JCI insight

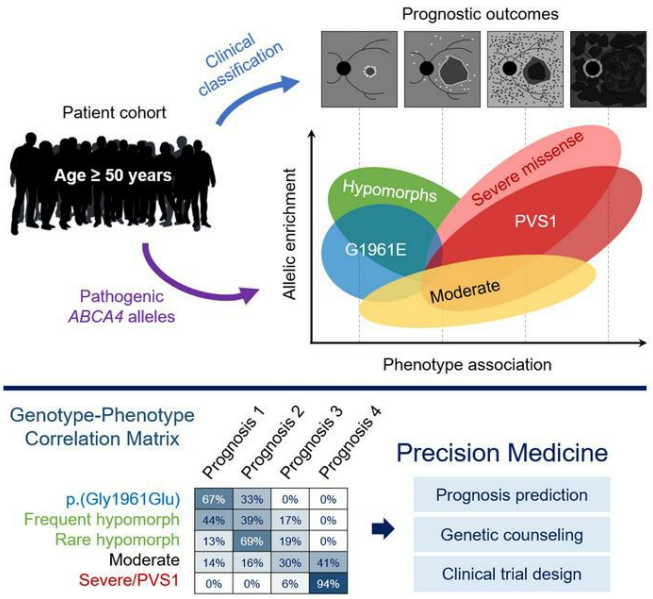
A genotype-phenotype correlation matrix for Stargardt/ABCA4 disease based on long-term prognostic outcomes

Winston Lee, … , Stephen H. Tsang, Rando Allikmets

JCI Insight. 2021. https://doi.org/10.1172/jci.insight.156154.

Clinical Medicine | In-Press Preview | Genetics | Ophthalmology

Graphical abstract



A genotype-phenotype correlation matrix for Stargardt/ABCA4 disease based on long-term prognostic outcomes

Winston Lee,[1,2] Jana Zernant,[2] Pei-Yin Su,[2] Takayuki Nagasaki,[2] Stephen H. Tsang,[2,3] Rando Allikmets[2,3]*

Affiliations:
[1]Deparment of Genetics & Development, Columbia University Irving Medical Center, New York, NY, USA
[2]Department of Ophthalmology, Columbia University Irving Medical Center, New York, NY, USA
[3]Department of Pathology & Cell Biology, Columbia University Irving Medical Center, New York, NY, USA

Keywords: ABCA4, Stargardt disease, retinal degeneration, genotype, phenotype, correlation, prognosis, hypomorph

*__Correspondence__:
Rando Allikmets, Ph.D.
Department of Ophthalmology,
Columbia University Irving Medical Center
Eye Research Annex Rm 202
635 W 165th St.
New York, NY 10032
+1 (212) 305-8989
Email: rla22@cumc.columbia.edu

Summary: We constructed a genotype-phenotype correlation matrix that provides quantifiable probabilities of long-term disease outcomes associated with specific ABCA4 genotypes from a large, age-restricted patient cohort

Role of funding source: The funding organizations for this study had no role in its overall design; acquisition and interpretation of data; preparation and revision of the manuscript.

Abstract

Background: >1,500 variants in the *ABCA4* locus underlie a heterogeneous spectrum of retinal disorders ranging from aggressive childhood-onset chorioretinopathy to milder, late-onset macular disease. Genotype-phenotype correlation studies have been limited in clinical applicability as patient cohorts are typically small and seldom capture the full natural history of individual genotypes. To overcome these limitations, we constructed a genotype-phenotype correlation matrix that provides quantifiable probabilities of long-term disease outcomes associated with specific ABCA4 genotypes from a large, age-restricted patient cohort.

Methods: The study included 112 unrelated patients ≥50 years of age in whom 2 pathogenic variants were identified after sequencing of the *ABCA4* locus. Clinical characterization was performed using the results of best-corrected visual acuity, retinal imaging and full-field electroretinogram testing.

Results: Four distinct prognostic groups were defined according to the spatial severity of disease features across the fundus. Recurring genotypes were observed in milder prognoses including those associated with a newly defined class of rare hypomorphic alleles. PVS1 (predicted null) variants were enriched in the most severe prognoses; however, missense variants comprised a larger than expected fraction of these patients. Analysis of allele combinations and their respective prognostic severity, showed that certain variants such as p.(Gly1961Glu), and both rare and frequent hypomorphic alleles, are “clinically dominant” with respect to patient phenotypes irrespective of the allele in *trans*.

Conclusion: These results provide much needed structure to the complex genetic and clinical landscape of ABCA4 disease and adds a tool to the clinical repertoire to quantitatively assess individual genotype-specific prognoses in patients.

Introduction

Pathogenic variants in the *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),(2) 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.(3) This phenotypic heterogeneity is matched by an equally extensive array of pathogenic variation across the ~140 kb-spanning *ABCA4* locus (1p22.1). To date, more than 1,500 disease-causing variants have been identified in patients.(4) Consistent with the model that clinical phenotypes are dependent on the residual activity of ABCA4 protein,(5, 6) 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.(6-10)

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 slow-progressing disease trajectory in patients who often present with transient phenotypes such as bull's eye maculopathy and occult macular dystrophy.(11, 12) 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 (MAF ≈ 0.5% in Europeans), and much higher in some ethnic groups.(13, 14) We

recently resolved this controversy, at least in most part, by showing that the contribution of an additional deep intronic variant, c.769-784C>T,(15, 16) present in *cis*, is required for clinical penetrance, particularly in homozygotes.(17) 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 general 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.(18, 19)

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, occult macular dystrophy, etc. 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 re-classified 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.

Results

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 1. Clinical data from the most recent visit for each patient were used in the study. Each patient was categorized into one of four "Prognosis" categories based on the observable spatial progression of ABCA4-associated disease features in the fundus (Figure 1A) by age 50 years or older. Patients categorized as having Prognosis 1 (n=28) had the mildest disease outcome (in the cohort) manifesting early RPE 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 1A, yellow arrowheads). All patients with Prognosis 3 (n=20) had multi-focal 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.

There were no significant differences in the mean age of patients between Prognosis groups (Supplemental Data 1). The mean age of symptomatic onset was earliest among patients with Prognosis 4 (17.1 years) compared to 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 1B, see Supplemental Data 2). Best-corrected visual acuity (BCVA) from the most recent visits were also poorest among patients with Prognosis 4 of which ~40% were counting fingers

or worse (P < 0.00001) (Figure 1C, see Supplemental Data 3). Comparatively, BCVA distributions were multi-modal among patients with Prognosis 1-3 most of whom had 20/200 (logMAR 1.00) or worse and 20/20 in cases with foveal sparing. Full-field ERG responses were largely unremarkable in Prognosis 1 and 2 (Figure 1D). Significant defects were found in Prognosis 3 (50% Lois Group II) and Prognosis 4 (93% Lois Group III). There were no significant differences in mean age of patients in across Prognosis groups (P = 0.254) (Supplemental Table 3).

Classification of p.(Gly1961Glu), p.(Asn1868Ile) and a 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 in Prognosis 1 and Prognosis 2 (Figure 2A). Despite the advanced age of this cohort, 3 patients (P12, P18 and P20) presented with early stage bull's eye maculopathy (Supplemental Figure 1A-1C). As we have previously shown, p.(Asn1868Ile) is highly associated with foveal sparing which is a major contributing factor to the delayed symptomatic onset age in most patients (Figure 2B).(19) 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), p.(Arg2030Gln), that have features in common with p.(Asn1868Ile), most notably, delayed symptomatic onset due to foveal sparing (Table 1, Figure 2C and 2D). Disease features in the fundus of these cases were confined to a delineable area around the vascular arcades in a reticular appearance studded along the peripheral boundary with elongated "tails"

projecting eccentrically in a radial pattern (Figure 3A and 3B and Figure 4A and 4B). Generalized dysfunction of the cone and rod systems were not detected on full-field electroretinogram (ffERG) testing (Figure 3C). 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 conflicting interpretations of pathogenicity (Supplemental Table 2). Furthermore, as has been observed with p.(Asn1868Ile), the allele *in trans* in these genotypes are mostly loss-of-function alleles, including an 8.4 kb deletion that was identified in P58 (Figure 4C). Considering these differences, we separated these mild *ABCA4* alleles into three classes: p.(Gly1961Glu), Frequent hypomorph and Rare hypomorph.

Classification of PVS1 and "severe" non-PVS1 alleles

The distribution of PVS1 (i.e., null or loss-of-function alleles) (Table 1) were skewed towards the most severe clinical phenotypes although at a lower-than-expected proportion. Genotypes with a PVS1 allele comprised ~1/3 of Prognosis 3 and Prognosis 4 cases while the remaining ~2/3 fraction 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).(19) Using our current dataset, 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 separate “severe” sub-class (Table 1).

Classification of moderate variants

After classifying 65% of alleles in the study cohort as either 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 to the other classified allele groups which skewed accordingly towards mild or severe Prognoses (Figure 5A). The coding effect in 93% of these alleles is missense (Figure 5B). The three non-missense variants in this group were an exonic in-frame duplication, deep intronic 15 nucleotide deletion and the known c.859-9T>C variant, which prior midi-gene studies in HEK293T cells have determined to have a “moderate” effect as the variant results in 75% of wild-type *ABCA4* RNA.(16) Considering the nonspecific genetic attributes of these alleles and their collectively uniform distribution across Prognosis categories, we classified them in a “Moderate” group.

Construction of a genotype-phenotype correlation matrix

We generated probability matrices representing correlations between the four clinical Prognosis categories (Prognosis 1-4) and all possible genotypic combinations for the following allele classes: p.(Gly1961Glu), Frequent hypomorph, Rare hypomorph, Moderate, Severe and PVS1 (Table 1, Figure 6). Genotypes consisting of either 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 6A-6C). Genotypes of these three allele classes were also the least heterogenous in

terms of prognostic distribution (P = 0.1164, two-sided FET, see Supplemental Data 4) compared to both Moderate and Severe/PVS1 genotypes (P < 0.001, two-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 these three allele classes exhibit a form of “clinical dominance” whereby the allele in trans, while necessary for disease expression, has minimal to no effect on the phenotypic variability.

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 prognostic severity for these allele combinations. For instance, Moderate allele genotypes with another Moderate allele in *trans* give a 43% probability of having Prognosis 1 whereas having a p.(Gly1961Glu) allele in *trans* increases the Prognosis 1 probability to 83% and having Severe or PVS1 allele in *trans* reduces the Prognosis 1 probability to 0-12% and increases the probability of Prognosis 3 and 4 to 44-55% (Figure 6D). Similar trends were true 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 two allele classes (Figure 6E and 6F). 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 these representing the genotypes of all patients across the study cohort (Figure 7).

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 that 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,(12, 20-24) 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, 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 which 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 clinical 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 slow-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 *X. laevis* tadpole retinae and HEK293T cells revealed no observable defects in sub-cellular localization.(25, 26) The latter study also showed that p.(Ala1038Val) mutant structure closely resembles the WT ABCA4 structure using single particle analysis (cryo-EM).(26) The clinical phenotypes of all well-characterized patients harboring p.(Ala1038Val),(26) p.(Arg2030Gln),(27, 28) p.(Pro1486Leu),(29) p.(Thre1526Met)(27, 28, 30) and p.(Ile1562Thr)(30, 31) 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 which has been proposed for ABCA4,(32) patient phenotypes are expected to vary according to the combined effects of both *ABCA4* alleles and indeed, the phenotypic outcome of Moderate and Severe/PVS1 alleles vary widely depending on the allele in *trans* (Figure 5D-5F). 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., foveal sparing (hypomorphs) and optical gap (p.(Gly1961Glu))) irrespective of the type of allele

in *trans*. This phenomenon may be partially explained by the non-penetrance of mild genotypes resulting in a 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 re-classified 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 ~39% correctly spliced mRNA in HEK293T cells.(16) Consistent with other clinical studies,(6, 28, 33, 34) we also found the variant to be exclusively associated with severe phenotypes (compound heterozygous in two patients with Prognosis 3 and three patients with Prognosis 4) which led us to the 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 Prognosis 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, biallelic PVS1 genotypes which are known to underlie the most severe ABCA4 disease phenotypes such as RP-like, ROC and cone-rod dystrophy were not included.(6-10) Prognostic assessment in these cases, however, usually unambiguous as visual

deterioration begins early in life and disease progresses rapidly. The four 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 multi-ethnic, 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 patients. Two major disease-causing variants of ABCA4, p.Gly1961Glu) and p.(Asn1868Ile) accounted for more than half of the genotypes (patients) with mildest prognoses. We also identified new class of rare hypomorphic variants among mild prognoses cases which, together with p.Gly1961Glu) and p.(Asn1868Ile), exhibit "clinical dominance" in their consistent clinical features irrespective of the allele in *trans*. We identify a large group of missense variants that are associated with the more severe prognoses, and clinically re-classified 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 assess disease severity in patients and as a framework for designing and selecting of patients for clinical trials.

Methods

Study subjects and clinical characterization

Patients diagnosed with Stargardt or *ABCA4*-related disease were recruited from the Department of Ophthalmology at Columbia University Irving Medical Center. In total, 112 unrelated patients harboring two 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 symptomatic onset is ~35 years (interquartile range 28-48 years).(19) Each patient underwent a complete ophthalmic examination by a retinal physician (SHT), which included slit-lamp and dilated fundus examination, best corrected visual acuity (BCVA; Snellen), color fundus photography, fundus autofluorescence (AF, 488-mn, 532-nm and 787-nm), spectral domain-optical coherence tomography (SD-OCT) scanning and full-field electroretinogram (ffERG) testing. Conversion of “counting fingers” (CF) and “hand motion” (HM) to logMAR units were 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 were conducted following pupil dilation (>7mm) with tropicamide (1%) and phenylephrine hydrochloride (2.5%). Fundus autofluorescence (488-nm) images and 9mm horizontal foveal SD-OCT scans were acquired with the Spectralis HRA+OCT (Heidelberg Engineering, Heidelberg, Germany). Ultra-widefield autofluorescence images were acquired with an Optos 200 Tx (Optos PLC, Dunfermline, United Kingdom). Full-field electroretinograms (ffERGs) were recorded silver impregnated fiber electrodes (DTL; Diagnosys LLC, Littleton, MA) on the Espion Visual Electrophysiology System (Diagnosys LLC, Littleton, MA, USA) in accordance with International Society for Clinical Electrophysiology of Vision (ISCEV) standards.(36)

ffERG classifications were assigned according to electrophysiological attributes described by Lois et al.(23, 37) 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 (I, II, III or IV) were determined by two independent graders (W.L. and P.Y.S.) using 55° AF (488-nm) images of each eye for all study patients. In patients with inter-ocular discordance, the Prognosis classification was assigned according to the more advanced eye. Discordant evaluations between graders were adjudicated by an additional grader (S.H.T.). 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 three graders were blinded to the *ABCA4* genotype of each patient at the time of Prognosis classification.

Molecular analyses

Screening of the ABCA4 gene was performed by next-generation sequencing (NGS) as previously described.(38, 39) 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(40) using pathogenicity scores of M-CAP, REVEL, Eigen, and CADD (v1.6). As a general guideline, pathogenic consequences are predicted for variants with scores over 0.025 for MCAP, 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).

Statistics

A detailed summary of all statistic calculations is provided in the Appendix. Comparison of mean characteristics between prognosis categories were determined by a One-way ANOVA test with post-hoc Tukey HSD and Kruskal-Wallis test. Significance was set at alpha level <0.05. Density plots were generated using the ggridges package in R version 4.0.4. Fisher's Exact Tests for Count Data (2x3 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 Medical Center. The study adhered to tenets set out in the Declaration of Helsinki.

Conflicts of interests

S.H.T. has received support from Abeona Therapeutics, Inc. and is a board member of Emendo Biotherapeutics, Nanoscope and Rejuvitas, Inc. The other authors declare no competing interests.

Author Contributions

W.L. designed the study, recruited study subjects, acquired and analyzed clinical data and wrote the manuscript; J.Z. performed sequencing, analyzed molecular data and critically revised the manuscript; P.Y.S. recruited subjects and acquired clinical data, T.N. assisted with molecular analyses, S.H.T. clinically examined study subjects, R.A., supervised the study, critically revised the manuscript and obtained research funding.

Funding

This work was supported, in part, by the National Eye Institute, NIH grants R01 EY028203, R01 EY028954, R01 EY029315, P30 19007 (Core Grant for Vision Research), the Foundation Fighting Blindness USA, grant no. PPA-1218-0751-COLU, and the unrestricted grant to the Department of Ophthalmology, Columbia University, from Research to Prevent Blindness.

Acknowledgements

The authors acknowledge Professor Joseph Terwilliger for helpful discussions and Jimmy Duong and Professor Wei-Yann Tsai of the Biostatistics, Epidemiology, and Research Design (BERD), Irving Institute for Clinical and Translational Research for helpful statistical consultation on the project. Lastly, all authors would like to thank the patients and their families for their immeasurable contributions to our *ABCA4* studies over the years.

References

1. Allikmets R, Singh N, Sun H, Shroyer NF, Hutchinson A, Chidambaram A, et al. A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. *Nat Genet.* 1997;15(3):236-46.
2. Stargardt K. Über familiäre, progressive degeneration in der maculagegend des auges. *Albrecht von Graefes Arch Klin Ophthalmology.* 1909;71:534–50.
3. Cremers FPM, Lee W, Collin RWJ, and Allikmets R. Clinical spectrum, genetic complexity and therapeutic approaches for retinal disease caused by ABCA4 mutations. *Prog Retin Eye Res.* 2020;79:100861.
4. Cornelis SS, Bax NM, Zernant J, Allikmets R, Fritsche LG, den Dunnen JT, et al. In Silico Functional Meta-Analysis of 5,962 ABCA4 Variants in 3,928 Retinal Dystrophy Cases. *Hum Mutat.* 2017;38(4):400-8.
5. Maugeri A, van Driel MA, van de Pol DJ, Klevering BJ, van Haren FJ, Tijmes N, 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.* 1999;64(4):1024-35.
6. Cremers FP, van de Pol DJ, van Driel M, den Hollander AI, van Haren FJ, Knoers NV, 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.* 1998;7(3):355-62.
7. Tanaka K, Lee W, Zernant J, Schuerch K, Ciccone L, Tsang SH, et al. The Rapid-Onset Chorioretinopathy Phenotype of ABCA4 Disease. *Ophthalmology.* 2018;125(1):89-99.

8. Martinez-Mir A, Paloma E, Allikmets R, Ayuso C, del Rio T, Dean M, et al. Retinitis pigmentosa caused by a homozygous mutation in the Stargardt disease gene ABCR. *Nat Genet.* 1998;18(1):11-2.

9. Lee W, Zernant J, Nagasaki T, Tsang SH, and Allikmets R. Deep Scleral Exposure: A Degenerative Outcome of End-Stage Stargardt Disease. *Am J Ophthalmol.* 2018;195:16-25.

10. Bertelsen M, Zernant J, Larsen M, Duno M, Allikmets R, and Rosenberg T. Generalized choriocapillaris dystrophy, a distinct phenotype in the spectrum of ABCA4-associated retinopathies. *Invest Ophthalmol Vis Sci.* 2014;55(4):2766-76.

11. Noupuu K, Lee W, Zernant J, Tsang SH, and Allikmets R. Structural and genetic assessment of the ABCA4-associated optical gap phenotype. *Invest Ophthalmol Vis Sci.* 2014;55(11):7217-26.

12. Cella W, Greenstein VC, Zernant-Rajang J, Smith TR, Barile G, Allikmets R, et al. G1961E mutant allele in the Stargardt disease gene ABCA4 causes bull's eye maculopathy. *Exp Eye Res.* 2009;89(1):16-24.

13. Guymer RH, Heon E, Lotery AJ, Munier FL, Schorderet DF, Baird PN, et al. Variation of codons 1961 and 2177 of the Stargardt disease gene is not associated with age-related macular degeneration. *Arch Ophthalmol.* 2001;119(5):745-51.

14. Lee W, Schuerch K, Zernant J, Collison FT, Bearelly S, Fishman GA, et al. Genotypic spectrum and phenotype correlations of ABCA4-associated disease in patients of south Asian descent. *Eur J Hum Genet.* 2017;25(6):735-43.

15. Sangermano R, Garanto A, Khan M, Runhart EH, Bauwens M, Bax NM, et al. Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. *Genet Med.* 2019;21(8):1751-60.

16. Sangermano R, Khan M, Cornelis SS, Richelle V, Albert S, Garanto A, et al. ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Res.* 2018;28(1):100-10.

17. Lee W, Zernant J, Nagasaki T, Molday LL, Su PY, Fishman GA, 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.* 2021.

18. Zernant J, Lee W, Nagasaki T, Collison FT, Fishman GA, Bertelsen M, 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.* 2018;4(4).

19. Zernant J, Lee W, Collison FT, Fishman GA, Sergeev YV, Schuerch K, et al. Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration. *J Med Genet.* 2017;54(6):404-12.

20. Di Iorio V, Orrico A, Esposito G, Melillo P, Rossi S, Sbordone S, et al. ASSOCIATION BETWEEN GENOTYPE AND DISEASE PROGRESSION IN ITALIAN STARGARDT PATIENTS: A Retrospective Natural History Study. *Retina.* 2019;39(7):1399-409.

21. Del Pozo-Valero M, Riveiro-Alvarez R, Blanco-Kelly F, Aguirre-Lamban J, Martin-Merida I, Iancu IF, et al. Genotype-Phenotype Correlations in a Spanish Cohort of 506 Families With Biallelic ABCA4 Pathogenic Variants. *Am J Ophthalmol.* 2020;219:195-204.

22. Lewis RA, Shroyer NF, Singh N, Allikmets R, Hutchinson A, Li Y, et al. Genotype/Phenotype analysis of a photoreceptor-specific ATP-binding cassette transporter gene, ABCR, in Stargardt disease. *Am J Hum Genet.* 1999;64(2):422-34.

23. Fujinami K, Lois N, Davidson AE, Mackay DS, Hogg CR, Stone EM, et al. A longitudinal study of stargardt disease: clinical and electrophysiologic assessment, progression, and genotype correlations. *Am J Ophthalmol.* 2013;155(6):1075-88 e13.

24. Fujinami K, Lois N, Mukherjee R, McBain VA, Tsunoda K, Tsubota K, et al. A longitudinal study of Stargardt disease: quantitative assessment of fundus autofluorescence, progression, and genotype correlations. *Invest Ophthalmol Vis Sci.* 2013;54(13):8181-90.

25. Wiszniewski W, Zaremba CM, Yatsenko AN, Jamrich M, Wensel TG, Lewis RA, et al. ABCA4 mutations causing mislocalization are found frequently in patients with severe retinal dystrophies. *Hum Mol Genet.* 2005;14(19):2769-78.

26. Zhang N, Tsybovsky Y, Kolesnikov AV, Rozanowska M, Swider M, Schwartz SB, et al. Protein misfolding and the pathogenesis of ABCA4-associated retinal degenerations. *Hum Mol Genet.* 2015;24(11):3220-37.

27. Cideciyan AV, Swider M, Aleman TS, Tsybovsky Y, Schwartz SB, Windsor EA, et al. ABCA4 disease progression and a proposed strategy for gene therapy. *Hum Mol Genet.* 2009;18(5):931-41.

28. Fakin A, Robson AG, Chiang JP, Fujinami K, Moore AT, Michaelides M, 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.* 2016;57(14):5963-73.

29. Riveiro-Alvarez R, Lopez-Martinez MA, Zernant J, Aguirre-Lamban J, Cantalapiedra D, Avila-Fernandez A, et al. Outcome of ABCA4 disease-associated alleles in autosomal recessive retinal dystrophies: retrospective analysis in 420 Spanish families. *Ophthalmology.* 2013;120(11):2332-7.

30. Shroyer NF, Lewis RA, Yatsenko AN, Wensel TG, and Lupski JR. Cosegregation and functional analysis of mutant ABCR (ABCA4) alleles in families that manifest both Stargardt disease and age-related macular degeneration. *Hum Mol Genet.* 2001;10(23):2671-8.

31. Yatsenko AN, Wiszniewski W, Zaremba CM, Jamrich M, and Lupski JR. Evolution of ABCA4 proteins in vertebrates. *J Mol Evol.* 2005;60(1):72-80.

32. Schindler EI, Nylen EL, Ko AC, Affatigato LM, Heggen AC, Wang K, et al. Deducing the pathogenic contribution of recessive ABCA4 alleles in an outbred population. *Hum Mol Genet.* 2010;19(19):3693-701.

33. Green JS, O'Rielly DD, Pater JA, Houston J, Rajabi H, Galutira D, et al. The genetic architecture of Stargardt macular dystrophy (STGD1): a longitudinal 40-year study in a genetic isolate. *Eur J Hum Genet.* 2020;28(7):925-37.

34. Klevering BJ, van Driel M, van de Pol DJ, Pinckers AJ, Cremers FP, and Hoyng CB. Phenotypic variations in a family with retinal dystrophy as result of different mutations in the ABCR gene. *Br J Ophthalmol.* 1999;83(8):914-8.

35. Schulze-Bonsel K, Feltgen N, Burau H, Hansen L, and Bach M. Visual acuities "hand motion" and "counting fingers" can be quantified with the freiburg visual acuity test. *Invest Ophthalmol Vis Sci.* 2006;47(3):1236-40.

36. McCulloch DL, Marmor MF, Brigell MG, Hamilton R, Holder GE, Tzekov R, et al. ISCEV Standard for full-field clinical electroretinography (2015 update). *Doc Ophthalmol.* 2015;130(1):1-12.

37. Lois N, Holder GE, Bunce C, Fitzke FW, and Bird AC. Phenotypic subtypes of Stargardt macular dystrophy-fundus flavimaculatus. *Arch Ophthalmol.* 2001;119(3):359-69.

38. Zernant J, Schubert C, Im KM, Burke T, Brown CM, Fishman GA, et al. Analysis of the ABCA4 gene by next-generation sequencing. *Invest Ophthalmol Vis Sci.* 2011;52(11):8479-87.

39. Zernant J, Xie YA, Ayuso C, Riveiro-Alvarez R, Lopez-Martinez MA, Simonelli F, et al. Analysis of the ABCA4 genomic locus in Stargardt disease. *Hum Mol Genet.* 2014;23(25):6797-806.

40. Wang K, Li M, and Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.

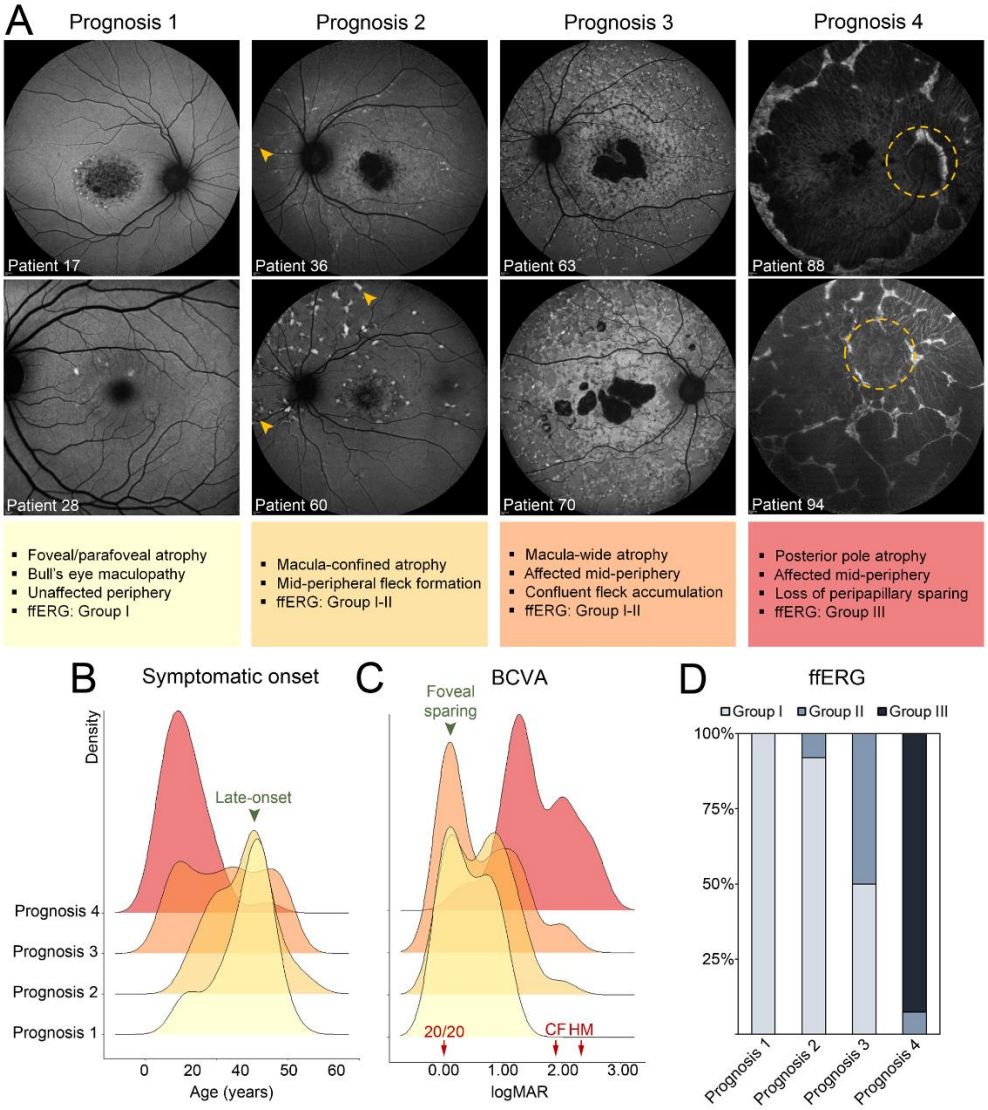


Figure 1: Clinical characteristics of four prognostic outcomes observed in 112 patients (>50 years of age) with ABCA4 disease. Each prognosis was defined according to observable spatial progression of disease features detected at the most recent visit in each patient. (A) Representative autofluorescence images and clinical descriptions of patients ineach Prognosis classification. Extramacular development of flecks in Prognosis 2 are indicted by yellow arrow heads. The position of the optic nerve is encircled by the dotted yellow line. (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). BCVA were presented in logMAR units with corresponding Snellen equivalents (20/20, counting fingers (CF) and hand motion (HM), red arrows) provided. (D) Proportion of full-field electroretinogram (ffERG) groupings according to the classification by Lois et al.(37) for each prognosis category. Group I, normal responses; Group II, attenuation of cone responses; Group III, attenuation of cone and rod responses.

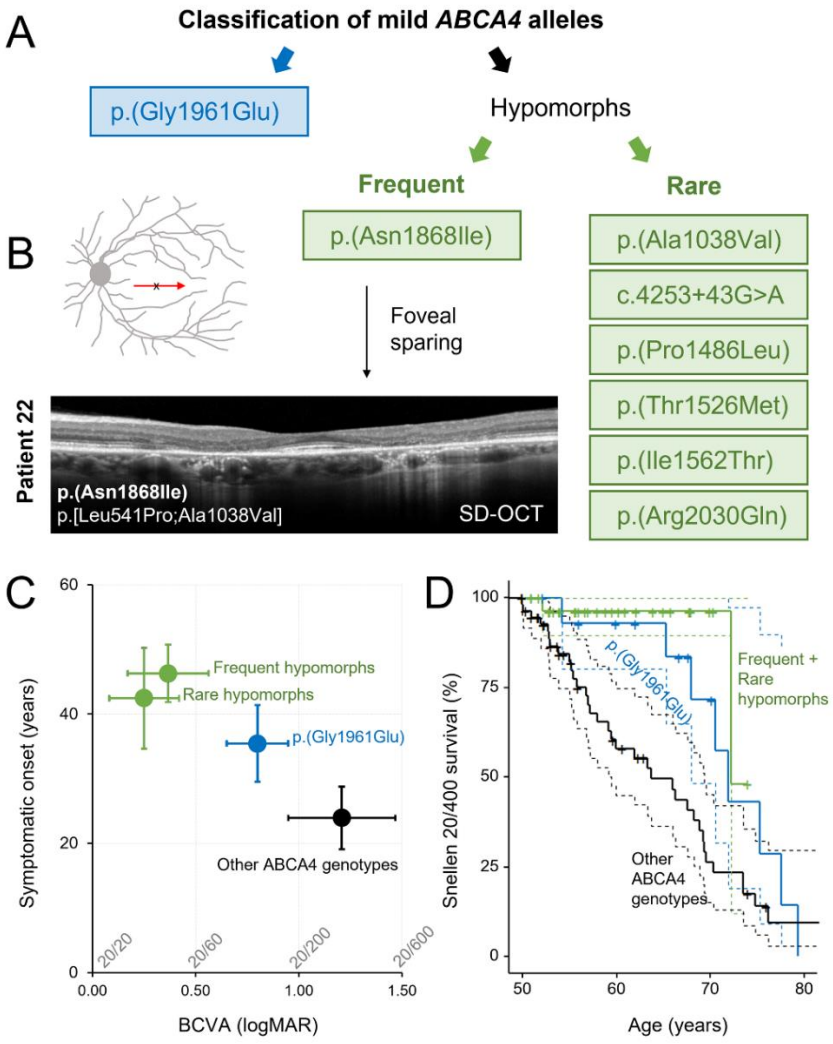


Figure 2: Classification and phenotypic characterization of mild *ABCA4* alleles. (A) Mild *ABCA4* alleles identified in patients with mild prognoses included p.(Gly1961Glu) and two hypomorphic allele sub-groups: frequent hypomorphs which consisted of p.(Asn1868Ile) and rare hypomorphs which consisted of p.(A1038V), c.4253+43G>A, p.(Pro1486Leu), p.(Thr1526Met), p.(Ile1562Thr), p.(Arg2030Gln). (B) Horizontal spectral domain-optical coherence tomography (SD-OCT) scan showing structural preservation of the fovea in Patient 22, an allele-specific sub-phenotype common amongst 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. BCVA 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), 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.

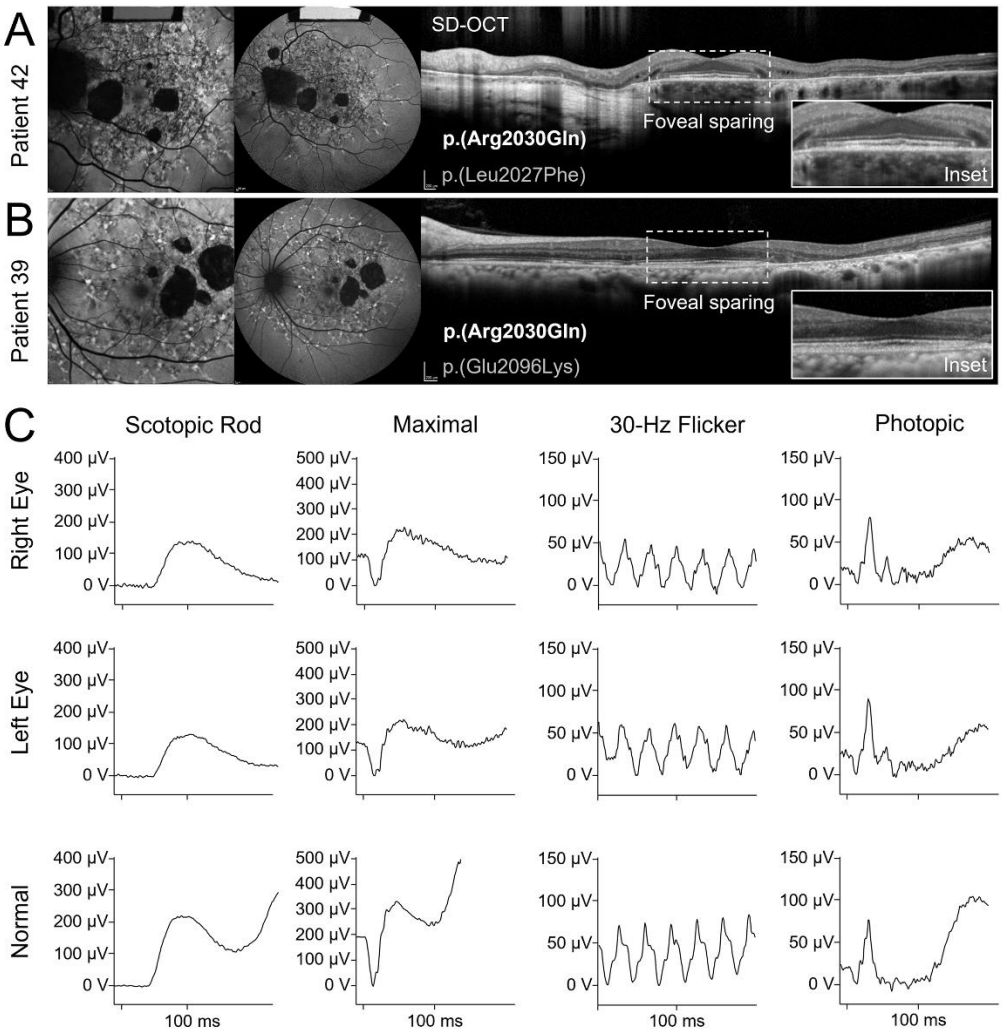
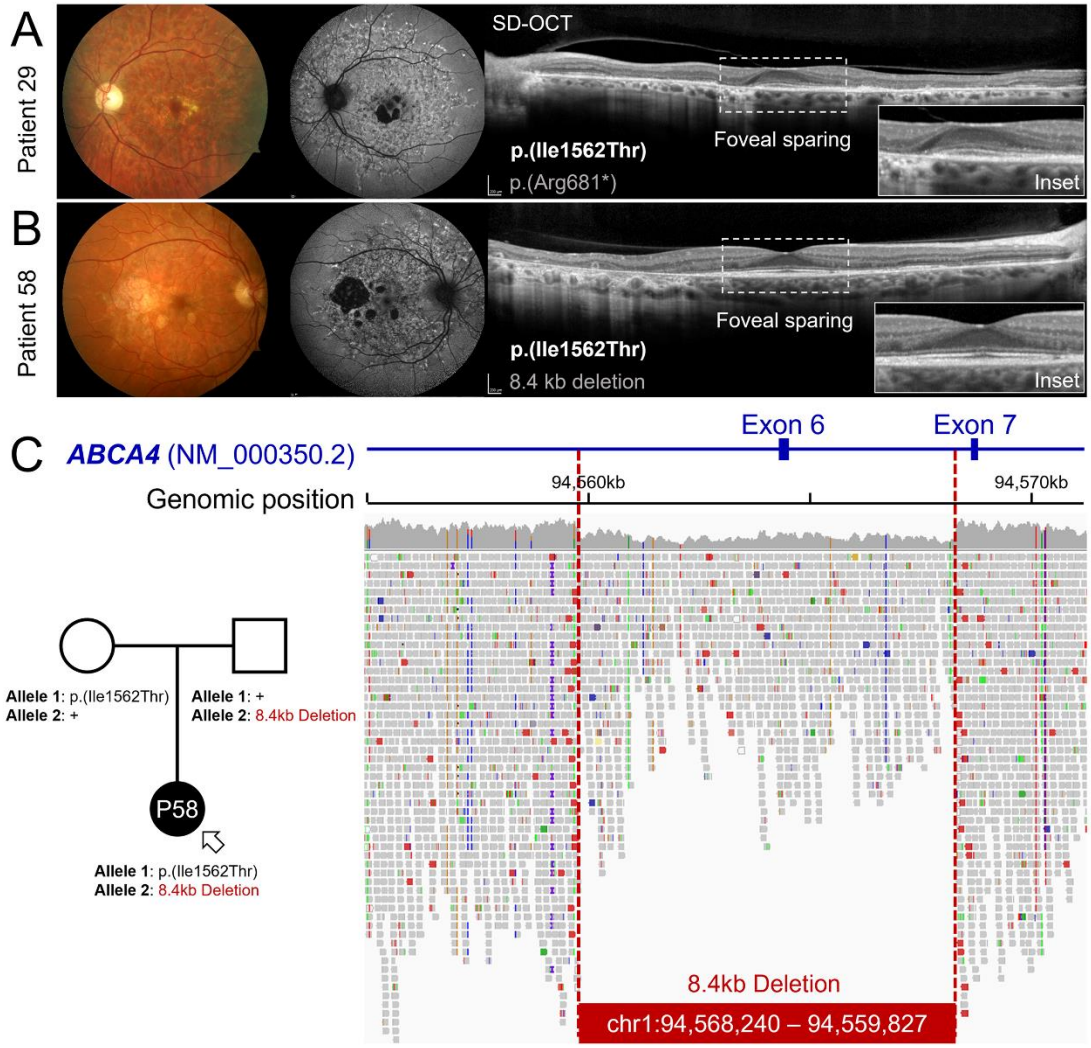
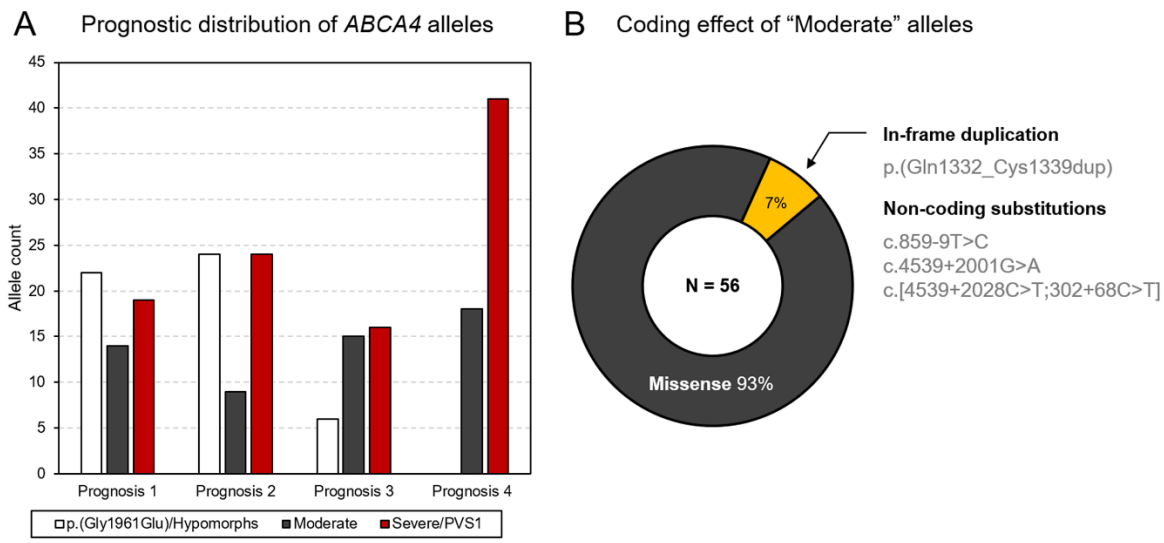


Figure 3: Retinal phenotype of the rare hypomorph p.(Arg2030Gln) variant of ABCA4 disease. Macular 30° autofluorescence, 55° autofluorescence images and horizontal spectral domain-optical coherence tomography (SD-OCT) scans of the (A) left eye of Patient 42 and (B) left eye of Patient 40. SD-OCT scans with enlarged insets of the fovea show preservation of outer retinal layers resulting in 20/20 vision in the eyes of both patents. Unimpaired 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 Patient 39 and representative waveforms from an age-matched healthy control eye.



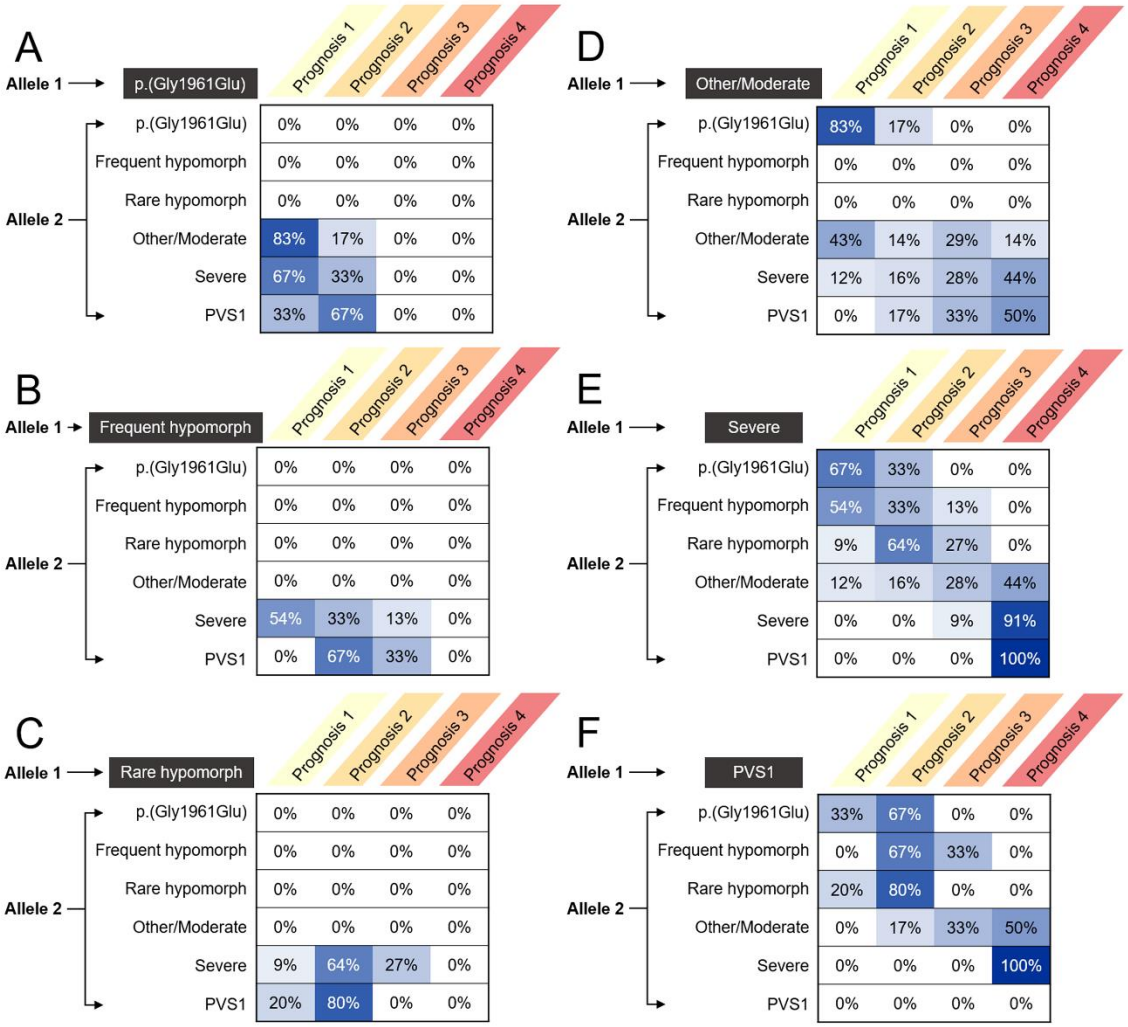
573

574 **Figure 4**: Retinal phenotype of the rare hypomorph p.(Ile1562Thr) variant of ABCA4
575 disease. Color fundus photographs, autofluorescence images and horizontal spectral
576 domain-optical coherence tomography (SD-OCT) of the (A) left eye of Patient 29 and (B)
577 right eye of Patient 58. SD-OCT scans with enlarged insets of the fovea show
578 preservation of outer retinal resulting in unimpaired 20/20 vision in the eyes of both
579 patents. (C) Pedigree showing segregation of the p.(Ile1562Thr) and large 8.4 kb deletion
580 alleles in Patient 58. Pileup of whole genome sequencing reads showing the approximate
581 size and genomic position of the *ABCA4* deletion which spans the entire length of exon
582 6.



583

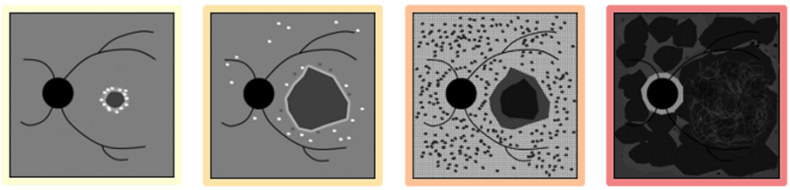
584 **Figure 5**: Clinical and genetic characteristics of moderate ABCA4 alleles in the study. (A)
585 Distribution of p.(Gly1961Glu) and hypomorphs (white bars), moderate (gray bars) and
586 Severe/PVS1 (red bars) alleles across Prognosis categories. (B) Coding effect of alleles
587 designated as "Moderate" in patients with ABCA4 disease.



588

589 **Figure 6**: Prognostic probabilities (%) of all possible combinations for each allele class:
590 (A) p.(Gly1961Glu), (B) Frequent hypomorph, (C) Rare hypomorph, (D) Moderate, (E)
591 Severe and (F) PVS1. Percentages represent the observed fraction of patients across
592 each Prognosis category for a given Allele 1 and Allele 2 combination.

ABCA4 Genotype-Phenotype Correlation Matrix



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 7: Genotype-phenotype correlation matrix based on the long-term prognostic outcomes of 112 genetically confirmed patients with ABCA4 disease patients. 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 Supplementary Table 2.

Table 1: Classification criteria and list of all pathogenic ABCA4 alleles in the study cohort

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	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

Footnotes: 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. The PVS1 classification was used in accordance with the ACMG/AMP Standards and Guidelines for the Interpretation of Sequence Variants.(37)