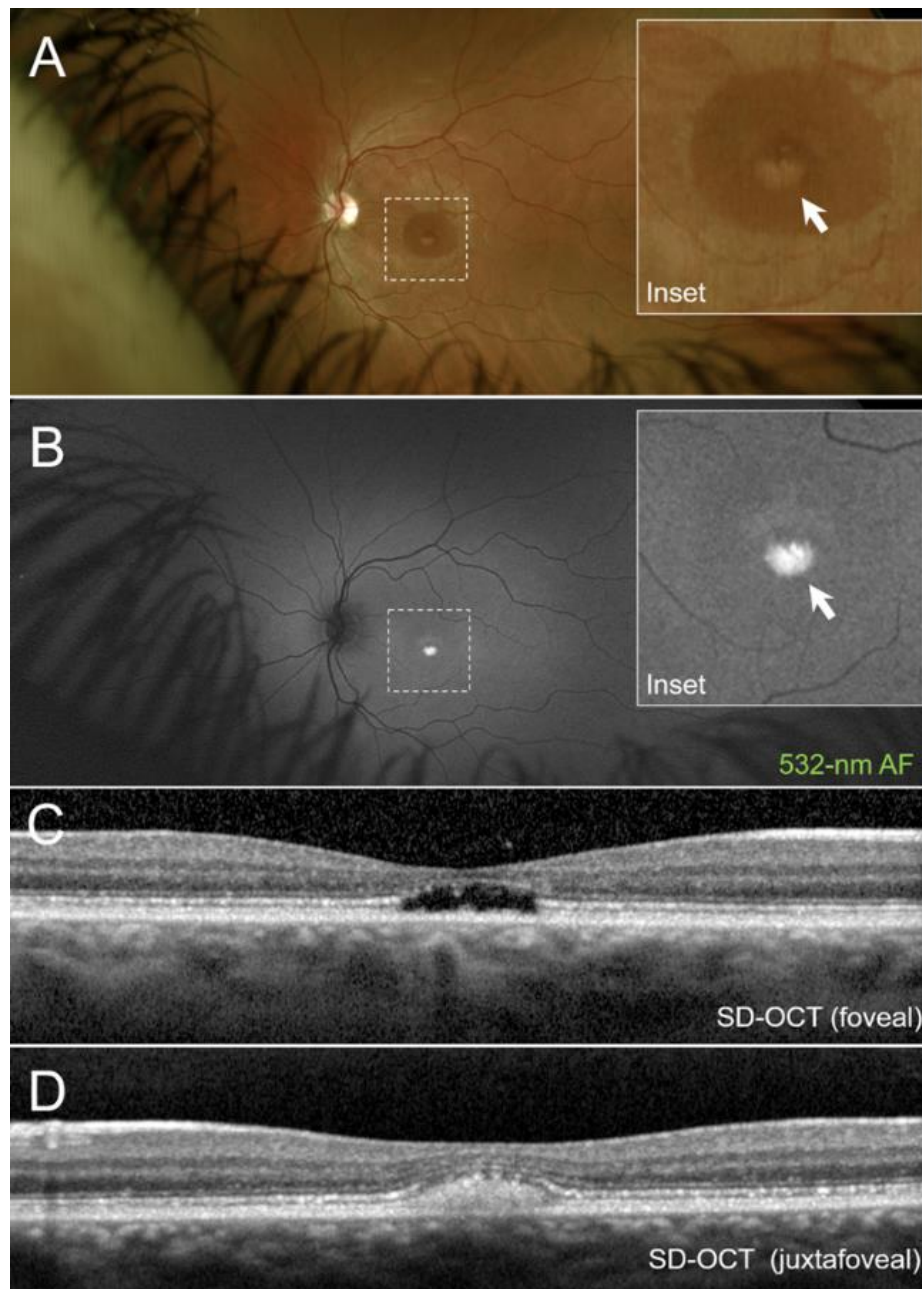
Electrophysiology: EOG and ffERG recordings

[yy] Clinical presentation of a foveomacular vitelliform lesion in 25-year-old Caucasian man with confirmed ABCA4 disease. Ultra-wide field photograph (A) and 532-nm autofluorescence image (B) of fundus in the left eye and enlarged insets showing the sub-foveal lesions with vitelliform fluid (white arrows). (C) Horizontal spectral domain-optical coherence tomography (SD-OCT) scans through the fovea showing focal loss of the inner segment ellipsoid band resulting in a cavitated optical gap lesion. (D) Horizontal SD-OCT scan through the vitelliform fluid accumulation at the base of the lesion (inferior to the fovea).

Electrooculogram (EOG) testing was performed and showed notable light rise following the dark

adaptation phase in both eyes, from 12.5μV to 18.9μV in the right eye and 13.3μV to 20.5μV in

the left eye (**Appendix Figure E2A**). The average Arden ratio (light peak-to-dark trough)

between both eyes was 1.62. Full-field ERG (ffERG) testing also showed no significant

generalized dysfunction of the cone and rod systems across the retina (**Appendix Figure E2B**).

Appendix Figure E2[zz]
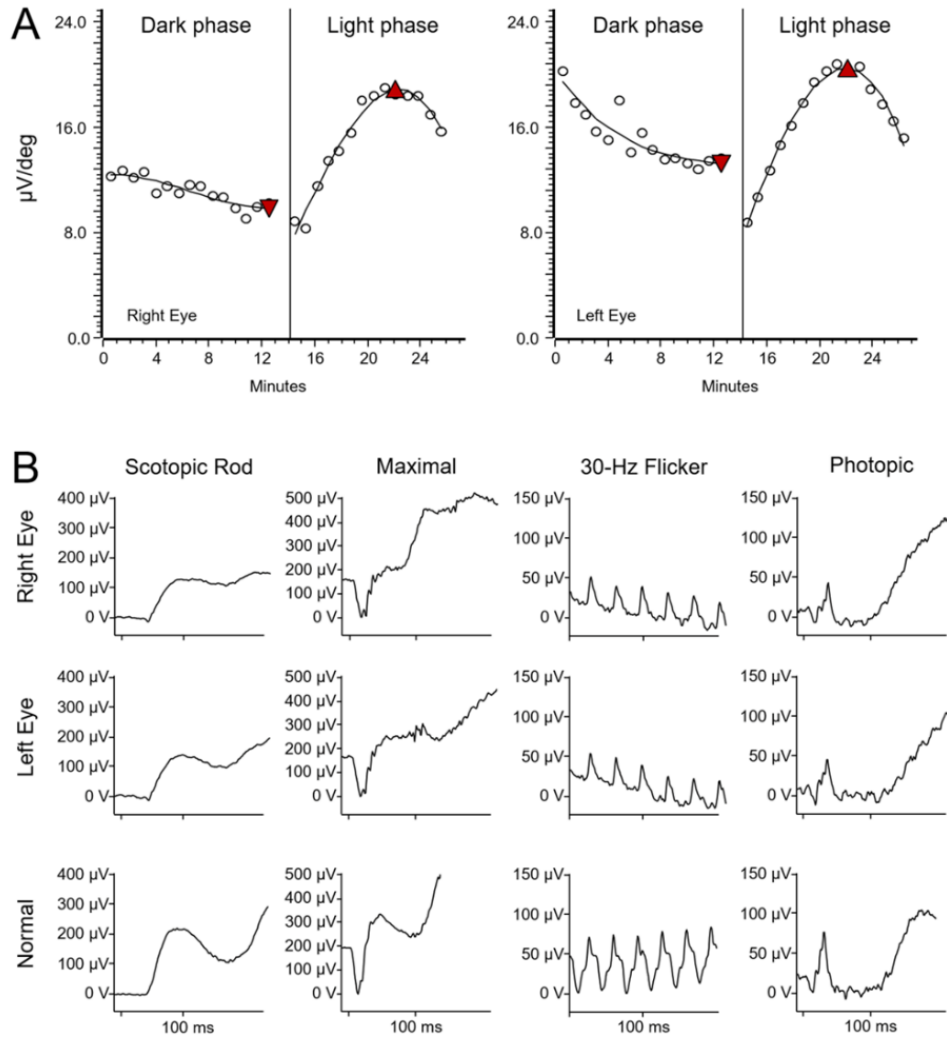
[zz] Electrooculogram (EOG) and full-field eletroretinogram (ffERG) testing results in a patient with foveomacular vitelliform-associated ABCA4 disease. (A) EOG recordings in the patient showing a light rise from dark trough to light peak (red triangles) in both eyes. (B) Full-field scotopic (dark-adapted 0.01 rod), maximal (dark-adapted 3.0 combined rod and cone), 30-Hz flicker and photopic (light-adapted 3.0 single flash cone) electroretinogram responses of the right and left eyes of the patient and representative waveforms from an age-matched healthy control eye. Abbreviations: μV, microvolts; deg, degrees (polar angle, where entire ocular globe ≈ 360° or 2π; the typical (monocular) visual field of a human eye ranges ~120° across the vertical meridian and ~160° across the horizontal meridian at the posterior pole); ms, microseconds.

Total retina, OS+, ONL+ and RPE layer thickness

Retinal layers on horizontal SD-OCT scans were segmented in both eyes and compared to the respective mean thicknesses ± 2 standard deviation intervals of healthy individuals/eyes (n = 25) (**Appendix Figure E3A-E3D**). Outer segment (OS+) layer thickness in both eyes were within normal limits. This layer was disrupted at the location of lesion in the fovea (**Appendix Figure E3B**). Significant thinning of the outer nuclear layer + outer plexiform layer (ONL+) was observed inside the lesion and, notably, within a ~1-2 mm radius outside of the foveal lesion (**Appendix Figure E3C**). No significant thinning was apparent in the RPE layer (**Appendix Figure E3D**).

Appendix Figure E3[aaa]
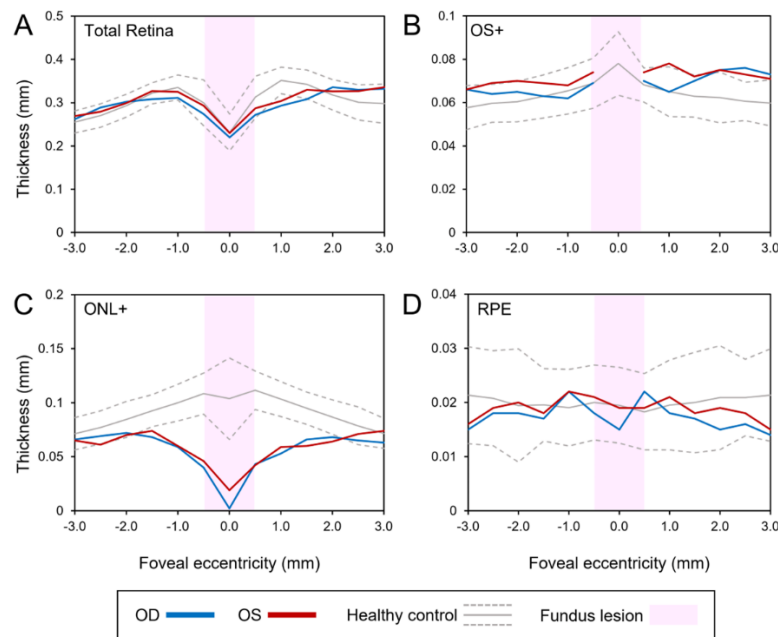
[aaa] Assessment of total retinal, outer nuclear layer + outer plexiform layer (ONL+), outer segment+ (OS+) and RPE thickness along horizontal spectral domain-optical coherence tomography (SD-OCT) scans of both eyes in the patient and 25 healthy individuals/eyes. Solid and dotted gray lines are the mean thickness values of healthy eyes ± 2 standard deviations, respectively. Pink shaded regions denote the location of the visible lesion in the fovea. Thicknesses were plot as a function of eccentricity from the fovea, 3 mm in the nasal (positive values) and temporal (negative values) directions. ONL+ is defined as the distance between the external limiting membrane (ELM)and the proximal edge of the outer plexiform layer (OPL); OS+ is defined as the distance between BM and the anterior boundary of the ellipsoid zone band.

En face analysis: ISe and RPE layers

On *en face* maps of ISe reflectivity in both eyes, an elliptical hyper-reflective halo was observed surround the non-reflective central lesion in the fovea (**Appendix Figure E4A**). No significant regions of abnormal reflectivity were apparent at the RPE layer (**Appendix Figure E4B**). When compared to other imaging modalities, this peri-lesional abnormality was not visible on SW-AF (**Appendix Figure E4C**) but manifestly apparent on NIR-AF as a homogeneous region of hyperautofluorescence (**Appendix Figure E4D**).

Appendix Figure E4[bbb]
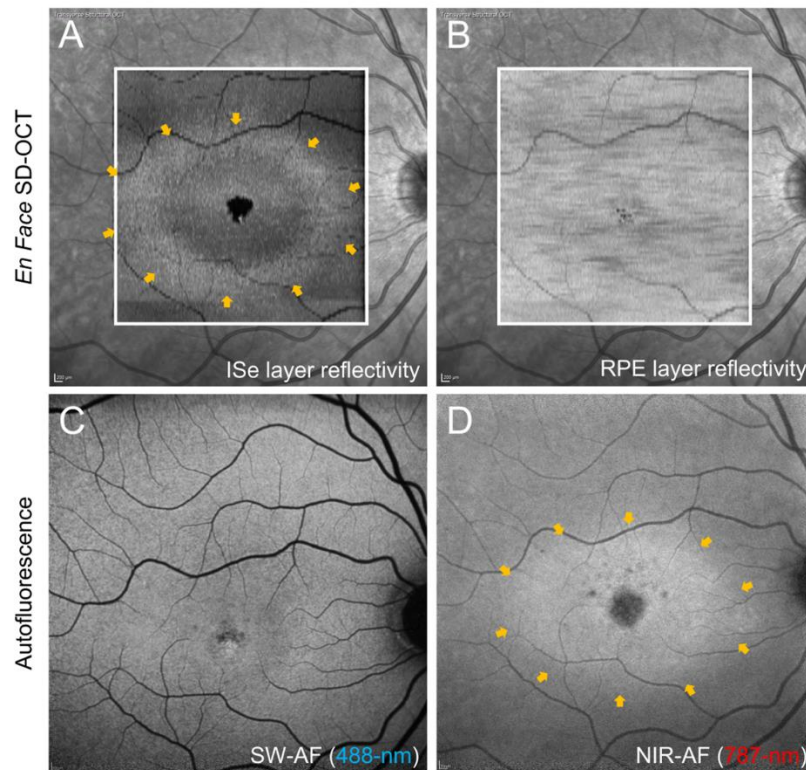
[bbb] Multi-modal imaging of the macula in the right eye of the patient with foveomacular vitelliform-associated ABCA4 disease. (A) 8mm x 8mm en face map of the photoreceptor-attributable inner segment ellipsoid (ISe) layer constructed from the transverse analysis of 97 spectral domain-optical coherence tomography (SD-OCT) scans. The en face map is overlaid onto a near-infrared reflectance image. Yellow arrows point to the edge of an abnormal region of reflectance around the foveal lesion. (B) 6mm en face map of the retinal pigment epithelium layer showing no apparent regions of abnormal reflectivity. (C) Short wavelength-autofluorescence (SW-AF, 488-nm) and (B) near infrared-autofluorescence (NIR-AF, 787-nm) images of the right eye. Yellow arrows in the NIR-AF image point to the edge of a homogeneous region of increased autofluorescence around the foveal lesion.

Long-term progression and disease outcome

The patient was monitored over a 35-month period to assess longitudinal changes. BCVA remained relatively stable throughout the course of follow up although the left eye slightly improved from 20/60 to 20/50 after 30 months (**Appendix Figure E5A**). Serial SD-OCT scans through the fovea showed no changes to the optical gap lesions in the fovea aside from the variable appearance of small, punctate spots across 4 visits in both eyes (**Appendix Figure E5B**). After months from initial presentation, the lesion in the left eye become vitelliruptive and most of the vitelliform fluid had resolved in the right eye (**Appendix Figure E3C**). At this time, small, well-delineated foci exhibiting an intense autofluorescence signal were observed inside the lesions on autofluorescence imaging. These foci were hyper-reflective on SD-OCT scans and traversed both photoreceptor-attributable layers. At 19 months, these intraretinal hyperreflective foci were no longer visible in the right eye and began migrating towards the inner retina in the left eye (**Appendix Figure E3C**). At this visit, an optical coherence tomography angiography (OCTA) map of the choriocapillaris layer showed a bright visibility of large, underlying choroidal vessels within the lesion in the right eye (**Appendix Figure E3A**). This underlying vessel visibility was less pronounced in the lesion of the left eye, in which two obstructive spots were present. After 30 months, vitelliform fluid had completely resolved in both eyes leaving empty optical gaps with residual hyperreflective puncta traversing the ELM.

Appendix Figure E5[ccc]
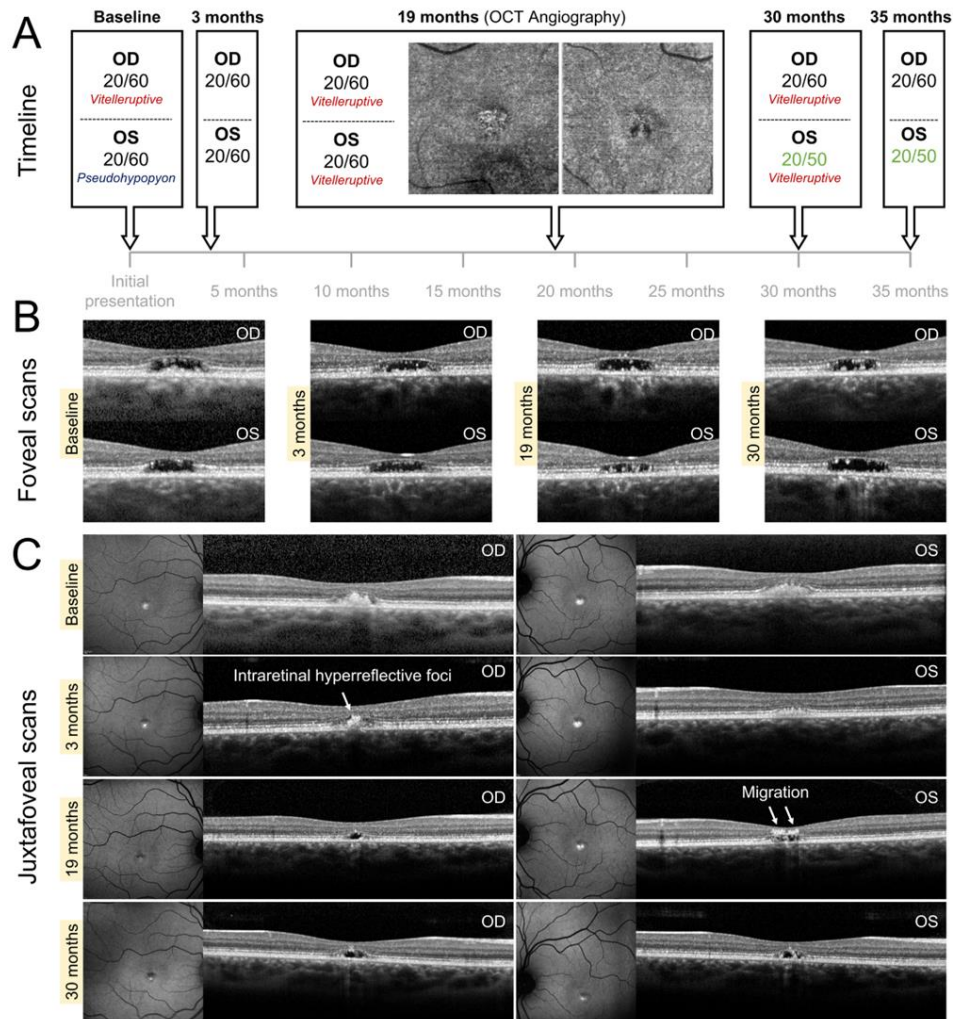
[ccc] Longitudinal analysis of the patient with foveomacular vitelliform-associated ABCA4 disease over a 35-month examination period. (A) Timeline providing the best-corrected visual acuity (BCVA) and vitelliform stage in the right (OD) and left (OS) eyes. 6mm x 6mm optical coherence tomography angiography (OCTA) of the choriocapillaris layer of the right and left eye were acquired 19 months from the baseline visit. (B) Enlarged horizontal SD-OCT scans through the fovea in both eyes across four visits from baseline to 30 months. (C) Enlarged horizontal SD-OCT scans through the vitelliform fluid accumulation at the base of the lesion (inferior to the fovea) across four visits from baseline to 30 months.

Exome sequencing analysis

Whole exome sequencing was performed in the patient and variant filtering was restricted to genes previously associated with retinal disease (RetNet, https://sph.uth.edu/retnet/). At an initial minor allele frequency (MAF) filter of ≤ 0.2, variants in two vitelliform-associated genes were identified: a homozygous missense variant, c.1609T>C (p.(Ser537Pro)), in *BEST1* (MIM# 607854) and two heterozygous missense variants, c.1898G>A (p.Arg633His) and c.1876C>T (p.Arg626Trp), in *IMPG1* (MIM# 602870). These variants were not considered pathogenic due to their extremely high MAF (>0.1) (**Appendix Table E2**). At MAF filter ≤ 0.01, 15 heterozygous variants in 12 genes remained of which half are associated with syndromic an/or non-syndromic retinitis pigmentosa (**Appendix Table E3**). Considering the underlying phenotype in the patient, the c.5882G>A (p.(Gly1961Glu)) and c.4139C>T (p.(Pro1380Leu)) variants in *ABCA4* were determined to be the causal variants. Interestingly, a heterozygous missense variant c.2644C>T (p.(Arg882Trp)) was identified in *RP1L1* (MIM# 608581) which causes a similar phenotype (occult macular dystrophy) although it was not predicted to be pathogenic (CADD-PHRED = 7.132, REVEL = 0.113, Eigen = -1.668).

Appendix Table E2[ddd]

Gene	Transcript ID	Genomic location	HGVS cDNA	HGNS protein	Zygosity	Variant type	Variant effect	RefSNP	MAF	CADD-PHRED	M-CAP	REVEL	Eigen
BEST1	NM_004183.3	11-61730553-T-C	c.1609T>C	p.(Ser537Pro)	Homozygous	Substitution	Missense	rs17185413	0.1769	13.68	-	0.085	-1.195
IMPG1	NM_001563.4	6-76640781-C-T	c.1898G>A	p.(Arg633His)	Heterozygous	Substitution	Missense	rs3734313	0.1875	17.8	-	0.071	-0.319
IMPG1	NM_001563.4	6-76640803-G-A	c.1876C>T	p.(Arg626Trp)	Heterozygous	Substitution	Missense	rs10943299	0.1195	0.045	-	0.047	-2.376

Appendix Table E3[eee]

Gene	Transcript ID	Genomic location	HGVS cDNA	HGNS protein	Zygosity	Variant type	Variant effect	RefSNP	MAF	CADD-PHRED	M-CAP	REVEL	Eigen
ABCA4	NM_000350.2	Chr 1:94473807	c.5882G>A	p.(Gly1961Glu)	Heterozygous	Substitution	Missense	rs1800553	0.0048	28.4	-	0.76	0.346
ABCA4	NM_000350.2	Chr 1:94496666	c.4139C>T	p.(Pro1380Leu)	Heterozygous	Substitution	Missense	rs61750130	0.0002	25.2	0.391	0.867	0.696
ADAM9	NM_003816.3	Chr 8:38869207	c.226G>A	p.(Glu76Lys)	Heterozygous	Substitution	Missense	rs61753672	0.0065	19.17	-	0.059	-0.617
ALMS1	NM_015120.4	Chr 2:73676961	c.3304C>G	p.(Pro1102Ala)	Heterozygous	Substitution	Missense	rs200257398	0.0006	15.28	0.014	0.036	-0.35
ARHGEF18	NM_001367823.1	Chr 19:7447680	c.725G>A	p.(Ser242Asn)	Heterozygous	Substitution	Missense	rs545108657	0.0005	11.73	0.015	-	-
CEP290	NM_025114.4	Chr 12:88454728	c.6401T>C	p.(Ile2134Thr)	Heterozygous	Substitution	Missense	rs117852025	0.0075	24.6	-	0.406	0.807
EYS	NM_001142800.2	Chr 6:65016935	c.6119T>A	p.(Val2040Asp)	Heterozygous	Substitution	Missense	rs201580493	0.0017	26.6	0.107	0.296	0.27
EYS	NM_001142800.2	Chr 6:65596611	c.2971C>T	p.(Leu991Phe)	Heterozygous	Substitution	Missense	rs201819948	0.0004	13.17	0.074	0.167	-0.976
EYS	NM_001142800.2	Chr 6:66204970	c.334G>A	p.(Val112Ile)	Heterozygous	Substitution	Missense	rs112609906	0.009	9.676	-	0.191	-0.778
GRM6	NM_000843.4	Chr 5:178418555	c.727G>T	p.(Val243Phe)	Heterozygous	Substitution	Missense	rs17078894	0.005	25.8	-	0.604	0.573
IFT172	NM_015662.3	Chr 2:27668298	c.4933G>A	p.(Val1645Ile)	Heterozygous	Substitution	Missense	rs149117098	0.0007	21.5	0.006	0.031	-0.268
LAMA1	NM_005559.4	Chr 18:7023207	c.2657C>T	p.(Ala886Val)	Heterozygous	Substitution	Missense	rs144738522	0.0021	29	0.033	0.33	0.429
RP1L1	NM_178857.6	Chr 8:10468964	c.2644C>T	p.(Arg882Trp)	Heterozygous	Substitution	Missense	rs148936402	0.0027	7.132	-	0.113	-1.668
TMEM126A	NM_032273.4	Chr 11:85365174	c.154A>G	p.(Ser52Gly)	Heterozygous	Substitution	Missense	rs140047528	0.0029	25.7	0.044	0.227	0.693
WFS1	NM_006005.3	Chr 4:6302799	c.1277G>A	p.(Cys426Tyr)	Heterozygous	Substitution	Missense	rs35218685	0.0026	19.7	0.474	0.49	-0.097

[ddd] Genetic and functional annotation summary of variants in genes associated with vitelliform lesions identified from whole exome sequencing in the patient. Pathogenicity thresholds: >20 for CADD-PHRED (v1.6), >0.025 for MCAP, >0.5 for REVEL and >0.5 for Eigen. Abbreviations: MAF, minor allele frequency (https://gnomad.broadinstitute.org/); HGVS, human genome variation society; RefSNP, reference SNPs.

[eee] Genetic and functional annotation summary of rare variants (≤1% MAF) in genes associated with retinal disease identified from whole exome sequencing in the patient.

Etiological correlation with the *ABCA4*-associated optical gap phenotype

Following complete resolution of vitelliform fluid in the patient, the remaining hypo-reflective subretinal space in the fovea was consistent in appearance with foveal cavitation, occult maculopathy or optical gap lesions described in a subgroup of Stargardt/ABCA4 disease as well as other inherited and non-inherited retinal conditions (**Appendix Table E4**). While no other cases with vitelliform lesions were found in our initial internal screening of confirmed ABCA4 disease patients, we identified and further analyzed 18 additional cases exhibiting foveal optical gap lesions (**Appendix Table 1, Appendix Table E5**).

Appendix Table E4[fff]

Clinical Entity	Gene	OCT Description	Reference
Spinocerebellar Ataxia Type 7	*SCA7*	Altered foveal lamination, abnormal area of low reflectivity	Aleman et al.[307]
Stargardt disease	*ABCA4*	Foveal optical gap	Cella et al.[119]
Occult macular dystrophy	*RP1L1*	Well-defined disruption of the inner segment–outer segment (IS-OS) junction of photoreceptors and the Verhoeff membrane (cone outer segment tips)	Park et al.[308]
Achromatopsia	*CNGA3*	Foveal bubble, optical empty cavity	Thiadens et al.[309]
Achromatopsia	*CNGB3*		
Blue cone monochromacy	*OPN1LW*	Focal disruption of the inner segment ellipsoid (ISe) near the foveal center	Carroll et al.[310]
Blue cone monochromacy	*OPN1MW*		
Occult macular dystrophy	*ABCA4*	Foveal cavitation	Sisk et al.[311]
Cone dystrophy with supernormal rod electrogram	*KCVN2*	Focal disruption of reflective band at IS/OS junction	Sergouniotis et al.[312]
Occult macular dystrophy	*RP1L1*	Focal disruption of the IS-OS junction	Ahn et al.[313]
Spinocerebellar Ataxia Type 7	-	Focal loss of foveolar photoreceptors, disruption of the inner segment/outer segment junction, and preservation of the external limiting membrane	Watkins et al.[314]
Stargardt disease	*ABCA4*	Optical gap	Noupuu et al.[42]
Achromatopsia	*CNGA3*	Optically empty space	Greenberg et al.[315]
Achromatopsia	*CNGB3*		
Achromatopsia	-		
Achromatopsia	*CNGA3*	Foveal morphology: inner segment ellipsoid (ISe) absence/disruption/hyporeflective	Sundaram et al.[316]
Achromatopsia	*CNGB3*		
Achromatopsia	*GNAT*		
Achromatopsia	*PDEC*		
Achromatopsia	-		
Cone dystrophy	*GUCY2D*	Abrupt disruption of the inner–outer segment junction layer in the subfoveal photoreceptors	Mukherjee et al.[317]
Cone-rod dystrophy	*POC1B*	Fovea: changes at inner-segment ellipsoid zones, suggesting loss of junctions between inner and outer segments	Roosing et al.[318]
Tamoxifen-induced retinopathy	(non-hereditary)	Pseudocystic foveal cavitation	Doshi et al.[319]
Achromatopsia	*ATF6*	Optical gap	Kohl et al.[320]
Maculopathy	*PRPH2*	[*Structural irregularities especially in the photoreceptor layer, which is in the area of the outer segments of the cones in the foveola are particularly pronounced*][‡]	Maertz et al.[321]
Stargardt-like macular dystrophy	*ELOVL4*	Subretinal pocket	Palejwala et al.[322]
Cone dystrophy/macular dystrophy	*GUCA1A*	Absent ellipsoid zone band with thinning of the outer nuclear layer in the fovea	Manes et al.[323]
Tamoxifen-induced retinopathy	(non-hereditary)	Hyporeflective foveal cavitation and focal photoreceptor disruption	Lee et al.[324]

[fff] Differential descriptions of foveal optical gap lesions on optical coherence tomography (OCT) in inherited and non-inherited retinal degenerations. ‡Description translated from German.

Appendix Table E5[ggg]

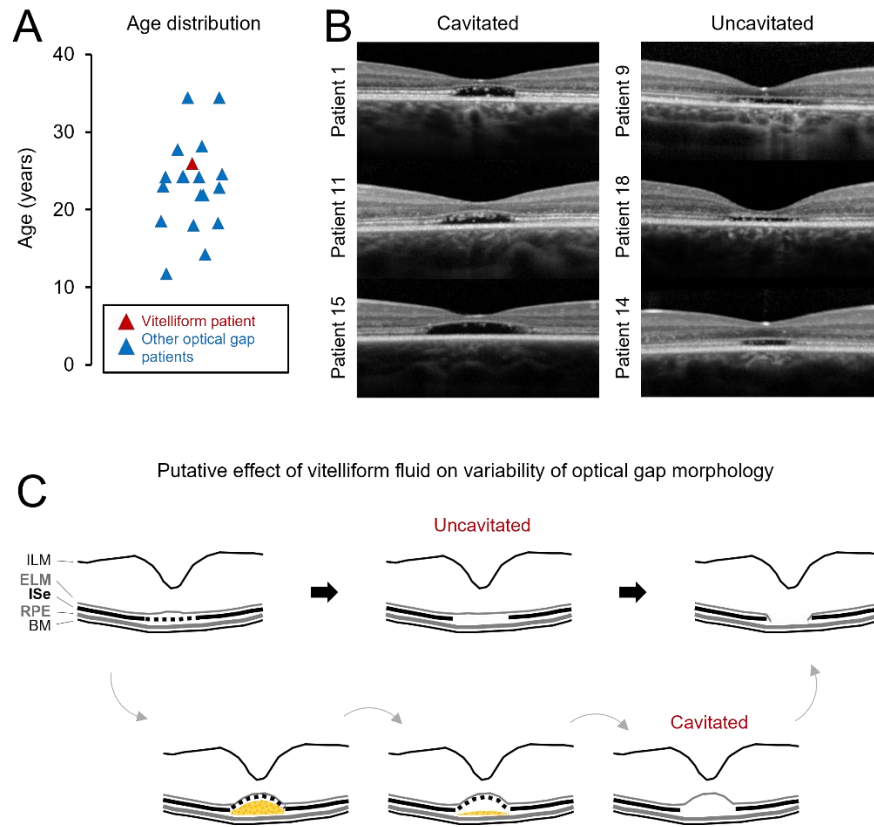| gDNA	cDNA (NM_000350.2)	Protein	Location	Allele count[‡]	Coding effect	dbSNP	MAF (gnomAD)	M-CAP	REVEL	Eigen	CADD-PHRED	DANN	Reported cases harboring variant
g.94577046_94577049dup	c.247_250dup	p.(Ser84Thrfs*15)	Exon 3	1	Frameshift	rs1005271380	-	-	-	-	-	-	Simonelli et al.[325] Maugeri et al.[326] Xu et al.[327] Fujinami et al.[328] Jiang et al.[329] Maggi et al.[330] Lee et al.[331]
g.94577010T>C	c.286A>G	p.(Asn96Asp)	Exon 3	1	Missense	rs61748529	0.000007967	0.283	0.768	0.923	23.3	0.997	Maia-Lopes et al.[332] Lambertus et al.[333] Nassisi et al.[334] Salles et al.[335] Fujinami et al.[336]
g.94528806A>G	c.1622T>C	p.(Leu541Pro)	Exon 12	2	Missense	rs61751392	0.0001627	0.58	0.983	0.999	26.8	0.999	Rozet et al.[337] Fishman et al.[338] Webster et al.[54] Briggs et al.[339] Gerth et al.[340] Wiszniewski et al.[137] Hargitai et al.[341] Cella et al.[119] Riveiro-Alvarez et al.[159] Fujinami et al.[342] Bertelsen et al.[43] Zhang et al.[190] Schulz et al.[53]
g.94520793A>T	c.2461T>A	p.(Trp821Arg)	Exon 16	1	Missense	rs61749433	0.00001062	0.512	0.87	0.962	28.6	0.997	Lewis et al.[186] Cideciyan et al.[343] Ścieżyńska et al.[344]
g.94509017T>C	c.3065A>G	p.(Glu1022Gly)	Exon 21	1	Missense	-	-	0.783	0.981	0.999	32	0.999	Zaneveld et al.[345] Jiang et al.[329] Oh et al.[346]
g.94496666G>A	c.4139C>T	p.(Pro1380Leu)	Exon 28	2	Missense	rs61750130	0.000233633	0.391	0.87	0.70	25.2	0.999	Lewis et al.[186] Briggs et al.[339] Shroyer et al.[193] Oh et al.[347] Fingert et al.[348] Hwang et al.[349] Chacón-Camacho et al.[350] Duncker et al.[23] Stone et al.[23] Lee et al.[331]
g.94496571G>A	c.4234C>T	p.(Gln1412*)	Exon 28	1	Nonsense	rs61750137	0.000007073	-	-	-	51	0.998	Maugeri et al.[326] Bertelsen et al.[43] Xu et al.[327] Kousal et al.[351] Fujinami et al.[328] Jiang et al.[329] Schulz et al.[53] Stone et al.[352] Paavo et al.[24]

[ggg] Summary of predicted pathogenicity and disease association of *ABCA4* alleles identified in the study cohort. ‡Allele counts exclude familial cases in the cohort. Pathogenicity thresholds: >20 for CADD-PHRED (v1.6), >0.025 for MCAP, >0.5 for REVEL and >0.5 for Eigen. Pathogenic scores are highlighted in pink. Abbreviations: MAF, minor allele frequency (https://gnomad.broadinstitute.org/); HGVS, human genome variation society; RefSNP, reference SNPs.

g.	c.	p.	Exon/Intron	N	Type	rs	Freq						References
g.94484082T>C	c.5196+1056A>G	p.(Met1733Valfs*2)	Intron 36	1	?	rs886044749	-	-	-	-	-	-	Stone et al.[352] Schulz et al.[53] Zernant et al.[47] Khan et al.[353] Del Pozo-Valero et al.[189]
g.94480241G>A	c.5318C>T	p.(Ala1773Val)	Exon 38	1	Missense	rs760549861	0.00007557	0.123	0.79	0.932	28.7	0.999	Chacón-Camacho et al.[350] Jiang et al.[329] Schulz et al.[53] Del Pozo-Valero et al.[189] Sun et al.[354]
g.94473807C>T	c.5882G>A	p.(Gly1961Glu)	Exon 42	11	Missense	rs1800553	0.004564289	-	0.76	0.35	28.4	0.998	Lewis et al.[186] Klevering et al.[355] Fishman et al.[338] Wiszniewski et al.[137] Kitiratschky et al.[356] Cideciyan et al.[357] Cella et al.[119] Burke et al.[55] Fujinami et al.[342] Lee et al.[37] Schulz et al.[53] Lee et al.[56] Birtel et al.[358] Nassisi et al.[334] Lee et al.[106] Lee et al.[284]
g.94466423A>G	c.6448T>C	p.(Cys2150Arg)	Exon 47	1	Missense	rs61750656	0.000003976	0.528	0.941	0.988	29.8	0.998	Briggs et al.[339] Simonelli et al.[359] Duncker et al.[360] Jiang et al.[329]
g.94466422C>T	c.6449G>A	p.(Cys2150Tyr)	Exon 47	1	Missense	rs61751384	2.82811E-05	0.525	0.92	0.80	32	0.998	Fishman et al.[338] Schulz et al.[53] Thiadens et al.[361] Sisk et al.[311] Oh et al.[347] Duncker et al.[23] Cideciyan et al.[192]
g.[94549781G>A; 94473807C>T]	c.[769-784C>T;5882G>A]	p.([Leu257Aspfs*3, Gly1961Glu])	Intron 6; Exon 42	2	(Complex allele)	-	-	-	-	-	-	-	Sangermano et al.[32] Runhart et al.[185] Lee et al.
g.[94528806A>G; 94508969G>A]	c.[1622T>C;3113C>T]	p.([Leu541Pro;Ala1038Val])	Exon12; Exon 21	3	(Complex allele)	-	-	-	-	-	-	-	Maugeri et al.[326] Rivera et al.[362] Fishman et al.[338] Wiszniewski et al.[137] Schulz et al.[53] Zhang et al.[190] Tracewska et al.[136] Zolnikova et al.[363] Lee et al.[284]
g.[94520784T>C; 94508969G>A]	c.[2470A>G;3113C>T]	p.([Ile824Val;Ala1038Val])	Exon16; Exon 21	1	(Complex allele)	-	-	-	-	-	-	-	Both variants are predicted pathogenic and have been reported as individual disease-causing alleles in patients: c.2470A>G[30] c.3113C>T[1,53,186,190,284]

Like the index patient, most of these additional cases harbored c.5882G>A (p.(Gly1961Glu)) and all together, the variant was highly enriched, occurring on one of the alleles in 17 out 19 (89.5%) cases (FET, p = 5.01E-08). The presenting age of the patient, 25-years-old, was also well within the range of the other optical gap cases from age 11.8 to 34.4 years (mean 23.3 years) (**Appendix Figure E6A**).

Hyper-reflective puncta traversing the ELM, similar to the ones observed in the patient after vitelliform resorption, were found in at least 2 other patients (P5 and P8). Variability in the size and morphology of the optical gap lesions was noted on SD-OCT. Larger cavitated lesions in which the ELM exhibits a rounded contour displacing the adjacent inner retinal layers like the one found in the patient were also observed in at least five other cases (P11, P13, P15, P17 and P18) (6/19, 31.5%) (**Appendix Figure E6B**). Significant changes in lesion morphology, including cavitation, were not noted in at least 4 patients (P6, P7, P10 and P12) who were followed longitudinally from early loss of the EZ band to incipient atrophy.

Appendix Figure E6[hhh]

A — Age distribution

B — Cavitated / Uncavitated (Patient 1, Patient 11, Patient 15 / Patient 9, Patient 18, Patient 14)

C — Putative effect of vitelliform fluid on variability of optical gap morphology

Discussion

The present case expands the clinical spectrum of both *ABCA4-* and vitelliform-associated

diseases. Vitelliform lesions are most prominently associated with the bestrophinopathies, which

are a group of retinal disorders caused by pathogenic variants in the *BEST1* gene.[152,364] The

presenting age and singular foveomacular lesion in each eye is most consistent with autosomal

dominant Best vitelliform macular dystrophy (VMD2, MIM# 153700). Aside from genetic

---

[hhh] Phenotypic analysis of optical gap lesions in the patient and 18 additional ABCA4 disease cases. (A) Distribution of the initial ages (years) at which each patient presented with optical gap lesions. (B) Structural variability of fully formed optical gap lesions in representative patients. (C) Proposed schematic diagram illustrating the effect of vitelliform fluid on the observed structural variability of optical gap lesions in ABCA4 disease. ILM, inner limiting membrane; ELM, external limiting membrane; ISe, inner segment ellipsoid; RPE, retinal pigment epithelium; BM, Bruch's membrane.

testing, which definitively confirmed ABCA4 disease, results of EOG testing provided one of the first of several findings that were inconsistent with VMD2. While the Arden ratio (1.65) was moderately subnormal for healthy eyes (≥1.8 to 2), it exceeds the ≤1.5 threshold and range typically reported in patients with VMD2.[365] The EOG has long been a reliable diagnostic test for bestrophinopathies although occasional findings of "normal" EOG results in VMD2 have been reported.[366,367] Electrophysiological testing in this patient was corroborative but ultimately did not contribute to diagnostic confirmation. Routine genetic testing is becoming increasingly available in clinics at diminishing costs. Following such recent trends, genetic testing may at some point, supersede the diagnostic utility of electrophysiological tests like the EOG for bestrophinopathies.

Abnormalities found outside of the foveomacular lesion in this patient which included photoreceptor layer thinning, ISe reflectivity changes on SD-OCT and the intensely bright region of NIR-AF around the perimeter of the lesion are also inconsistent with disease localization in the fundus of VMD2 patients. Levels of SW-AF (488-nm) and NIR-AF (787-nm) levels, as well as the thickness of retinal layers, were thoroughly investigated by two studies in large cohorts of VMD2.[368,369] Among other findings, the authors showed that similar underlying disease changes such as cellular layer thinning and lipofuscin accumulation do not occur in non-lesion areas. On the contrary, subclinical detection of early structural and functional deterioration in non-lesion areas of the fundus in ABCA4 disease patients is a very well-documented phenomenon.[370,371]

Vitelliform lesions have also been described in patients with mutations in other genes such as *PRPH2* (MIM# 179605),[372] *IMPG1* (MIM# 602870)*/IMPG2* (MIM# 607056),[373,374] *ABCC6* (MIM# 603234),[375] and non-inherited disorders such as age-related macular degeneration (AMD).[376] Interestingly, many of these conditions mimic one another and ABCA4 disease in

various ways. For instance, scattered hyperautofluorescent deposits found in each of these conditions sometimes resemble one of the various fleck patterns ABCA4 disease.[2] In *PRPH2*-associated pattern dystrophy and recessive bestrophinopathy (MIM# 611809), these deposits and other disease features have even been noted to exhibit the peripapillary sparing phenomenon.[301,377] While each of these conditions are genetically very distinct, the list of clinical commonalities between them, which now includes development of vitelliform fluid, strongly suggests that a mechanistic commonality may be shared between them, the implications of which should be further explored in treatment options and genetic modifier studies.

The development of vitelliform lesions in ABCA4 disease, while an evidently rare clinical observation, is actually congruous with its known disease mechanism. Dysfunction of the ABCA4 protein, which resides in photoreceptor outer segment disc membranes, leads to a perpetual build-up of visual cycle retinoids and subsequent synthesis of A2E in the disc lumen.[378] This substance is then deposited in RPE cells where it accumulates as lipofuscin, the presence of which is the principal cause cellular degeneration.[15,16] The precise composition of vitelliform material likely varies depending on the disease setting however clinicopathologic studies in patients with bestrophinopathy[379-382] and adult-onset vitelliform macular dystrophy[383-386] confirm lipofuscin and un-phagocytized outer segment debris to be a major component. Contrary to the milder, more macular-confined disease changes seen in conditions like VMD2 or AMD, the augmented accumulation of lipofuscin in ABCA4 disease is typically more widespread and manifests heterogeneous features such as the confluent formation of flecks across the posterior pole.[22] Interestingly, flecks undergo similar time-dependent changes as vitelliform lesions, whereby reflective subretinal deposits traversing photoreceptor-attributable bands exhibit a strong autofluorescence signal that extinguishes as the subretinal deposits resorb

overtime.[120,122] The precise etiological mechanism of fleck formation is largely unknown although several compelling explanations have been proposed.[122,387] Whether or not fleck and vitelliform formation are related pathological processes is indeed an intriguing question but beyond the scope of this study. What can be said, however, is that the aggressive disease changes characteristic of moderate to severe phenotypes of ABCA4 disease are likely not conducive to the formation of subfoveal vitelliform lesions typically seen in diseases like VMD2.

An exception to this may be found in patients with milder, slow-progressing phenotypes of ABCA4 disease—specifically individuals, like the patient, with the c.5882G>A (p.(Gly1961Glu))-associated optical gap phenotype. Both optical gap lesions and a mild disease setting imparted by c.5882G>A (p.(Gly1961Glu)) genotypes provide an ideal cellular milieu for the vitelliform lesion formation. Structurally, c.5882G>A (p.(Gly1961Glu)) optical gaps are stable, enclosed subretinal spaces within which vitelliform fluid can be deposited and retained for a period of time. This arrangement appears anatomically analogous to the occurrence of acquired vitelliform lesions within the space of subfoveal tractional neuroretinal detachments seen in patients with a epiretinal membranes.[388]

Our analysis of additional ABCA4 disease cases with optical gap lesions identified some features that may be indicative of prior vitelliform deposition. Intraretinal hyper-reflective foci found in the patient following the resorption of vitelliform fluid was also noted in several other cases. Similar deposits, histopathological shown to contain lipofuscin and melanolipofuscin granules, have been described in acquired vitelliform lesions.[384] The shape and size of optical gap lesions varied across patients with some, including the one found in the vitelliform patient, exhibiting a an enlarged, cavitated shape. It is unclear how and why this cavitation occurs in some lesions but not others. Based on our observations, the additional accumulation of vitelliform fluid in this

276

space may be a contributing factor (**Appendix Figure E6C**). It should also be noted however that similar cavitated optical lesions are also found in other retinal disorders (e.g., cone/cone-rod dystrophies; see **Appendix Table E4**) which contradicts this hypothesis. Understanding the mechanistic nuanced effect of vitelliform fluid in the retina is beyond the limitations of this study and future studies in large patient cohorts will be necessary to address these questions.

In summary, we present an in-depth analysis of confirmed case of ABCA4 disease who, to the best of our knowledge, has the first documented foveomacular vitelliform lesion. This clinical manifestation is phenotypically similar to those seen in other associated diseases and its resolution, which we documented over a 35-month period progresses through characteristic vitelliform stages. Aside from genetic testing, ABCA4-associated vitelliform lesions are likely associated with the optical gap subphenotypes which we hypothesize is the reason their infrequent documentation. Lastly, ABCA4-associated vitelliform lesions can be distinguished from those of other conditions like autosomal dominant bestrophinopathy by EOG testing (Arden ratio >1.55), as well as the presence of subclinical abnormalities in fundus areas eccentric to the lesion.

Materials & Methods

All study subjects provided written and signed informed consent to participate in this study. All study procedures are described in protocol #AAAI9906 which was approved by the Institutional Review Board at Columbia University Medical Center. The study adhered to tenets set out in the Declaration of Helsinki. The patient underwent a complete ophthalmic examination that included slit-lamp and dilated fundus exam, best corrected visual acuity (BCVA; Snellen), multi-modal imaging and electrophysiological testing. Electrooculogram (EOG) and full-field electroretinogram (ffERG) recordings were conducted using the Espion Visual

Electrophysiology System (Diagnosys LLC, Littleton, MA, USA) according to International Society for Clinical Electrophysiology of Vision (ISCEV) standards.[108,389]

Imaging across all modalities were conducted following maximum pupil dilation (>7mm) with tropicamide (1%) and phenylephrine hydrochloride (2.5%). Short wavelength (488-nm) and near infrared-autofluorescence (787-nm) images and spectral domain-optical coherence tomography (SD-OCT) scans were acquired using the Spectralis HRA+OCT scanning laser ophthalmoscope (SLO) (Heidelberg Engineering, Heidelberg, Germany). Ultra-widefield (optomap) color and 532-nm autofluorescence images were acquired with an Optos 200 Tx SLO (Optos PLC, Dunfermline, United Kingdom). Optical coherence tomography angiography (OCTA) images were acquired using a Zeiss AngioPlex Cirrus HD-OCT 5000 (Zeiss Meditec Inc, Dublin, California, USA). *En face* 3mm x 3mm images of the choriocapillaris (CC) layer image were constructed from raster scans that were automatically segmented and manually adjusted when necessary.

The thickness of the total retina, outer nuclear layer + outer plexiform layer (ONL+), outer segment+ (OS+) and RPE layers were manually measured using the caliper tool in the HEYEX software (Heidelberg Engineering) in both eyes each of family member and compared to the mean ± 2 standard deviations (SD) of 25 eyes of 25 healthy subjects (mean age, 47.5 years; range, 13.5–84.4 years). TREC+ is defined as the distance between Bruch's membrane (BM) to the border of the outer plexiform/ inner nuclear layer; OS+ is defined as the distance between BM and the ellipsoid zone. Measurements were carried out on horizontal SD-OCT scans (high-resolution, 9 mm) at 0.5 mm intervals up to 3 mm along the nasal and temporal axes (17 positions total). The horizontal diameter of the hypoautofluorescent lesion in each eye was measured, averaged and mapped on each thickness plot. *En face* images of the inner segment

278

ellipsoid (ISe) slab were generated using the transverse analysis module of the HEYEX software. A dense raster scan of 97 high resolution B-scans were acquired in each eye and the contour of the ISe band was manually segmented for each individual scan.

Genomic DNA was extracted from peripheral blood lymphocytes from the patient and the whole exome sequencing was performed by Psomagen (Rockville, MD) using the SureSelect Human All Exon V8 (Agilent, Santa Clara, CA). Sequence reads were aligned to the hg19 reference genome with BWA, and processed with GATK according to the best practices recommendations. After variant calling, we narrowed our analyses to variants in genes previously associated with retinal disease (https://sph.uth.edu/retnet/) at a minor allele frequency (MAF) ≤ 0.2 based on the gnomAD database (assessed December 2021). We then performed functional annotation on the 221 resulting variants with ANNOVAR (https://annovar.openbioinformatics.org/en/latest/) using pathogenicity scores from the dbnsfp 4.2a dataset (https://sites.google.com/site/jpopgen/dbNSFP). After filtering and evaluating variants from genes associated with vitelliform lesions (**Appendix Table 1**), we further restricted the MAF filter to ≤0.01 and evaluated the remaining 15 variants for causality (**Appendix Table 2**). The phase of the *ABCA4* variants identified in each patient were either confirmed by parental screening (**Appendix Figure E7**) or imputed based on their co-occurrence patterns in the general population (**Appendix Figure E8**, https://gnomad.broadinstitute.org/variant-cooccurrence).

Appendix Figure E7[iii]

Phase determination of *ABCA4* variants by familial segregation

---
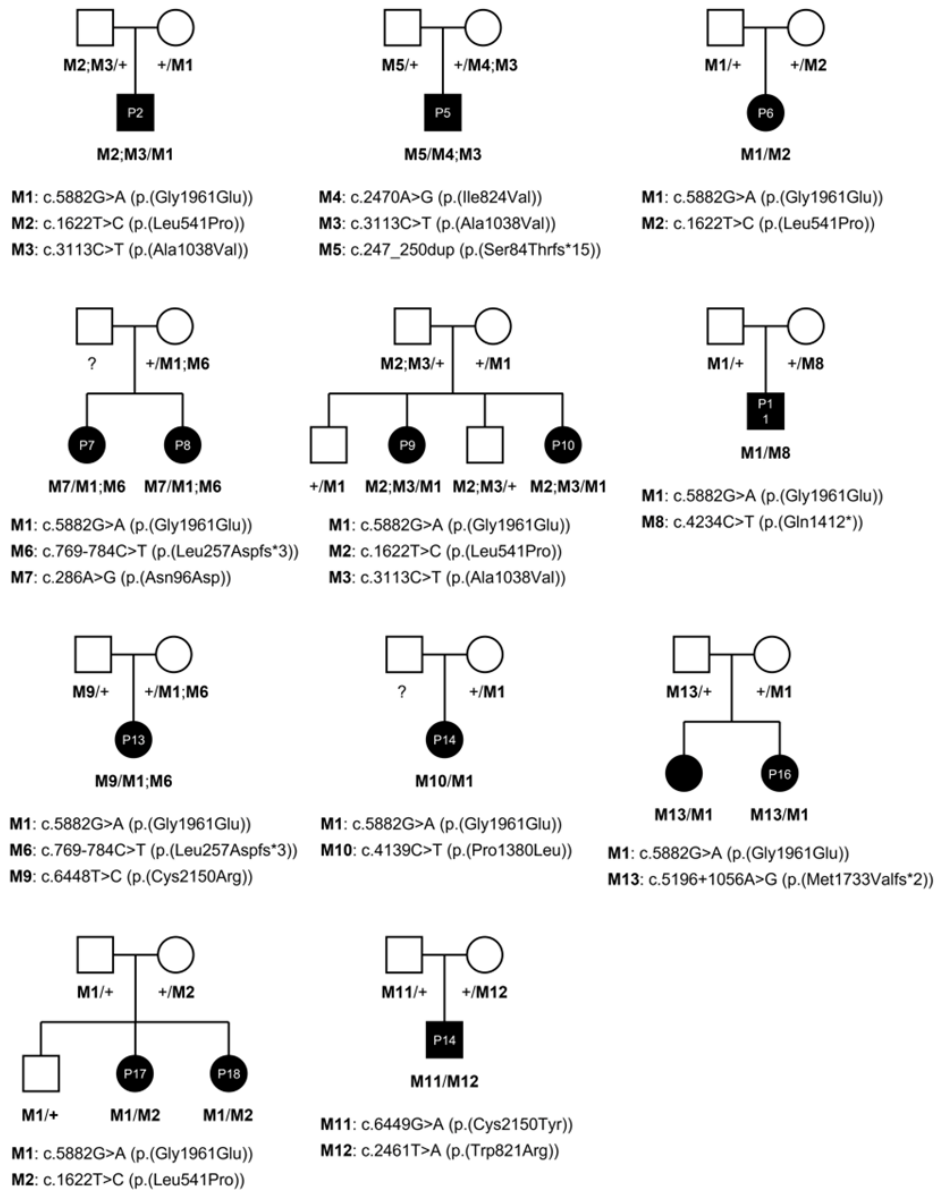
[iii] Phase determination of *ABCA4* variants by familial segregation. DNA samples collected from parents and siblings of patients were screened. The segregation of each *ABCA4* variant on the maternal and paternal allele is shown for 14 patients across 11 familial pedigrees.

Appendix Figure E8[iii]
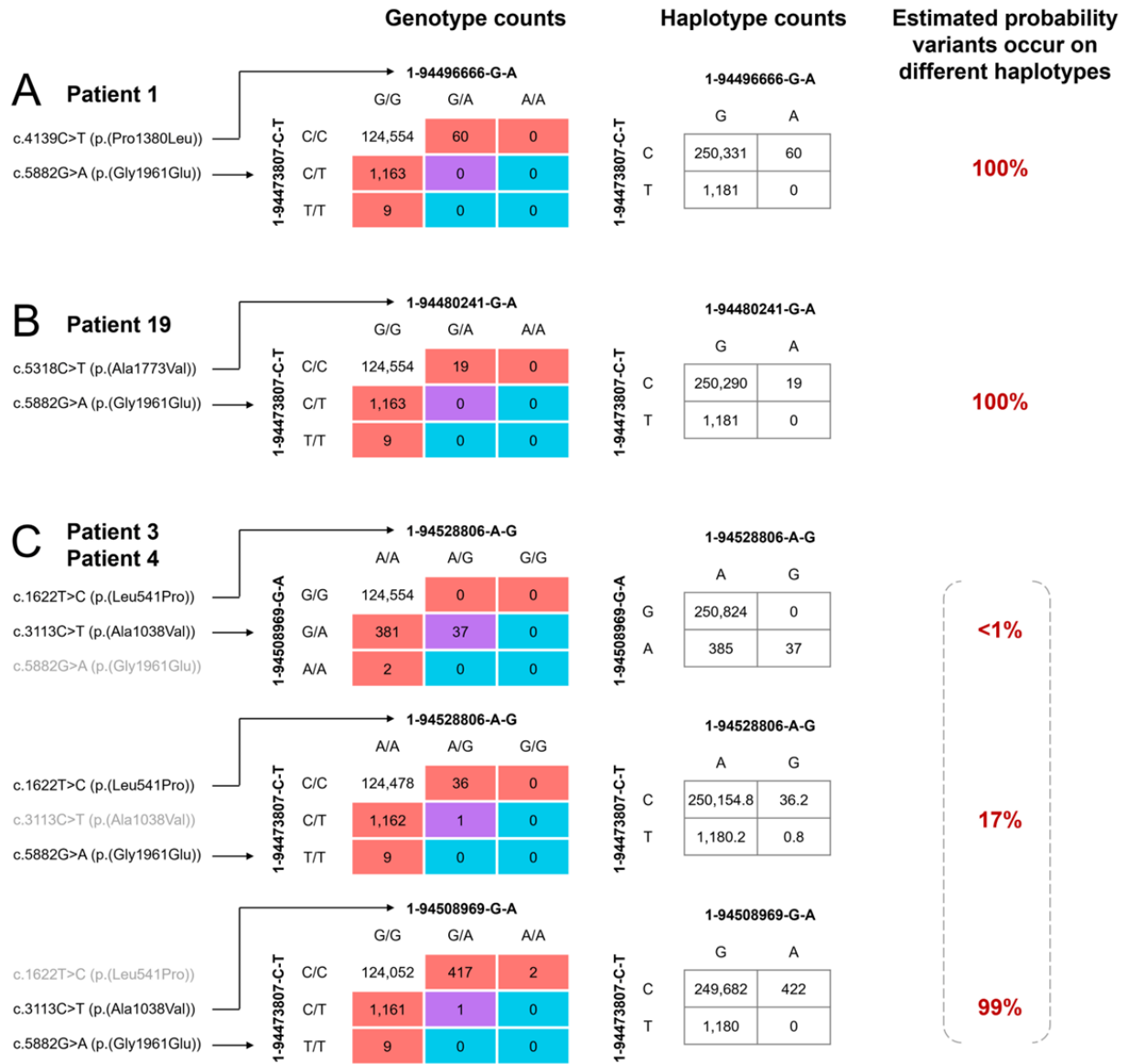
[iii] Analysis of co-occurrence patterns in the general population of *ABCA4* variants found in study patients in whom familial screening for phase determination was not available. Analysis of co-occurrence patterns in the general population of *ABCA4* variants found in study patients in whom familial screening for phase determination was not available. Genotype counts are presented in 3x3 tables listing the number of individuals (exomes) in the gnomAD database (https://gnomad.broadinstitute.org/gene/ENSG00000198691?dataset=gnomad_r2_1, accessed March 2022) with these genotype combinations. Red squares are individuals with variants falling on different haplotypes; blue squares are individuals with variants falling on the same haplotype; purple squares are indeterminant. Haplotype counts are estimated based on an expectation-maximization algorithm (https://gnomad.broadinstitute.org/news/2021-07-variant-co-occurrence-phasing-information-in-gnomad/). Percentages are based on expectation-maximization probability distributions.

Thickness plots and standard deviation intervals of retinal layers in patients and healthy controls were generated using the statistical computing software R version 4.0.4 (https://www.r-project.org/). Thickness measurements were performed by two independent graders. Intraclass correlation coefficients (ICC), calculated using the irrICC package (https:// cran.r-project.org), showed excellent intergrader agreement for the patient (ICC>0.9978) and controls (ICC>0.9996). A Fisher's Exact Test (FET) for Count Data was used to assess the enrichment of the c.5882G>A (p.(Gly1961Glu)) allele in patients with and without the optical gap phenotype.

Appendix F: Identification of a pathogenic 1,500 bp *Alu* insertion in *ABCA4*[kkk]

The patient was a woman who presented in her early 20's with disease characteristics in the macula of both eyes consistent with recessive Stargardt disease (STGD1, MIM #248200). No other systemic issues or family history were reported.

Direct sequencing of the *ABCA4* locus revealed a single inherited pathogenic variant, c.3322C>T, p.(Arg1108Cys) (CADD-PHRED = 32) (rs61750120) (**Supplemental Table 6**). Targeted long-read sequencing using adaptive sampling on the Oxford Nanopore platform[36] was performed in this patient in collaboration with investigators from the University of Washington and Seattle

Children's Research Institute.[lll] A ~1,500 bp insertion was captured in two long reads (**Appendix Figure F1**) with the following sequences:

> ae1e45de-1f16-4804-886a-da01aa0b646c[mmm]

AAACGTGCCTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCATTCAAACAAATA TTTATTAAACACCTTCTTTCTTTAAATTCCAAATCTCTCATTGTTTGAGGCAGTCCAA ATGC   TCAGCTGTCCTAAGGCTGTTCTACTCTCCATACAAAGCTTTAGCTTCTCCAG CATATCCCCATTACCACATCTCATTTTCCATCCAGGATTCTTCCTATTTCCATATTTTT ACCTATTTTACTGAAATGATCTCTCCTGCTCTTCAATTAATTCAAATCTGAATCATTT CTTCAAAGCCTGCTGAAGTCACCTCGCCCCCGGACCTC   TGGCCTAATTGCCTTCAT

[kkk] Text and figures from this chapter have been recently published, in part, in Miller et al. (2021) *American Journal of Human Genetics*. 2021 Aug 5;108(8):1436-1449.
[lll] Danny E. Miller, M.D., Ph.D., Timothy Cherry, Ph.D. and Evan Eichler, Ph.D.
[mmm] Sequence in green corresponds to Repeat L2a, family L2 (LINE), sequence in blue corresponds to Repeat L2c, family L2 (LINE) (131/148 bp (~89%) sequence identity) and sequence in red corresponds to Repeat *Alu*Jr, family Alu (SINE) (265/298 bp (~89%) sequence identity).

GCCATTCTGTGACATTATAATATTGGAGCTTAAGTTGCATTCCACACAAACACTTGA

TATTTCTTTGATTTGAATAGTTCTATTTTTATTATTGTTAATAGAGACAGGACCTTCCT

CTGTTTGGAGTGCAGGACACAACCATAGCTCACTGCAGCCTTGCAAACCTCTGGGCT

CATGTCCCTCCCTTCT   TCAGTCTCTCAAGTATCTAGTACTACAAATATATGCCACC

ACACCCAGATACTTTTATTGTTTGTAGAGATAGGGTCTCACTTGGATGTTCAGTTAAT

CTCGAATTTACTTCATAATCCTCTCTCATCTCAGCCTCCCAAAGCACTGAGACTATAG

GCATCAGCCATCAAGTGCCTGGCCTTGTTCTCTTATTTTTGTTTTTTTTTT   TTTTTTT

TTTTTTTTTTATTATACTCTAGTTTTAGTACATGTGCACACGTGTGTAGTTAGTTACA

TATGTATACATGTGCTATGCTGGTGCGCTGCACCTCCACTAATGTAATATCTATGACT

AGGGCATATCTTCCCAATGCTATTTCTCTCTCGACCCCACCACAGTCCCCAGAGTGT

GATATTCCCTTCCTGTGTCCATGTGA   TCTCTCATTGTTCAATTCCCACCCATGAGCA

GAGAATATGCGGTGTTTGGTTTTTTGTTCTTGCGATAGTTTACTGAGAATGATGGTTT

CAATTTCATCCATGTAAGTCCCTACAAAGGATATGAACTCATCATTTATGGCTGCAT

ATATTCCATGTGTATATGTGCCACACGCTTAATCCAGTCTATCATTGTTGGACATTTA

GTT   ATCTGCTCTTTAAAGAGTCATAATAAAATGGCAATTGATGACGCCAAAGGTC

TCATAGCAATGATTTATAATTACTGGTAAGGCACCTGTGCAATGGGTATTTGCTGGG

TCAAATGGTATTTCTAGCCCTAGATCCCTGATGAATTCACATCTCCACAATGGTTGA

ACTAGTTTACAGTCCCACCAACAGTGCATGTGGCAGCAGTC   TTTATTTCTCCGCAT

CCTCTCCAGCACCTGTTTCTGACTTTTAATGATTGCCATTTAACTGGAGGGAGATGAT

ATCTCATAGTGGGTTTTGATTTGCATTTACGCCAGTGATGATGAGCATTACTTTATGC

TTTGGCTGCATA

>b56e36f4-2d00-4bc6-9656-598bb5be6f19[nnn]

GTTTGTTTTTCATTCAACAAATATTGTTGTTTACCTTCTTTCTTTAAATTCCAAATTCC

TCATTGAGGGCAGTCCCATACTCAGCTGTCCAAGGCTGTTCTCTACAAGCTTAGTTTC

TCCAGCATCTAGCCACCAC   ATCTCTGTTTTCTATCAGGATTCTTCCTATTTCCATAT

TTTACTCATTTGCTGAAATGATCTTCCTTTCCTCCACCAATTCAAATCTGAATCTATT

CCCAAAGCTAGCTGAAGTCACTCGCCCCCGGACCTGTCCTAATTGCCTTCATGCCAT

TCTGTGTGACATTGTTCTCTAATATTGAGCTTAAGTTGCATTCCCACACAAACA   CT

TGATATTTCTTGATTTGAATAGTTCTATTTTAATGTGTTAGAGACAGGACCTTCCTCT

GTTTGGAGTGCAGTGACACAACCATAGCTCACTGCAGCCTTCAAGCTCCTATCAAGG

GGTCCTCAACTTCTTCAGTCTCTCAAGTGTCTAGTACTACAAATATATGCCACCACAC

CCAGATAATTTTTATTGTTTTGTAGAGATAGG   GTCTCACGGTGGCTGTTCAGGCTG

GTCTCGAACTACAAGCTGGGCAATCCTCTCATCTCAGCCTCAGCACTTTGAGACTAT

AGGCATCGCGCCATCATGCCTGGCCTGTTCTCTTATTTTTGTTTAAATTTTTTTTTGTG

CTCTAAGTTTTGGGGTACATGTGCACATTGTGCAGGTTAGTTACATATGTATATGTAC

TGTGCTGGT   CTTTGCACCCACTAATGTGTCATCTAGCATTGGGCATATCTCCCAAT

GCTGTCTCCCTCCCCTCCCGACCTTTGAGTCCCCAGGGTGATATTCCCCTTCCCTGTG

TCCATGTGATCTCATTGTTCAATTCCCACCTATGGTGAGAGGTGTTTGGTTTTATTTC

TTGATCAGTTTCACACTGAGAATGATGGTTTCCAATTTCATCCA   TGTCCCTACAAA

GGATATGAACTCATCTCATTTTTATATGGCTGCATAGTATTCCATGGTGTATATTGCC

ACATTTTCTTAATCAATCTATCATTGTTGGACATTTGGGTTGGTTCCAAGTCTTACCT

ATTGAATAGTGCCGCAATAAACATGGCATGCAGCATAGCATGAGTAGCATGATTTAT

ACTCATTTGGGTGTATCTGGTA   ATGGGATAGCTGGGTCAAATGTTCATGGTTCTAG

ATCCTGAGGAATCGCCACACTGACTTCCACAATGGTTTGAACTAGTTTACAGTCCAC

CAACAGTGTAAAAGTGTTCCTATTTCTCCGCATCCTCTCAACTTCCTGTTGTTCCCTG

ACTTTTTGTGATTGCCATTCTAACTGGTGTGAGATGATATCTACATGTGTTTGATTTG

CATTTCTCTGATGGCCAGTGAGCATTTCTTCATGTGTTTTTGGCTGCATAACGTGCG

CCT

Appendix Figure F1[ooo]
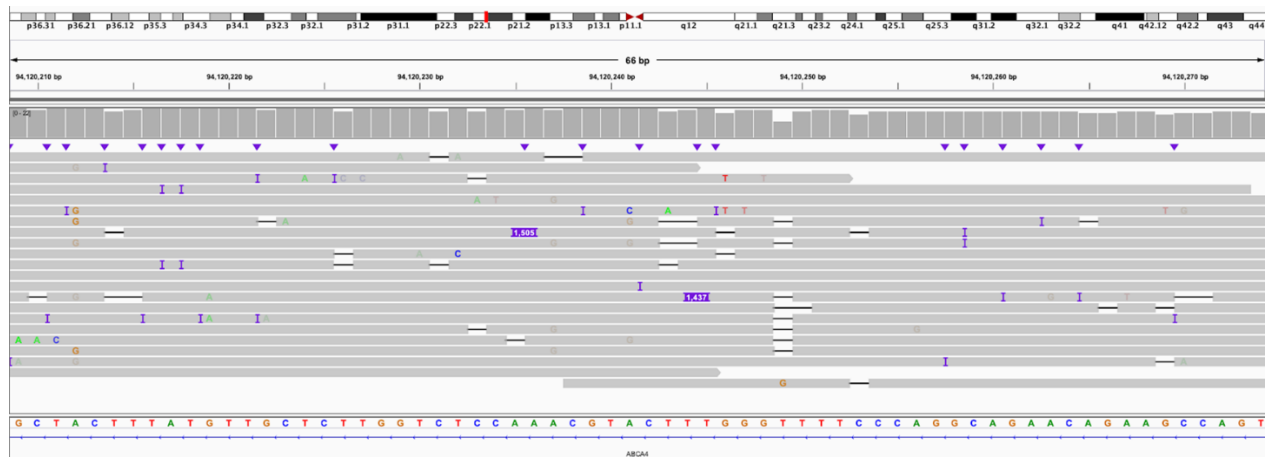
Further analyses revealed this ~1,500 bp variant to be a composite retrotransposable element

insertion (**Appendix Figure F1**) consisting of *Alu*J short interspersed nuclear element (SINE)

and partial L2a, L2c, L2d2, and L1HS long interspersed nuclear elements (LINEs) mapping to

Intron 1 of *ABCA4*.

While the pathogenicity of this insertion is of unknown significance (PM3, PP3 and PP4), *in

silico* analyses predicts aberrant splicing between Exons 1 and 2. The SV is ultra-rare, being

absent from all accessible population genetic databases (gnomAD, BRAVO). Specifically,

---

[ooo] Integrated Genomics Viewer (IGV) display of long reads showing two reads (purple) that include the ~1,500 bp insertion in intron 1 of ABCA4. Several reads on either side of the region include the insertion and are soft clipped. Genomic region shown spans chr1:94,120,209 - 94,120,274.

SpliceAI predicts a splice-altering consequence of the pre-mRNA by the introduction of a *de novo* donor site by the insertion event (Δ score = 33%, high recall). The ~1,500 bp insertion sequence itself contains two strong alternative splice donors and six alternative splice acceptors (**Appendix Figure F2**). The relative strength of these alternative donors and acceptors indicates a high probability of competition with canonical sites. One alternative donor site in particular exhibits a stronger splice signal than the canonical donor site at the Exon 1–Intron 1 boundary is the most probable site of alternative splicing leading to the inclusion of the 5' portion of the intron in the final mRNA transcript.

Appendix Figure F2[ppp]

Splice donor	SpliceSiteFinder-like	MaxEntScan	NNSPLICE	GeneSplicer	Human Splice Finder
Canonical	81.70	8.90	1.00	5.20	84.70
Alternative	90.10	8.00	0.90	-	95.00
Alternative	81.50	7.20	0.80	0.60	90.80

Splice acceptor	SpliceSiteFinder-like	MaxEntScan	NNSPLICE	GeneSplicer	Human Splice Finder
Canonical	95.25	10.87	0.99	12.03	92.11
Alternative	93.50	10.00	1.00	3.20	90.60
Alternative	88.54	7.30	1.00	1.50	79.83
Alternative	77.80	6.10	0.60	2.30	78.50
Alternative	83.20	5.90	0.70	-	85.80
Alternative	71.70	5.70	-	1.60	79.50
Alternative	74.60	6.90	0.60	-	78.30

[ppp] Length of color bars (blue = donor, green = acceptor) are proportioned to the respective scales of each algorithm: SpliceSiteFinder-Like (0-100), MaxEntScan (0-16), NNSPLICE (0-1), Gene Splicer (0-15), Human Splice Finder v.3.1 (0-100).

We statistically determined the phase of these two variants by showing that a marker SNP, rs2184339 (T>C), on the same allele as the Alu insertion is in linkage equilibrium with the c.3322C>T, p.(Arg1108Cys) (rs61750120) using frequency data from 2,504 unrelated individuals in the 1000 Genomes Project. Observed haplotype frequencies of SNPs rs61750120 and rs2184339:

	C	T
A	0	3
G	1003	4002

Individual allele frequencies for rs61750120 and rs2184339 are: $P(A)$ = 3 out of 5,008 total alleles = 0.001; $P(C)$ = 1,003 out of 5,008 total alleles = 0.2. No individuals (haplotypes) were found harboring both the A and C alleles: $P(AC)$ = 0 out of 2,504 = 0.

$$D = P(AC) - P(A)P(C)$$

$$D = 0 - 0.001 \times 0.2$$

$$D = -0.0002$$

A coefficient of linkage disequilibrium (D) of -0.0002 indicates a weak or no linkage disequilibrium between the two alleles being studied. The D statistic ranges from -1 to 1, where a value of 0 indicates that the two alleles are in linkage equilibrium, meaning that their inheritance is independent of each other.[qqq] It's important to note that the interpretation of the D statistic depends on the population studied and the frequency of the alleles being analyzed.[rrr]

[qqq] A negative value of D indicates that the two alleles are less likely to be inherited together than expected by chance, while a positive value of D indicates that they are more likely to be inherited together than expected by chance. However, a D value close to zero, such as -0.0002, suggests that the two alleles are not strongly associated with each other and are likely to be inherited independently.
[rrr] The coefficient of linkage disequilibrium D is not always a convenient measure of linkage disequilibrium because its range of possible values depends on the frequencies of the alleles it refers to. This makes it difficult to compare the level of linkage disequilibrium between different pairs of alleles.

Normalizing D by dividing it by the theoretical maximum difference between the observed and expected haplotype frequencies as follows:

$$D' = \frac{D}{D_{max}}$$

where,

$$D_{max} = \begin{cases} max\{-P(A)P(C), -(1 - P(A))(1 - P(C))\} & when \ D < 0 \\ min\{P(A)(1 - P(C)), (1 - P(A))P(C)\} & when \ D > 0 \end{cases}$$

Given that $D < 0$, we set D_{max} equal to $max\{-P(A)P(C), -(1- P(A))(1- P(A))\}$ or $max\{-P(A)P(C), -P(G)P(T)\}$, where $P(G)$ and $P(T)$ are the alternative (major) allele frequencies of rs61750120 and rs2184339, respectively.

$$max\{-0.0002, -0.7991\}$$

$$D' = \frac{-0.0002}{-0.0002}$$

$$D' = 1$$

The value of D depends on allele frequences making it difficult to interpret. An alternative to D' is the correlation coefficient between pairs of loci (R), usually expressed as its square, R^2:

$$R^2 = \frac{D^2}{p_A(1 - p_A)p_C(1 - p_C)}$$

$$R^2 = \frac{(-0.0002)^2}{0.001(1 - 0.001)0.2(1 - 0.2)}$$

$$R^2 = 0.00025025$$

We also run a contingency table test such as X^2 where the test statistic is computed based on the observed and expected haplotype frequencies as if there is no linkage disequilibrium (H_0):

$$X^2 = \sum \frac{N(\alpha \times \delta - \gamma \times \beta)^2}{(A)(G)(C)(T)}$$

where,

	C	T
A	$\alpha = 0$	$\beta = 3$
G	$\gamma = 1003$	$\delta = 4002$

$$X^2 = \sum \frac{N_{alleles}(\alpha \times \delta - \gamma \times \beta)^2}{(A)(G)(C)(T)}$$

$$X^2 = \sum \frac{5{,}004(0 \times 4002 - 1003 \times 3)^2}{(0 + 3)(1003 + 4002)(0 + 1003)(3 + 4002)}$$

$$X^2 = 0.751761198$$

This X^2 statistic gives a p-value = 0.3859 ($df = 1$), indicating no linkage disequilibrium between rs61750120 and rs2184339 (i.e., p.(Arg1108Cys) and Alu insertion marker are estimated to be on separate alleles).

For calculations with very small sample sizes (e.g., number of cases is no greater than 15 (and consequently $N_{alleles} < 30$), one should apply Yates's correction for continuity to the X^2 test, where 0.5 is subtracted from the numerical difference between the observed and expected frequencies:[sss]

$$X^2 = \sum \frac{|(N_{alleles}(\alpha \times \delta - \gamma \times \beta)| - 0.5)^2}{(A)(G)(C)(T)}$$

[sss] A Fisher's exact test is also appropriate for such situations.

Appendix G: Statistical Determination of Disease Severity in ABCA4 disease

To determine the principal clinical feature (covariates) that distinguish prognostic outcome categories in **Chapter 5**, we performed a Cox proportional hazard model on a cohort of patients (n=149) in whom the caontinuous and categorical variables were available: sex, age of onset (AO, years), choroidal thickness, fleck distribution, and DDAF presence. We found that both AO and fleck distribution were the largest predictors of disease severity:

covariates	coef	exp(coef)	se(coef)	z	Pr(>|z|)	
sex	0.049535	1.050782	0.386003	0.128	0.89789	
ao	-0.074741	0.927984	0.017439	-4.286	1.82E-05	***
choroid	0.002021	1.002023	0.001907	1.06	0.289364	
fleck	0.544367	1.723518	0.143544	3.792	0.000149	***
DDAF	-0.043937	0.957015	0.017384	-2.527	0.01149	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

We then assessed the time-dependent effect of these covariates by fitting this data to Aalen's additive regression model.[ttt] We found that fleck distribution appears to have an increasing effect on the hazard ratio over time, (**Appendix Figure G1**, green curve). In other words, the importance of fleck distribution proportionally increases as a determinant of severity as we progress to more advanced stages of the disease, while AO (**Appendix Figure G1**, red curve) and the presence of a DDAF lesion(s) (**Appendix Figure G1**, turquoise cuve), although initially significant, become less important.

[ttt] Assumption is that the cumulative hazard H(t) for an individual can be expressed as a(t) + X B(t), where a(t) is a time-dependent intercept term, X is the vector of covariates for the subject (possibly time-dependent), and B(t) is a time-dependent matrix of coefficients.
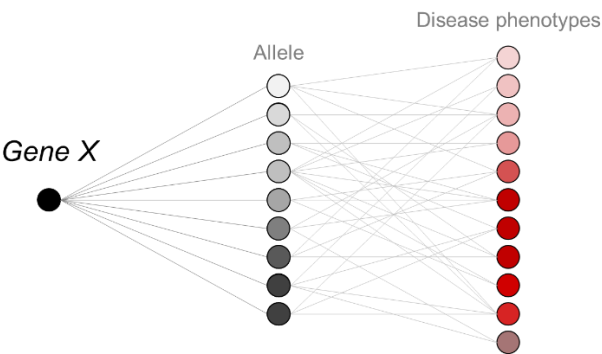
Appendix Figure G1[uuu]
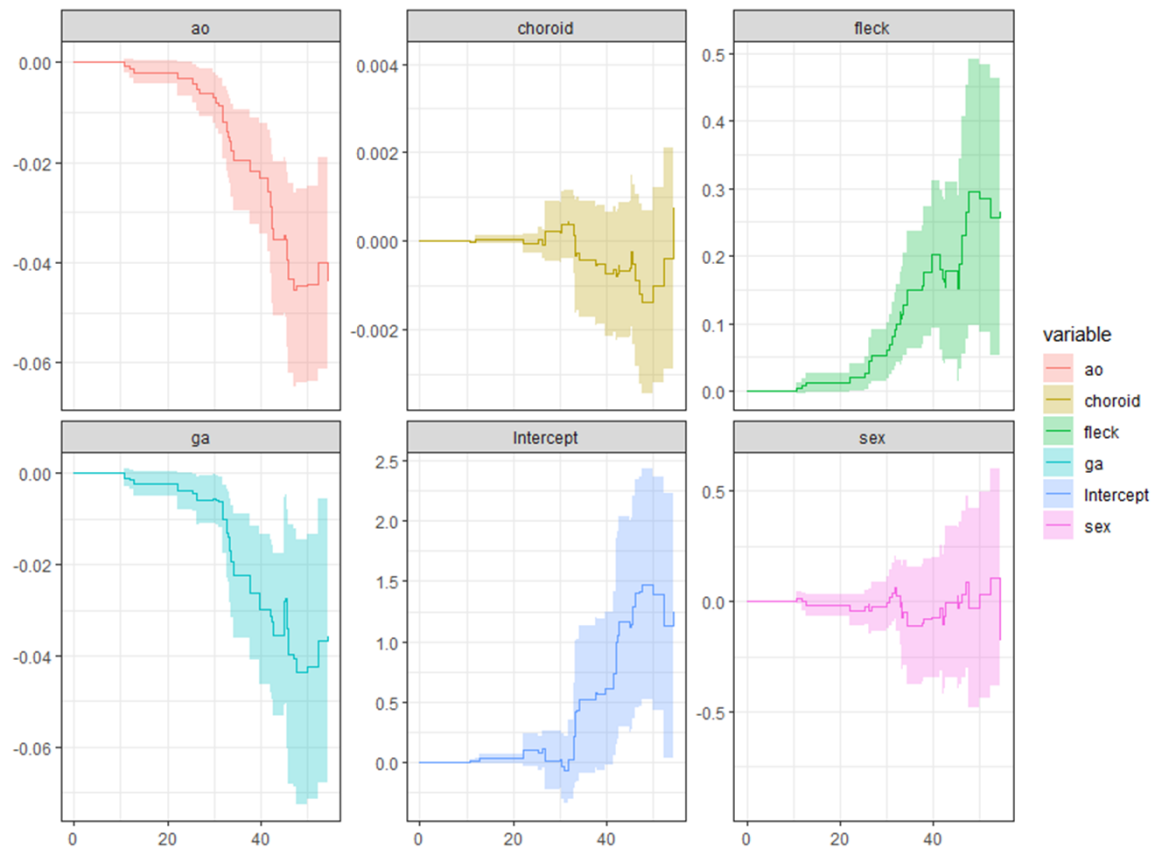
[uuu] Estimated cumulative regression functions based on Aalen's additive model showing the time-dependent effects of covariates age of onset (ao), choroidal thickness (choroid), distribution of flecks in the fundus (fleck), DDAF (ga) and sex on the hazard ratio in 149 patients with ABCA4 disease. Shaded regions around curves are 95% pointwise confidence intervals.

Appendix H: Definition and graphical representation of monogenic disease architecture

Common linear representation of monogenic disease architecture for a given Gene X:

Realistic representation for autosomal dominant diseases:

Realistic representation for autosomal dominant diseases: