TALLINN UNIVERSITY OF TECHNOLOGY
DOCTORAL THESIS
36/2019

ABCA4 Locus as a Model for Complex Genetics in Monogenic Diseases

JANA ZERNANT

TAL TECH PRESS

TALLINN UNIVERSITY OF TECHNOLOGY
Faculty of Science
Department of Chemistry and Biotechnology: Division of Gene Technology

This dissertation was accepted for the defence of the degree: 29.05.2019

Supervisors: William and Donna Acquavella Professor
Rando Allikmets, PhD
Departments of Ophthalmology, and Pathology & Cell Biology
Columbia University, New York, NY, USA

Associate Professor Andres Veske, PhD
Department of Chemistry and Biotechnology:
Division of Gene Technology
Tallinn University of Technology, Estonia

Opponents: Senior Investigator Michael Dean, PhD
Division of Cancer Epidemiology & Genetics
Laboratory of Translational Genomics
National Cancer Institute, Gaithersburg, MD, USA

Senior Researcher Tarmo Annilo, PhD
Estonian Genome Center Science Center
University of Tartu, Estonia

Defence of the thesis: 21.06.2019, Tallinn

Declaration:
Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology has not been submitted for doctoral or equivalent academic degree elsewhere.

Jana Zernant

---

signature

European Union
European Regional
Development Fund

Investing
in your future

ISSN 2585-6898 (publication)
ISBN 978-9949-83-454-9 (publication)
ISSN 2585-6901 (PDF)
ISBN 978-9949-83-455-6 (PDF)

TALLINNA TEHNIKAÜLIKOOL
DOKTORITÖÖ
36/2019

ABCA4 lookus kui monogeensete haiguste kompleksse geneetika mudel

JANA ZERNANT

TAL TECH
KIRJASTUS

To the patients with ABCA4 disease and their families

CONTENTS

INTRODUCTION

Inherited retinal dystrophies present a genetically heterogeneous group of disorders affecting visual functions in as many as 1 in 3,000 individuals around the world. Currently, more than 300 genes encoding a wide variety of proteins with different cellular functions have been associated with inherited retinal disease phenotypes.

Mutations in the *ABCA4* gene cause autosomal recessive Stargardt disease (STGD1, MIM #248200), the most frequent form of inherited juvenile macular degeneration.[1] Stargardt disease is characterized by slowly progressing central vision loss, with variable age of onset (usually in the 2nd decade), variable disease course and mostly poor final visual outcome.[2, 3]

Phenotypes caused by mutations in the *ABCA4* gene range from late onset mild cases of central atrophy to very early onset, retinitis pigmentosa (RP)-like, panretinal degeneration.[4-10] Phenotypic variability is caused by extensive genetic variation in the *ABCA4* gene, with >1000 definitely or possibly disease-causing variants described in the coding sequences and splice sites[11] (JZ and RA, unpublished data).

While the understanding of disease-associated genetic variation in *ABCA4* has substantially improved since the discovery of the gene in 1997, diagnostic sequencing of the *ABCA4* gene in patients diagnosed with STGD1 disease was, until recently, successful in only 60%–70% of cases where both disease-causing *ABCA4* alleles were identified. The remaining cases included patients with one *ABCA4* mutation (20%–25%) or no mutations (10%–15%) identified.[12] These fractions depend on many variables, most importantly the quality of the clinical diagnosis and the ethnic composition of the cohort. The general population frequency of *ABCA4* alleles is high—1:20 people carry a potentially pathogenic *ABCA4* allele.[13] However, many *ABCA4* variants are more common in patients with specific geographic and ethnic backgrounds, while absent in others.[14, 15]

Identifying the exact genetic cause of a disease and correlating it with clinical phenotype facilitates proper diagnosis, which is crucial for prognosing disease progression, genetic counseling of patients, and selecting patients for clinical trials.

The thesis starts with a brief overview of Stargardt disease and continues with the analysis of genetic causality in the *ABCA4* locus, the main subject of my research. I am presenting my results as a part of a comprehensive review on the topic, combined and discussed with the concurrent and most recent data from the literature.

AIMS OF RESEARCH

This research has been dedicated to:

1) Developing comprehensive targeted screening methods for the analysis of the *ABCA4* gene;
2) Determining genetic variation of *ABCA4*-associated disease in previously not characterized ethnic cohorts;
3) Finding the 'missing' *ABCA4* mutations in genetically unsolved monoallelic STGD1 patients, by a combination of next-generation sequencing, comparative genomic hybridization arrays (aCGH), and *in silico*, allele frequency, and segregation analyses in families.

LIST OF PUBLICATIONS

The present thesis is based on the following publications, referred to in the text by their Roman numerals:

I **Zernant J**, Collison FT, Lee W, Fishman GA, Noupuu K, Yuan B, Cai C, Lupski JR, Yannuzzi LA, Tsang SH, Allikmets R (2014). Genetic and Clinical Analysis of *ABCA4*-Associated Disease in African American Patients. Hum Mutat. 35(10):1187-94.

II **Zernant J**, Xie YA, Ayuso C, Riveiro-Alvarez R, Lopez-Martinez MA, Simonelli F, Testa F, Gorin MB, Strom SP, Bertelsen M, Rosenberg T, Boone PM, Yuan B, Ayyagari R, Nagy PL, Tsang SH, Gouras P, Collison FT, Lupski JR, Fishman GA, Allikmets R (2014). Analysis of the *ABCA4* genomic locus in Stargardt disease. Hum Mol Genet. 23(25):6797-806.

III **Zernant J***, Lee W*, Collison FT, Fishman GA, Sergeev YV, Schuerch K, Sparrow JR, Tsang SH, Allikmets R (2017). Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration. J Med Genet. 54(6):404-412.

* Authors contributed equally

AUTHOR'S CONTRIBUTION TO THE PUBLICATIONS

I Design and application of *ABCA4* array screening and targeted next-generation sequencing of the *ABCA4* gene in patients of African American decent, analysis and interpretation of the identified DNA variants, comparison to Caucasian population, drafting of the manuscript.

II Design and application of targeted next-generation sequencing of the *ABCA4* locus in patients; analysis and interpretation of the identified DNA variants, *ABCA4* RNA analysis from retina, drafting of the manuscript.

III Sequencing, analysis and interpretation of the identified DNA variants, drafting of the manuscript.

ADDITIONAL RELATED PUBLICATIONS

Allikmets R, **Zernant J**, Lee W. - Penetrance of the ABCA4 p.Asn1868Ile Allele in Stargardt Disease. Invest Ophthalmol Vis Sci. 2018 Nov 1;59(13):5564-5565.

Zernant J, Lee W, Nagasaki T, Collison FT, Fishman GA, Bertelsen M, Rosenberg T, Gouras P, Tsang SH, Allikmets R. - Extremely hypomorphic and severe deep intronic variants in the *ABCA4* locus result in varying Stargardt disease phenotypes. Cold Spring Harb Mol Case Stud. 2018 Aug 1;4(4). pii: a002733. doi: 10.1101/mcs.a002733. Print 2018 Aug.

Albert S, Garanto A, Sangermano R, Khan M, Bax NM, Hoyng CB, **Zernant J**, Lee W, Allikmets R, Collin RWJ, Cremers FPM - Identification and Rescue of Splice Defects Caused by Two Neighboring Deep-Intronic *ABCA4* Mutations Underlying Stargardt Disease. Am J Hum Genet. 2018 Apr 5;102(4):517-527.

Tanaka K, Lee W, **Zernant J**, Schuerch K, Ciccone L, Tsang SH, Sparrow JR, Allikmets R. - The Rapid-Onset Chorioretinopathy Phenotype of ABCA4 Disease. Ophthalmology. 2018 Jan;125(1):89-99.

Lee W, Schuerch K, **Zernant J**, Collison FT, Bearelly S, Fishman GA, Tsang SH, Sparrow JR, Allikmets R.- Genotypic spectrum and phenotype correlations of *ABCA4*-associated disease in patients of south Asian descent. Eur J Hum Genet. 2017 Jun;25(6):735-743.

Cornelis SS, Bax NM, **Zernant J**, Allikmets R, Fritsche LG, den Dunnen JT, Ajmal M, Hoyng CB, Cremers FP.- In Silico Functional Meta-Analysis of 5,962 *ABCA4* Variants in 3,928 Retinal Dystrophy Cases. Hum Mutat. 2017 Apr;38(4):400-408.

Lee W, Xie Y, **Zernant J**, Yuan B, Bearelly S, Tsang SH, Lupski JR, Allikmets R. - Complex inheritance of ABCA4 disease: four mutations in a family with multiple macular phenotypes. Hum Genet. 2016 Jan;135(1):9-19.

Fujinami K, **Zernant J**, Chana RK, Wright GA, Tsunoda K, Ozawa Y, Tsubota K, Robson AG, Holder GE, Allikmets R, Michaelides M, Moore AT. - Clinical and molecular characteristics of childhood-onset Stargardt disease. Ophthalmology. 2015 Feb;122(2):326-34.

Bertelsen M, **Zernant J**, Larsen M, Duno M, Allikmets R, Rosenberg T. - Generalized choriocapillaris dystrophy, a distinct phenotype in the spectrum of *ABCA4*-associated retinopathies. Invest Ophthalmol Vis Sci. 2014 Apr 29;55(4):2766-76.

Fujinami K, **Zernant J**, Chana RK, Wright GA, Tsunoda K, Ozawa Y, Tsubota K, Webster AR, Moore AT, Allikmets R, Michaelides M. - *ABCA4* Gene Screening by Next-Generation Sequencing in a British Cohort. Invest Ophthalmol Vis Sci. 2013 Oct 11;54(10):6662-74.

Riveiro-Alvarez R, Lopez-Martinez MA, **Zernant J**, Aguirre-Lamban J, Cantalapiedra D, Avila-Fernandez A, Gimenez A, Lopez-Molina MI, Garcia-Sandoval B, Blanco-Kelly F, Corton M, Tatu S, Fernandez-San Jose P, Trujillo-Tiebas MJ, Ramos C, Allikmets R, Ayuso C. - Outcome of *ABCA4* Disease-Associated Alleles in Autosomal Recessive Retinal Dystrophies: Retrospective Analysis in 420 Spanish Families. Ophthalmology. 2013 Nov;120(11):2332-7.

Cella W, Greenstein VC, **Zernant-Rajang J**, Smith TR, Barile G, Allikmets R, Tsang SH. - G1961E mutant allele in the Stargardt disease gene *ABCA4* causes bull's eye maculopathy. Exp Eye Res. 2009 Jun 15;89(1):16-24.

Kong J, Kim SR, Binley K, Pata I, Doi K, Mannik J, **Zernant-Rajang J**, Kan O, Iqball S, Naylor S, Sparrow JR, Gouras P, Allikmets R. - Correction of the disease phenotype in the mouse model of Stargardt disease by lentiviral gene therapy. Gene Ther. 2008 Oct;15(19):1311-20.

Hargitai J, **Zernant J**, Somfai GM, Vamos R, Farkas A, Salacz G, Allikmets R.- Correlation of clinical and genetic findings in hungarian patients with stargardt disease. Invest Ophthalmol Vis Sci. 2005 Dec;46(12):4402-8.

Simonelli F, Testa F, **Zernant J**, Nesti A, Rossi S, Allikmets R, Rinaldi E. - Genotype-phenotype correlation in Italian families with Stargardt disease. Ophthalmic Res. 2005 May-Jun;37(3):159-67.

Rando Allikmets and **Jana Zernant** – From Gene Chips to Disease Chips – New Approach in Molecular Diagnosis of Eye Diseases. DNA Microarrays by Taylor and Francis Group. 2005:83-97.

Simonelli F, Testa F, **Zernant J**, Nesti A, Rossi S, Rinaldi E, Allikmets R. - Association of a homozygous nonsense mutation in the *ABCA4 (ABCR)* gene with cone-rod dystrophy phenotype in an Italian family. Ophthalmic Res. 2004 Mar-Apr;36(2):82-8.

Jaakson K, **Zernant J,** Kulm M, Hutchinson A, Tonisson N, Glavac D, Ravnik-Glavac M, Hawlina M, Meltzer MR, Caruso RC, Testa F, Maugeri A, Hoyng CB, Gouras P, Simonelli F, Lewis RA, Lupski JR, Cremers FP, Allikmets R - Genotyping microarray (gene chip) for the *ABCR (ABCA4)* gene. Hum Mutat. 2003 Nov;22(5):395-403.

ABBREVIATIONS

A2E – N-retinylidene-N-retinylethanolamine
ABCA4 – ATP binding cassette subfamily A member 4
aCGH – array comparative genomic hybridization
all-t-ral – all-trans-retinal
AMD – age-related macular degeneration
APEX – arrayed primer extension
arCRD – autosomal recessive cone-rod dystrophy
arRP – and autosomal recessive retinitis pigmentosa
ATP – adenosine triphosphate
bp – base pair
CADD – Combined Annotation Dependent Depletion
CNV – copy number variant
COS7 – CV-1 in Origin with SV40 genes
DGGE – denaturing gradient gel electrophoresis
dHPLC – denaturing high-performance liquid chromatography
DNA – deoxyribonucleic acid
EIAV – equine infectious anemia lentivirus
ER – endoplasmic reticulum
ESP – Exome Sequencing Project
EVS – exome variant server
FDA – U.S. Food and Drug Administration
ffERG – full-field electroretinalgrams
GA – geographic atrophy
HA – heteroduplex analysis
HEK293 – human embryonic kidney
hESC – human embryonic stem cell
iPSC – induced pluripotent stem cell
kb – kilo base
KI – knock-in
KO – knock-out
LCA – Leber's Congenital Amaurosis
MAC – membrane attack complex
MAF – minor allele frequency
MIM – Mendelian Inheritance in Man
MLPA – Multiplex Ligation-dependent Probe Amplification
NGS – next-generation sequencing
N-ret-PE – N-retinylidene-phosphatidylethanolamine
PCR – polymerase chain reaction
PE – phosphatidylethanolamine
PolyPhen – Polymorphism Phenotyping
PPC – photoreceptor progenitor cell
RBP – retinal binding protein
RDH – retinol dehydrogenase
RFLP – restriction fragment length polymorphism
RHO – rhodopsin
ROS – rod outer segment

RP – retinitis pigmentosa
RPE – retinal pigment epithelium
RT-PCR – reverse transcriptase polymerase chain reaction
SIFT – Sorting Intolerant from Tolerant
SSCP – single strand conformation polymorphism
STGD1 – Stargardt disease type 1, Stargardt disease
VCM – visual cycle modulator
WES – whole exome sequencing
WGS – whole genome sequencing
WT – wild type

1 STARGARDT1/ABCA4 DISEASE

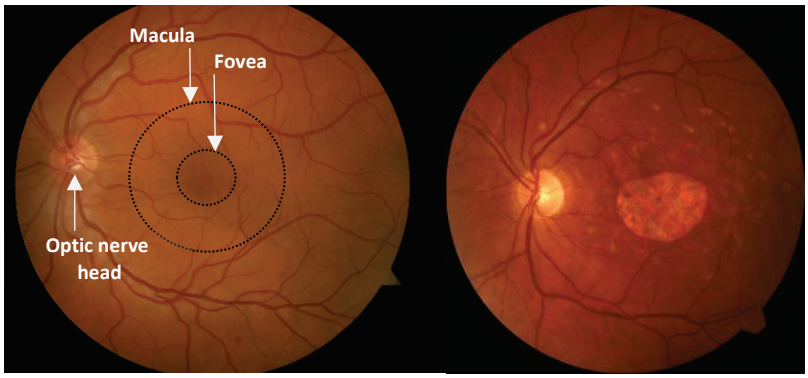
Inherited retinal dystrophies (IRD) comprise a genetically and clinically diverse group of disorders characterized by progressive and irreversible vision loss over time. Visual deterioration is caused by degeneration of cells in the outer retina, specifically photoreceptors and retinal pigment epithelia (RPE) that are responsible for the detection of light entering the eye. Disease phenotypes vary widely in their clinical manifestations such as age of onset, pathogenesis, symptom progression and inheritance pattern.

To date, DNA variants in more than 300 genes have been associated with various forms of IRDs (The Retinal Information Network, RetNet, https://sph.uth.edu/retnet/, accessed February, 2019). The majority code for proteins that are functionally relevant to the retina (i.e., visual cycle) although genes involved in more ubiquitous processes such as neural development and ciliary function have also been identified. The cellular origin of genetic dysfunction within the retina largely determines the clinical pattern of each disease. For instance, mutations in genes such as rhodopsin (RHO) cause widespread degeneration of rods, night blindness and progression from the periphery (rod-cone dystrophies), whereas disorders that predominantly affect cones begin in the macula causing central vision loss (cone/cone-rod dystrophies). Each disease is etiologically distinct; however the disease manifestation between two distantly related disorders can overlap in clinical appearance. Likewise, different mutations in the same gene can give rise to multiple different disease phenotypes illustrating the level of complexity in these disorders and challenges that arise in diagnosing and managing patients in the clinic.[16]

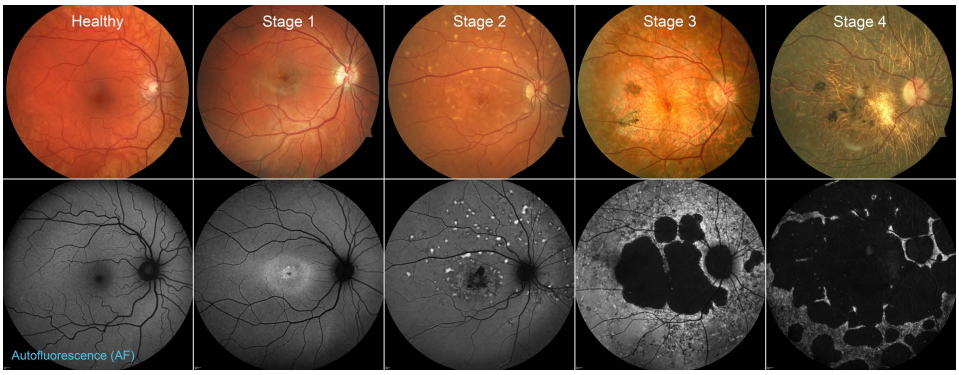
1.1 Clinical presentation of Stargardt disease

Stargardt disease type 1 (STGD1, MIM #248200) is the most common form of juvenile-onset retinal degeneration.[1] STGD1 accounts for approximately 7% of all IRDs, with estimated prevalence of 1 in 8,000 to 10,000[17, 18] It is caused by autosomal recessive mutations in the photoreceptor-specific, ATP-binding cassette, sub-family A, member 4, *ABCA4* gene.[1] Since the discovery of the gene, >1,000 variants within all 50 coding exons, canonical splice sites and deep intronic regions underlying varying disease phenotypes have been reported.

The symptoms of STGD1 typically begin with deterioration of central vision due to localized degeneration of photoreceptors and RPE, within the macula (Figure 1A). The onset of disease occurs most often within the first two decades and is indirectly correlated with disease severity - patients with a symptomatic onset of <10 years progress rapidly and to the point of near-blindness,[3] whereas milder cases remain asymptomatic well into the 4th and 5th decades of life.[19, 20] In addition to outwardly expanding macular atrophy, the disease also exhibits several accompanying pathognomonic features seen on ophthalmic examination on various retinal imaging modalities. Biomarkers associated with the overproduction of lipofuscin include the appearance of yellow, pisciform flecks across the fundus, a dark choroid on fundus angiography and significantly increased levels of 488-nm autofluorescence at the posterior pole.[21, 22]



A



B

Figure 1. Presentation of Stargardt disease. ***A*** *Fundus photographs of a healthy eye (left) compared to Stargardt disease eye (right).* ***B*** *Fundus photographs (top row) and autofluorescence (AF) images (bottom row) documenting the natural history of ABCA4/Stargardt disease according to the clinical classification system proposed by Fishman et al. The healthy fundus exhibits normal retinal vasculature, uniform pigmentation with marginal visibility of underlying choroidal vessels and the darkened central fovea. Stage 1: Early disease changes typically manifest as confined mottling of the RPE on fundoscopy that correspond to parafoveal hyperautofluorescence resulting in loss of central vision. This stage is often clinically referred to as bull's eye maculopathy. Stage 2: Progression from early stage is marked by the centrifugal appearance of yellow, autofluorescent flecks as a result of lipofuscin accumulation. The presence of both macular atrophy and flecks is the most recognizable feature used to diagnose Stargardt disease. Stage 3: Over time, fleck accumulation becomes visibly confluent, leaving behind reticular patches of atrophic RPE and photoreceptor debris. The degenerative central lesion progresses to the extent of choroidal visibility and complete autofluorescence signal attenuation. Peripapillary sparing, another diagnostic feature of Stargardt disease, is most discernible at this stage. Stage 4: Large coalescing lesions atrophy eventually appear throughout the posterior pole at the most advanced stages of the disease. Degeneration at this final stage is most characterized by loss of peripapillary sparing, involvement of underlying choroidal vessels (termed chorioretinal atrophy), an overall beaten bronze appearance of the fundus, attenuation of retinal vessels and optic disc pallor. Significant visual field loss occurs; however, patients often retain functional use of the peripheral retina.*

Several classification systems for the STGD1 disease progression have been described based on clinical and electrophysiological tests. Fishman et al. categorized STGD1 patients into 4 progressive stages (Figure 1B) based on the observed distribution of disease changes in the fundus: Stage 1 defines cases in which the disease is confined within the macula, Stage 2 is indicative of early extramacular transitioning, Stage 3 exhibits the extension of atrophic changes outside the macula and Stage 4 represents widespread RPE and chorioretinal atrophy across the posterior pole.[23] Not all cases, especially those with later onset, progress to the Stages 3-4.[13]

Another classification has been proposed using standard electrophysiological tests. Full-field electroretinalgrams (ffERG) provide a collective measure of cone and rod response to light stimuli across the entire retina. A reduction in the ffERG response is indicative of significant photoreceptor loss across a large portion of the retina; however, localized changes within the macula are not detectable. Unlike the cross-sectional Fishman classification, ffERG grouping, posed by Lois et al. is stationary and therefore more prognostically informative. The study identified three non-transitional groups in a relatively large cohort of patients: Group I patients exhibit no functional attenuation, Group II patients show attenuation and marked delay in cone responses and patients in Group III present with losses in both the cone and rod system.[24]

Currently, no FDA-approved treatment exists for *ABCA4*-associated retinal dystrophies, but several promising therapeutic options based on gene and stem cell therapy, and pharmacotherapy are in clinical trials.[25]

1.2 Disease mechanism

Visual perception starts with light initiating the phototransduction events in the light-sensitive rod and cone photoreceptor cells in the retina that ultimately trigger nerve impulses passing through the optic nerve to the brain (Figure 2).

The *ABCA4* gene encodes for a photoreceptor-specific ATP-binding transporter, localized in the outer segment disc membranes of the rod and cone photoreceptors.[26, 27] Recently, ABCA4 expression was demonstrated also in retinal pigment epithelium (RPE), at 1% level of ABCA4 expression in photoreceptors.[28]

In phototransduction cascade, all-trans-retinal (all-t-ral) is released from rhodopsin molecule in rod and cone outer segment disc membranes after photoexcitation. Spontaneous and reversible binding of all-trans-retinal molecules to membrane phospholipid phosphatidylethanolamine (PE) molecules forms N-retinylidene-phosphatidylethanolamine (N-ret-PE), a compound that is unable to diffuse across the membrane. The proposed function of the ABCA4 protein is to act as transporter that actively flips N-ret-PE molecules from the disc lumen side to the cytoplasmic side of photoreceptor disc membranes, where all-trans-retinal will be accessible for rhodopsin regeneration cycle through retinol dehydrogenase (RDH) and subsequent visual cycle reactions (Figure 3).[29]

Variants in the *ABCA4* gene impairing the transport function of the ABCA4 protein lead to the accumulation of N-ret-PE and all-trans retinal in the discs. These compounds react with each other to form bisretinoid molecules of N-retinylidene-N-retinylethanolamine (A2E) that are progressively accumulating over time in RPE cells as lipofuscin deposits after photoreceptor outer segment disc shedding and daily phagocytosis of these discs by the RPE. The accumulation of lipofuscin leads to apoptosis of RPE cells via combination of several proposed mechanisms: through its action as detergent to disrupt intracellular

membranes, through light-induced generation of disruptive free radicals, and by activating the complement system.[29, 30]

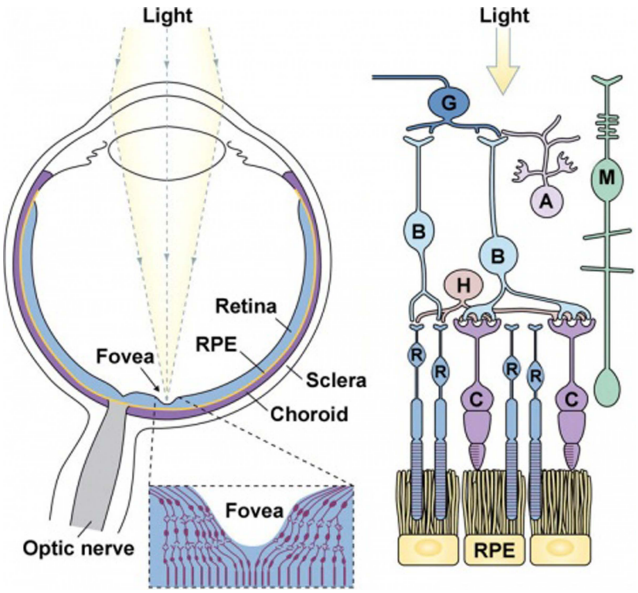


Figure 2. Diagrams of the eye and the organization of cells in retina. *Retina is the light sensing tissue lining the inner surface of the posterior part of the eye. Fovea is the cone-rich area responsible for central vision. The retinal pigment epithelium (RPE) is between retina and choroid. The apical microvilli of RPE cells enclose the outer segments of rod (R) and cone (C) photoreceptors. Other cell types in retina: bipolar cell (B), horizontal cell (H), amacrine cell (A), ganglion cells (G), Müller cell (M) (Sung CH, Chuang JZ. J Cell Biol. 2010 Sep 20; 190(6): 953–963).*

The described model suggests that lipofuscin accumulates only in light conditions, in which case restricting light exposure (i.e., living mostly in darkness) would have a therapeutic effect. Yet, no difference in A2E/lipofuscin accumulation was recorded between dark-reared and cyclic light-reared *Abca4*$^{(-/-)}$ or *Abca4*$^{(wt/wt)}$ mice,[31] suggesting an additional translocation function of ABCA4 for 11-cis-retinal, which is continuously supplied to the rod photoreceptors for rhodopsin regeneration and outer segment renewal. The suggested role for ABCA4 is importing 11-cis-retinal from the cytosole to the rod outer segment disc membrane, to protect the cytosolic enzymatic machinery from the excess of this reactive retinaldehyde. The accumulation of lipofuscin may arise also from defects of the import function, in addition to flipping the all-trans-N-ret-PE. This, the alternating access model is consistent with the high affinity of the lumenal domain of ABCA4 for all-trans-retinal, and cytoplasmic domain for 11-cis-retinal. ABCA4 may also function to regulate the direction of transport of these molecules.[31, 32]

In short, the ABCA4 protein plays an essential role in recycling of visual cycle retinoids to provide a continuous supply of 11-cis-retinal used in photoactivation by rhodopsin, removing the all-trans-retinal after light exposure and, therefore, preventing the toxic accumulation of these retinoid isoforms (Figure 3).

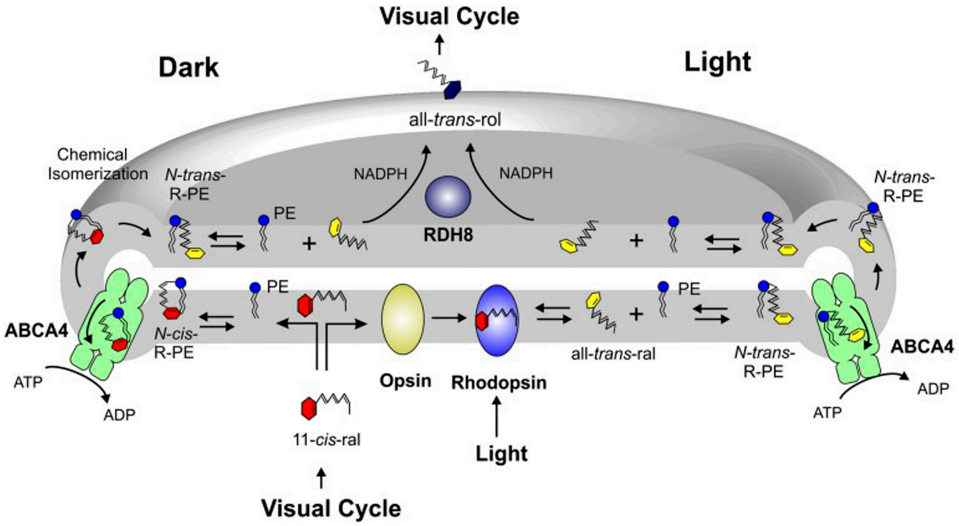


Figure 3. Proposed function of ABCA4 in rod outer segment disk membranes in light and dark conditions. *All-trans-retinal (all-trans-ral) produced from the photobleaching of rhodopsin, and excessive 11-cis-retinal (11-cis-ral) reversibly react with membrane phosphatidylethanolamine (PE) to form N-trans-R-PE and N-11-cis-retinylidene-PE (N-cis-R-PE), respectively. ABCA4 is flipping these compounds to the cytoplasmic leaflet of disc membrane, preventing the accumulation in disc lumen side and enabling all-trans-retinal to be reduced by RDH8 and enter the visual cycle. N-cis-R-PE is isomerized to its all-trans isomer N-trans-R-PE (Quazi F, Molday RS. Proc Natl Acad Sci U S A. 2014 Apr 1;111(13):5024-9).*

Photoreceptors depend on healthy RPE cells for survival, the degeneration of RPE due to lipofuscin accumulation will lead to subsequent photoreceptor degeneration and vision loss, as experienced by some STGD1 patients.[29] In some cases, however, degeneration of photoreceptors sequentially precedes RPE dysfunction in the earlier stages of disease progression, it appears to be highly associated with certain mutations in *ABCA4*.[33]

STGD1 is a loss-of-function disease, one functional copy of the *ABCA4* gene resulting in only 50% of the protein is sufficient for the relatively normal ABCA4 protein function. Parents of STGD1 patients as obligate heterozygous carriers for *ABCA4* pathogenic variants do not exhibit increased RPE lipofuscin accumulation and do not present with the disease until advanced age.[34, 35]

1.3 Animal models

1.3.1 Knock-out (KO) mouse model

The *Abca4*$^{-/-}$ KO mouse was created as an animal model of Stargardt disease for functional analyses and developing of therapeutic strategies. Mouse retinas resemble more closely human peripheral retina, since they do not contain macula, the cone-rich area for central and color vision in humans. Although *Abca4*$^{-/-}$ mice exhibit some features of human retinopathies including accumulation of A2E and lipofuscin in the RPE, they do not develop the most important feature of Stargardt disease, retinal degeneration, at a young age nor in their lifetime.[36]

The amount of accumulated A2E can be quantified from posterior eye cups in HPLC analysis and used as a measure in phenotype characterization.[36] The accumulation of lipofuscin can also be determined *in vivo* by quantitative autofluorescence (qAF) measurements.[37]

A preclinical study in a *Abca4*$^{-/-}$ KO mouse model of Stargardt disease was successful in utilizing lentiviral based *ABCA4* gene therapy vector. Subretinal delivery of the lentiviral vector carrying a functional copy of human *ABCA4* cDNA to the *Abca4*$^{-/-}$ mouse eyes effectively transduced photoreceptors and RPE cells, and substantially reduced disease-associated A2E accumulation compared to untreated and mock-treated control eyes. One year after the injection, the treated eyes of *Abca4*$^{-/-}$ mice showed A2E accumulation at comparable levels to wt controls, whereas mock-treated or untreated eyes accumulated 3-5 times more A2E.[38]

Other treatment strategies focusing on pharmacologic inhibition of the steps in A2E synthetic pathway have also been successfully tested on *Abca4*$^{-/-}$ mouse. Compounds to inhibit vitamin A, as the precursor for 11-cis-retinal, delivery to the eye, fenretinide and A1120 were shown to inhibit A2E accumulation in the *Abca4*$^{-/-}$ mouse model.[39, 40] Supplementation with C20-D_3-vitamin A, a deuterated form of vitamin A, reduced the A2E accumulation in treated *Abca4*$^{-/-}$ mice through slowing the rate of vitamin A dimerization.[41]

1.3.2 Knock-in (KI) mouse models

Only a small fraction (21/580, 3.6%) of STGD1 patients represent ABCA4 knock-outs harboring 2 nonsense mutations i.e. premature stop mutations, frameshifts or severe splice site mutations. The majority of STGD1 is caused by two missense mutations, and by combinations of missense with nonsense mutations. ABCA4 knock-out model does not ideally reflect the disease progress caused by missense mutations with variable expected residual ABCA4 protein expression and function.

Abca4$^{PV/PV}$ KI mouse model, homozygous for a relatively common founder mutation in STGD1 patients of German descent, the complex allele c.[1622T>C;3113C>T] (p.[L541P;A1038V]),[42] was generated to investigate the pathological role and previously reported misfolding effect of this variant *in vivo*. While mutant ABCA4 RNA levels approximated WT ABCA4 RNA levels, only trace amounts of mutant ABCA4 protein were noted in the retina. Analysis of the two protein variants separately and combined in heterologous mammalian cells, revealed substantial changes and misfolding for P and PV mutants, but only slight changes for the V mutant alone, compared to WT ABCA4. These results correlate with the observed disease phenotypes in STGD1 patients, where the complex allele and p.L541P alone have been associated with severe, and p.A1038V with milder forms of STGD1.[43]

Another KI mouse model for ABCA4 was created for c.2894A>G (p.N965S) variant, the most frequent disease allele in Stargardt-related retinopathies in the Danish population, causing moderate to severe disease phenotypes.[44] p.N965S ABCA4 variant was expressed at 50% level of WT ABCA4 expression, resulted in partial misfolding of the ABCA4 protein with some retention in the endoplasmic reticulum of photoreceptors, and in complete loss of substrate-dependent ATPase activity.[45]

Similar to *Abca4*$^{-/-}$ mice, *Abca4*$^{PV/PV}$ mice and the p.N965S KI mice showed substantial A2E and lipofuscin accumulation in their RPE cells but no retinal degeneration up to 12 months of age.[43, 45]

Besides *in vivo* functional characterization of pathogenic variants of ABCA4, knock-in models for STGD1 will be useful in exploring therapeutic options targeting specific functional defects caused by specific mutations.

1.3.3 Stargardt disease dog model

A naturally occuring retinal degeneration phenotype was described in Labrador retriever dogs. Vision of the affected dogs at the age of 10-12 years was impaired in daylight and dimlight conditions, but they still retained some vision throughout their lifetime. Whole-genome sequencing of the affected sib-pair and their unaffected parents revealed a frameshift insertion c.4176insC in the *ABCA4* gene, leading to a premature stop codon, p.F1393Lfs*1395.[46]

Similar to the clinical features of human STGD1, the accumulation of autofluorescent lipofuscin in the RPE throughout the fundus, bilateral diffuse retinal degeneration and atrophic regions in the RPE were observed in the affected dog. Unlike mouse retina, the dog has a cone rich visual streak, functionally analogous to human macula. Cone function was severely affected, with better preserved rod function, reflecting cone photoreceptors being affected prior to rods, as also seen in STGD1 patients. In addition to the very similar clinical manifestations of *ABCA4*-associated phenotype in dogs compared to humans, the size of the canine eye is also comparable to the human eye.

Canine models have successfully been used for experimental gene therapy for other retinal degenerative diseases, such as retinitis pigmentosa (RP) and Leber's Congenital Amaurosis (LCA).[47-49] The currently only FDA-approved gene therapy vector for RPE65-mediated LCA was developed and tested on a spontaneously occuring RPE65 mutant dog.[50]

The discovered dog homozygous for a *ABCA4* loss-of-function mutation represents the only known large animal model for human STGD1.[46]

1.4 Therapies in development

Currently, there is no approved treatment for the Stargardt disease. Vitamin A supplementation should be avoided in patients with *ABCA4*-associated retinopathies or macular dystrophies associated with lipofuscin accumulation in the RPE, in order to not increase the already progressive bisretinoid accumulation.[40] Investigational therapies for STGD1 include visual cycle modulators (VCMs), complement inhibitors, gene therapy, and human embryonic stem cell therapy for regeneration of the RPE.[25]

1.4.1 Visual cycle modulators

The visual cycle is a series of enzymatic reactions in photoreceptors and RPE, converting all-trans-retinal back to 11-cis-retinal for rhodopsin regeneration. Dysfunction of enzymes in the visual cycle leads to several IRDs, due to the inability to replenish 11-cis-retinal or to remove the accumulation of various retinoid products. The proposed function of ABCA4 is to flip spontaneously forming 11-cis-retinal and all-trans-retinal conjugates with membrane PE, to the cytoplasmic side of photoreceptor outer segment disc membranes. Defects in this function lead to the accumulation of these molecules inside the discs giving rise to formation of toxic lipofuscin bisretinoids like A2E. The purpose of VCMs is to inhibit the accumulation of lipofuscin and ultimately slow down the progression of STGD1.[25]

The cleavage of a C20 carbon–hydrogen bond in the retinaldehyde-PE conjugate is the rate determining step in vitamin A dimerization.[51] Replacing the C20 hydrogen atoms of vitamin A with deuterium atoms (i.e. C20-D_3-vitamin A) makes this bond harder to cleave

and hinders vitamin A dimerization and lipofuscin formation. *Abca4*$^{-/-}$ mice raised on diets containing the deuterated form of vitamin A, exhibited an 80% reduction in A2E, fewer lipofuscin granules in the RPE, and improved eye function.[41] Alkeus Pharmaceuticals is currently conducting a Phase 2 clinical trial, assessing the efficacy and long-term safety and tolerability of **ALK-001** (deuterated form of vitamin A) in the progression of Stargardt disease (ClinicalTrials.gov Identifier: NCT02402660).

Isotretinoin (Accutane), a drug for the treatment of severe acne, has also been shown to impede lipofuscin formation in a mouse model, by inhibiting the visual cycle enzyme 11-cis-RDH, thus slowing the synthesis of 11-cis-retinaldehyde and regeneration of rhodopsin. *Abca*$^{-/-}$ and WT mice, treated with isotretinoin, showed decreased production of A2E and lipofuscin granules in the retina, suggesting that isotretinoin may delay visual loss in STGD1 and other retinal diseases linked to lipofuscin accumulation.[52] However, Isotretinoin is not a valid candidate for treatment of STGD1 due to serious side effects even during short-term treatment and, therefore, is not applicable for long-term administration as required for STGD1.[53]

Emixustat (Acucela Inc) is a retinylamine derivative that inhibits the visual cycle enzyme encoded by the *RPE65* gene, reducing the conversion of all-trans-retinyl ester to 11-cis-retinol and slowing the accumulation of A2E. It was initially developed to potentially slow the progression of geographic atrophy (GA) in age-related macular degeneration (AMD), in which trials there were no statistically significant differences between treatment and placebo groups.[25] Emixustat is currently assessed for safety and efficacy as a potential treatment for STGD1, in a Phase 3 trial (ClinicalTrials.gov Identifier: NCT03772665).

VM200 is explored as aldehyde sequester to decrease excessive levels of all-trans-retinal.[54] VM200 is a primary amine that by forming inactive compound with all-trans-retinal, is blocking the further formation of A2E, and preserving the retinal structure in *Abca4*$^{-/-}$ mice.[55] VM200 is currently in preclinical trials.[25]

Fenretinide (Sirion Therapeutics) and **A1120** are antagonists to retinol binding protein (RBP), competing with vitamin A for binding to RBP, leading to reduced levels of circulating available RBP in the serum, and reduced vitamin A delivery to the RPE. Both compounds showed decreased formation of A2E in *Abca4*$^{-/-}$ mice.[39, 56] Fenretinide is a synthetic vitamin A analogue, and in Phase 2 clinical trial, assessing its efficacy in the treatment of dry AMD patients, several visual adverse events were reported, including night blindness and visual disturbance, with no significant visual benefit in the treatment group.[25] A1120 is a nonretinoid compound, unlikely to cause these visual side effects. Clinical trials for A1120 are under development.[25]

1.4.2 Complement inhibition

A2E and other bisretinoids are shown to have a proinflammatory effect by activation of the complement system in RPE cells, leading to cell death via membrane attack complex (MAC). Complement inhibition with C5 antibody was shown to protect against RPE cell death in a cell culture, after the cells were exposed to bisretinoid-induced complement pathway activation.[57] **Zimura** (Ophthotech Corporation) is a C5 complement inhibitor to prevent the formation of MAC, and consequently reduce cell death. It is delivered by intravitreal injection and is being explored as a potential treatment option, besides STGD1 (Phase 2, NCT03364153), for a number of eye diseases, including both forms of AMD.[25]

1.4.3 Gene supplementation therapy

The eye is an ideal organ for gene therapy because of its small size, isolated and easily accessible location, immune-privileged status due to blood-retina barrier, and the availability of imaging techniques to directly visualize the retinal cell layers and observe structural phenotypic changes. The contralateral eye can be left untreated and serve as a control. The goal of gene therapy for patients with Stargardt disease is to introduce a functional *ABCA4* gene to the retina, to restore normal ABCA4 protein function and to prevent disease progression. Virus-based vectors and non-viral carriers have been recently developed to deliver larger genes and efficiently transduce photoreceptor cells. Gene delivery to the cells of retina is accomplished via subretinal injection, delivering the vector between the photoreceptors and the RPE.[25]

The *ABCA4* cDNA (6.8 kb) exceeds the 4.5–5.0 kb capacity of adeno-associated virus (AAV) vectors that have proved most successful for other ocular genes, leading to utilization of an equine infectious anemia lentivirus (EIAV) for gene transfer. Subretinal injection of EIAV-*ABCA4* effectively transduced photoreceptors and RPE cells in a knock-out *Abca4*$^{-/-}$ mouse model.[38] The safety and distribution of EIAV-*ABCA4* was further tested in larger animals, in rabbits and macaques. Both species presented with slightly higher but localized and transient intraocular inflammation in eyes receiving the subretinal injection, with no systemic or ocular tissue toxicity.[58] The clinical Phase 1/2 trials (ClinicalTrials.gov Identifier: NCT01367444 and NCT01736592) are ongoing for 46 patients with Stargardt disease. The trials investigate short- and long-term safety and tolerability of escalating doses of SAR422459 (Sanofi; the lentiviral-based vector carrying *ABCA4* cDNA).

To overcome the insert size limitations of AAV vectors, a dual AAV strategy to deliver the *ABCA4* gene has been proposed, where the gene is split between two vectors, with overlapping sequences facilitating recombination and generation of the full-length *ABCA4* gene in the host cell. The expression of full-length ABCA4 protein in the photoreceptor outer segments of *Abca4*$^{-/-}$ mice, sufficient to reduce A2E accumulation, was demonstrated.[59]

Methods of non-viral gene supplementation involving carriers such as compacted DNA nanoparticles, naked DNA, and DNA-encapsulating liposomes have also been investigated. Non-viral gene delivery would reduce the risk of immune responses associated with viral vectors but, in general, would not provide long-term ABCA4 expression and, therefore, would require repeated administration. Compacted DNA nanoparticles formulated with polyethylene glycol-substituted polylysine (CK30PEG) are highly efficient in transfecting postmitotic cells with large capacity, and exhibit minimal toxicity even after repeated dosing. Subretinal delivery of *ABCA4* cDNA via compacted nanoparticles, to the eyes of STGD1 mouse model, resulted in *ABCA4* transgene expression for up to 8 months, and reduced lipofuscin granules on histopathologic evaluation.[60]

Successful transgene deliveries with phenotype rescue in mouse models merit further evaluation of these methods as viable options for gene therapy in STGD1 patients. Of all suggested treatment options, gene therapy would be an ideal treatment for STGD1 as a recessive loss-of-function disease. Even partial, up to one half, restoration of the ABCA4 protein function in photoreceptor cells would theoretically cure the disease, since carriers of *ABCA4* mutations (for example, parents of STGD1 patients) have no visual symptoms. However, the gene therapy treatment should be administered at early stages of the disease when most photoreceptor cells are still alive.

1.4.4 Stem-cell therapy

Stem-cell therapy harbors considerable potential in treatment of Stargardt disease, to replace the already lost RPE and photoreceptor cells with cells that carry functional copies of *ABCA4* gene. Like with gene therapy, the eye offers a relatively uncomplicated candidate for stem cell manipulations because of the immunological isolation, easy access, possibility to directly nonivasively visualize the transplanted cells, and because of the small size.

Preclinical studies of human embryonic stem cells (hESCs) derived RPE cells transplanted into the subretinal space of mouse models of retinal degeneration confirmed successful integration of grafted cells into the RPE, without evidence of tumor formation or spread to other areas, and with mice demonstrating improved visual function.[61]

Phase 1/2 clinical trials were developed to test safety and preliminary signs of efficacy of subretinal transplantation of hESC-derived RPE cells in patients with STGD1 and AMD (NCT01345006/NCT01469832). No adverse events were reported related to the stem-cell therapy, although some patients experienced complications probably related to the surgical procedure or the immunosuppressive regimen. Visual outcomes remained inconclusive due to the small sample size, relatively short follow-up time of 1 year, and lack of a control group. A follow-up study (NCT02445612) is assessing long-term safety and tolerability following the same cohort of patients up to 5 years after stem cell implantation.[62]

As an alternative to hESC-derived allograft transplants and life long immunosuppressive treatment, patient-derived induced pluripotent stem cells (iPSC) can be differentiated into RPE cells to be used as autograft transplant with no projected immune rejection. Human iPSC-derived RPE cells were injected into subretinal space of RP mouse model, where the graft assimilated into the host retina without disruption, leading to improved visual function in recipients, and supporting the feasibility of autologous iPS cell transplantation.[63]

The overarching problem with transplanting RPE cells, in addition to technical difficulties, is that STGD1 is a disease of photoreceptors. Transplanting RPE, even if it integrates and functions, could provide some support to photoreceptors, but would not target the primary diseased cells. However, generating differentiated functional mature photoreceptors from stem cells remains still challenging.[64] Most recent stem cell-based clinical trials for AMD are transplanting pre-grown batches (’sandwiches’) of RPE/photoreceptor layers, which could prove scientifically and technically more feasible. In summary, while theoretically very intriguing, stem cell therapies of retinal diseases still need extensive development before these would expect to have substantial effect on inherited retinal diseases.

2 THE *ABCA4* LOCUS

The *ABCA4* gene was identified in 1997 by Allikmets et al as the causal gene for Stargardt disease.[1] The coding sequence of 6822 basepairs in 50 exons encodes for a ATP-binding cassette (ABC) superfamily transmembrane protein expressed in photoreceptors and retinal pigment epithelium.[28, 65]

2.1 Analysis of the *ABCA4* coding sequences

First *ABCA4* gene analyses in Stargardt disease patient and control cohorts were carried out by a combination of indirect DNA variant detection methods using single strand conformation polymorphism (SSCP) analysis, heteroduplex analysis (HA), denaturing gradient gel electrophoresis (DGGE), and denaturing high-performance liquid chromatography (dHPLC), followed by direct Sanger sequencing of the amplicons with altered mobility patterns in those assays.[42, 66-68] Identification of two disease alleles per patient is expected for a recessive disorder, applications of indirect DNA variant detection methods achieved detection of 30-60% of all possible disease-associated alleles in patient cohorts, reported with extraordinary high allelic heterogeneity and carriers of pathogenic variants in control groups. Direct sequencing of the *ABCA4* gene in patients enabled 70% of disease-associated alleles to be identified.[69]

Besides Stargardt disease, mutations in the *ABCA4* gene were also associated with clinically more severe retinal dystrophies autosomal recessive cone-rod dystrophy (arCRD) and autosomal recessive retinitis pigmentosa (arRP).[4-6, 70] Approximately 80% of patients had onset of visual symptoms before 20 years of age, but disease onset in 5th and 6th decade were also reported.[71]

The c.2588G>C (p.[G863A,G863del]) variant was identified in 37% of STGD1 patients of Dutch and German origin, along with 2.9% carrier frequency in control individuals from the Dutch population.[67] The second allele in compound heterozygote STGD1 patients with p.[G863A,G863del] variant was more often a null allele (frameshift, premature stop, splice site mutation), and no homozygous STGD patients for the p.[G863A,G863del] variant were observed. This led to the conclusion that mild *ABCA4* variants in both alleles do not result in STGD1 disease phenotype, and the c.2588G>C (p.[G863A,G863del]) variant was categorized as a mild frequent founder mutation in the Western European population.[67]

Classifying *ABCA4* variants as 'mild', 'moderate', and 'severe' based on the predicted effect of the mutation on the ABCA4 protein function suggested a model for the *ABCA4*-associated disease, which correlated the variability in retinal disease phenotypes and age of onset with predicted residual ABCA4 protein function of an individual mutation.[66, 67, 72] Accordingly, the more severe the effect of a mutation on ABCA4 protein, the more aggressive and earlier onset the disease phenotype. Several family reports of early onset STGD1 or CRD associated with *ABCA4* null mutations, and analysis of a childhood-onset STGD1 patient cohort revealing a higher proportion of definitely or possibly deleterious *ABCA4* variants compared to adult-onset STGD1 cohort support this model.[73 74, 75]

With high carrier frequency of a possibly pathogenic variant in 1/20 people across populations, variation in the *ABCA4* locus emerged as the most prevalent cause of inherited retinal dystrophies.[67, 76]

2.1.1 *ABCA4* gene and age-related macular degeneration (AMD)

The observation that combinations of milder *ABCA4* alleles are associated with later ages of disease onset up to the age range for the diagnosis of age-related macular degeneration (AMD) in 6th-7th decade, and some phenotypic similarities between STGD1 and AMD, including the accumulation of pigmented substance in the RPE and progressive macular atrophy suggested *ABCA4* gene as a candidate gene for AMD.[77]

Indeed, c.5882G>A (p.G1961E) variant was identified with statistically significantly higher frequency in AMD patients than in age-matched controls.[78] Since c.5882G>A (p.G1961E) was also reported as the most frequent mutation (10% of all ABCA4 mutations identified) in STGD1 cohort from North America[71], and the clinical evaluation of STGD1 family members revealed a high frequency of AMD in first and second-degree relatives, it was hypothesized that some heterozygous *ABCA4* mutations may enhance susceptibility to AMD, and monoallelic family members of STGD1 patients may be at increased risk.[79, 80]

This hypothesis was disputed by other studies[81], but the prevalence and role of heterozygous ABCA4 alleles in AMD remains to be unequivocally determined. It has also been suggested that many of AMD cases are actually misdiagnosed late-onset STGD1 patients.[82]

The two major susceptibility loci for AMD - CFH and ARMS2 were reported in 2005.[83, 84]

2.2 *ABCA4* genotyping microarray

Allelic heterogeneity in *ABCA4*-associated disorders has substantially complicated genetic analyses and molecular diagnosis of patients. *ABCA4* gene presents an extremely challenging target for diagnostic applications due to the size (50 exons) and vast allelic heterogeneity. Mutations have been reported in all 50 exons, every screening of a considerable STGD1 patient cohort adds novel rare variants with unknown effect, and the three most common mutations (p.G1961E, p.[G863A,G863del], and p.A1038V) have each been described in only ~10% of STGD1 patients in a distinct populations.

The application of conventional indirect variant detection methods (SSCP, HA, DGGE, dHPLC) in combination with direct sequencing is labor-intensive and relatively inefficient. Due to technical limitations, the sensitivity and mutation detection rate of the different indirect variant detection methods is not uniform for all analyzed amplicons, potentially resulting in false-negative data.[42, 68]

To overcome these limitations and to generate a high-throughput, cost-effective screening tool we designed, in collaboration with Asper Biotech, Ltd (Tartu, Estonia), a genotyping microarray (gene chip) for the *ABCA4* gene, which included all reported disease-associated variants (around 400 in 2003), enabling fast and simultaneous screening for all known *ABCA4* variants.[85] The *ABCA4* gene chip utilized arrayed primer extension (APEX) technology – a rapid solid-phase genotyping method combining the efficiency of a microarray-based assay with Sanger sequencing.[86] The array was validated on an extensive cohort of 136 confirmed STGD1 patient samples, previously screened by SSCP and/or heteroduplex analyses.[71] The microarray screening detected numerous additional alleles, that went undetected in the previous SSCP/HA assays, bringing the mutation detection rate from 55% to 70% of all possible disease-associated alleles.[85]

In further evaluation of the *ABCA4* gene array, by screening several previously not analyzed STGD1 patient cohorts of different European ethnic groups 53%-60% of all possible disease-associated ABCA4 alleles were identified, similar to the best results

employing indirect mutation detection technics. By design, the *ABCA4* gene array tested for only known, previously reported variants, identifying most disease-associated *ABCA4* alleles in STGD1 patient cohorts of same or close ethnicity as the cohort where the initial *ABCA4* variant data was obtained from. Conversely, lower scores of detected disease-associated *ABCA4* alleles were expected in screening previously not characterized cohorts.[85] The *ABCA4* gene array was continuously updated for screening the ever growing number of new reported *ABCA4* pathogenic variants. The final version of the array screened for 800 disease-associated or rare *ABCA4* variants.

On average, the *ABCA4* gene array identified 1-2 pathogenic *ABCA4* variants in ~75% of screened ABCA4-related disease cases of European descent.[44, 85, 87] Patients, for whom two *ABCA4* disease-associated variants were identified, were considered 'solved', while patients with one or no *ABCA4* mutations detected by the array were subjects of further *ABCA4* gene analysis by dHPLC or DGGE and subsequent sequencing to elucidate the previously unreported disease-causing variants.[88, 89]

ABCA4 gene array served for scientific and diagnostic purposes as effective, high-through-put, accurate, and affordable first pass screening method until the development of next-generation sequencing technologies.

2.3 Interpretation of sequence variants

A growing number of variants are continuously being discovered in the human genome, including in the *ABCA4* locus. In inherited Mendelian disorders, determining if a DNA variant is disease-causing or benign, is based on several types of evidence: variant frequency in patient *vs* control populations, functional studies, computational prediction, and segregation of the variant with the disease phenotype in families. Based on the combined evidence, clinical significance of any given sequence variant falls along a gradient, ranging from almost certainly pathogenic to almost certainly benign.

The American College of Medical Genetics and Genomics (ACMG) set guidelines for the interpretation of sequence variants in clinical diagnostics, recommending classifying the variants as 'pathogenic', 'likely pathogenic', 'uncertain significance', 'likely benign', and 'benign', based on evaluating and combining the weight of each evidence datapoint. According to the guidelines, population frequency data and functional studies provide generally stronger evidence, while segregation data and *in silico* prediction are considered supporting evidence. There are limitations and caveats to every evidence category. For instance, with very rare variants, case-control comparisons may not reach statistical significance due to low power, functional assays may not reflect the biological environment perfectly, computational prediction from different algorithms may be contradicting, and family members are not always available for segregation analysis. With conflicting evidence between benign and pathogenic, the variant is classified as variant of uncertain significance.[90]

To avoid misinterpretation, careful expert judgment must be applied when assessing the combination of exisiting evidence for a given variant. Continually expanding knowledge on variant frequencies in larger populations with genetically different backgrounds, and on disease associations is often leading to re-classification of some variants.

2.3.1 Variant analysis by *in silico* predictions

Many computational tools predicting the effect of sequence variants are available to assist in their interpretation. Use of this type of evidence has become a routine part in annotation pipelines for variant assessment in gene-focused projects, or in whole exome

and whole genome sequencing. The two main categories of such *in silico* tools include those that predict whether a missense change is damaging to the resultant protein function or structure, and those that predict a variant's potential to affect splicing. Of note, synonymous and missense variants can also affect splicing by creating new competing splice sites or reducing the original splice sites in surrounding sequences.

The algorithms for predicting the impact of a missense change are based on one or several of the following criteria: evolutionary conservation of an amino acid or nucleotide, the location and context within the protein sequence, and the biochemical consequence of the amino acid substitution. Newer meta-predictors have incorporated many previous algorithms.[90, 91]

The available prediction programs have been evaluated and compared to each other by assessing their ability to predict "known" disease-causing variants. 18 prediction algorithms were compared on 14,819 benign or pathogenic missense variants from the ClinVar database (aggregated genomic variant data in relation to human health), with wide variability in concordance among different combinations of algorithms, and with particularly low concordance for benign variants. Only 3.2% of the benign and 41.5% of pathogenic variants had concordant assertions across all 18 algorithms. Predictions from five more commonly used algorithms (Polyphen, SIFT, CADD, PROVEAN and MutationTaster) resulted in higher concordance, with 79% concordance for pathogenic variants and only 33% for benign variants.[91]

There is little consensus among clinical labs on how many and which algorithms to use for missense variant interpretation. Different combinations of algorithms lead to discordant interpretations between different labs. Some studies have adhered to strict concordance guideline, others have used a majority vote rule between applied algorithms. The guidelines for variant interpretation recommend using multiple programs since different programs have their own strengths and weaknesses depending on the algorithm and in many cases performance can also vary by the gene and protein sequence. However, more algorithms lead to more discordant calls between them. An alternative is to use newer metapredictors (*e.g.* REVEL, VEST3, MetaLR, MetaSVM, Condel, Mcap, Eigen, and CADD) which combine multiple individual predictors to generate a score, and have shown high overall performance characteristics.[91]

The five most common tools for splice site evaluation are NNSplice, MaxEntScan, GeneSplicer, HumanSplicingFinder, and SpliceSiteFinder-like, because of their user-friendly and high throughput access provided by the bioinformatics tool Alamut. Their splice site detection methods are based on algorithms using position weight matrices, binary string artificial neural network, maximum entropy distribution, and maximal dependence decomposition.[92] When compared with 20,000 known splice sites and 20,000 false splice sites, in terms of accuracy, sensitivity, specificity, and precision, GeneSplicer provided the best overall performance, and the position weight matrices based methods, HumanSplicingFinder and SpliceSiteFinder-like, scored the lowest on donor regions due to an extremely high false positive rate (http://blog.goldenhelix.com/nfortier/splice-site-algorithms/).

According to Alamut combined assessment, c.5461-10T>C and c.5714+5G>A, the two most frequent intronic disease-associated *ABCA4* variants, are predicted to reduce their respective acceptor and donor sites by 4.3% and 33.4%, respectively. *In vitro* RNA analysis from photoreceptor progenitor cells (PPC) and 'midigenes' revealed a complete, 100% exon skipping for the c.5461-10T>C variant, and 40% of correctly spliced mRNA for the c.5714+5G>A variant.[93, 94]

The combination of predictions from different *in silico* tools is considered as a single piece of evidence in sequence interpretation, because of overlapping criteria in their algorithms. *In silico* predictions should be evaluated with caution, and be used only in combination with other evidence to make a final clinical assertion.

The possible effect of all identified *ABCA4* variants was assessed using following *in silico* prediction programs: SIFT,[95] Polyphen2,[96] Align-GVGD,[97] MutationTaster,[98] SpliceSiteFinder,[99] MaxEntScan,[100] NNSPLICE,[101] GeneSplicer,[102] and Human Splicing Finder.[103] All of these algorithms except for Polyphen2 were accessed via Alamut software (http://www.interactive-biosoftware.com).

2.3.2 Functional analysis of *ABCA4* missense variants

While variants introducing frameshifts, premature stop codons, and changes to canonical splice sites can be generally regarded as knock-out alleles resulting in no functional ABCA4 protein as their consequence, the majority of disease-associated variants in *ABCA4* are missense variants with uncertain effect and possibly very variable influence on ABCA4 protein function. Aside from the certain population- or ethnic group-specific disease alleles, which account for 20-30% of the disease load in STGD1 patients from these populations, most of the disease-associated *ABCA4* variants are rare in both patients and in the general population. Therefore, it is often impossible to determine the pathogenicity of a variant based on differences in allele frequencies between patients and controls in the modest size patient cohorts in clinics. The established correlation from genotype-phenotype analyses, between the disease severity and residual ABCA4 protein function, also presents a challenge to determine the actual pathogenic effect of different ABCA4 missense variants on the protein function.[104]

Despite the *ABCA4* gene mutations being the most frequent cause in Mendelian macular dystrophies, knowledge of the biochemical, structural, and functional properties of this transmembrane protein remains limited, due to its large size, and rapid loss of its stability and activity after removal from native rod outer segment (ROS) disk membranes. Stimulation of the ATPase activity of ABCA4, in the presence or absence of all-*trans*-retinal and *N*-retinylidene-PE as potential substrates, served as the only available functional assay, until a transport assay was developed measuring the increased transfer of retinal from donor to acceptor liposome vesicles, coupled with the rate of flipping the N-retinylidene-PE in donor vesicles from the inside to the outside of the membrane bilayer by ABCA4.[105, 106]

To characterize the biochemical defects of different ABCA4 protein variants, their expression level, cellular localizations, ATP binding, retinoid binding, and basal and substrate-stimulated ATPase activities have been determined from analyses in transiently transfected COS7 and HEK293 cell cultures, expressing human mutant and WT ABCA4 cDNA clones. The expression level of ABCA4 variant proteins in cell cultures was assayed by immunoblotting the cell lysates for ABCA4, and efficiency of ATP binding was measured by photoaffinity labelling with α-32P azido-ATP. To examine the effect of variants on the ATP hydrolysis, immunopurified ABCA4 protein variants were reconstituted into lipid membranes, and tested for basal and retinal-stimulated ATPase activities, compared to WT ABCA4. The expression pattern and localization of ABCA4 variant proteins in cell cultures was visualized by immunofluorescence microscopy. To date, functional defects of ~60 different ABCA4 missense variants have been described, demonstrating considerable variation in these assays.[104-109]

ABCA4 protein expression in COS7 and HEK293 cells, transfected with WT human *ABCA4* cDNA, appears partially in intracellular vesicle-like structures, but to a variable

extent between different experiments also in endoplasmic reticulum (ER), a localization typically seen for incorrectly folded membrane proteins.[104, 109] However, ABCA4 purified from transiently expressing HEK293 cells, compared to ABCA4 prepared from bovine rod outer segments, exhibited nearly identical profiles of all-trans-retinal stimulated ATPase activity, indicating of a fully functional form of ABCA4 protein also in cell culture.[104] In general, all analyzed ABCA4 protein variants showed either WT-like localization to vesicles and ER, or primarily to ER, indicative of some protein misfolding and retention.[108]

To overcome the limitations and inconsistency of a transient expression system, four missense variants, associated with more severe early-onset STGD1 phenotype, the complex allele p.[L541P;A1038V], as well as p.L541P and p.A1038V individually, p.R602W, and p.C1490Y, were analyzed in the photoreceptors of 14 days old transgenic Xenopus laevis tadpoles by immunofluorescent labeling. All variants except for p.A1038V, caused retention of the mutant ABCA4 protein in the photoreceptor inner segment and cell bodies, likely by impairing correct protein folding. p.A1038V variant exhibited correct localization to ROS like WT ABCA4.[109] Knock-in mouse model for p.[L541P;A1038V] corroborated these results, only trace amounts of mutant ABCA4 protein were noted in the outer segments.[42]

Immunofluorescence of the retina of another knock-in mouse for the p.N965S variant, revealed partial (~40%) retention of the mutant ABCA4 protein to the photoreceptor inner segments, likely due to protein misfolding. ATPase activity of p.N965S ABCA4 protein purified from mouse retinal membranes measured 50% of WT basal ATPase activity, but was insensitive to substrate all-trans-retinal addition.[45]

Experiments with transgenic animals demonstrating severe cellular localization defects of several ABCA4 protein variants *in vivo*, indicated that functional characterization of *ABCA4* mutant alleles assessing only ATP activity may be inadequate.[109]

Two of the most frequent disease-associated ABCA4 missense variants in European STGD1 patients, p.G1961E and p.A1038V, showed expression levels and cellular localization patterns similar to WT ABCA4, while the expression of p.[G863A,G863del] was substantially reduced, especially for the p.G863A protein variant. The p.A1038V variant displayed somewhat reduced basal and retinal-stimulated ATPase activity compared to WT ABCA4 protein, consistent with the milder disease phenotype associated with this variant. On the contrary, p.G1961E and p.[G863A,G863del], showed drastically reduced basal ATPase activity with little or no substrate stimulation, completely disagreeing with the mild, late-onset retinal disease phenotype associated with these ABCA4 variants.[32, 104, 108]

Two ABCA4 missense variants, p.T901A and p.R943Q, considered benign based on no statistical differences in allele frequencies between patients and the matched general population, displayed poor expression level with ER retention and 100-fold decrease in binding affinity for 11-cis-retinal, respectively.[32, 105]

Overall, biochemical analyses were consistent with assessed ABCA4 missense variants associated with more severe retinal disease phenotypes, whereas discrepant results between *in vitro* functional tests and actual phenotypes in STGD1 patients were observed when analyzing more frequent and milder missense variants.

2.3.3 *In vitro* analysis of *ABCA4* canonical and non-canonical splice site variants

To assess the effect of variants in canonical and non-canonical splice sites, 'minigene' constructs are generated using commercial exon trapping vectors to clone the PCR-amplified regions encompassing the potential splice affecting variants of interest and the adjacent exon(s) from the patient's genomic DNA. RNA is isolated from

transfected cells (COS7, HEK293) and splicing is analyzed by RT-PCR and sequencing from both, WT and variant clones.[42, 110]

At first, the most frequent intronic disease-associated ABCA4 variant c.5461-10T>C in intron 38, did not reveal any splice defect in minigene analysis, while three other analyzed intronic variants c.3050+5G>A, c.4253+5G>A, c.5714+5G>A resulted in complete or partial missplicing. The c.5461-10T>C variant was classified as pathogenic with unknown effect because of other strong evidence for its pathogenicity.[42] The c.5461-10T>C variant is always segregating with the disease phenotype in families, is very rare in the general population (MAF<0.001) and no other possibly disease-associated *ABCA4* variants have been identified on the same chromosome.[12]

In order to determine the effect of c.5461-10T>C in the most biologically relevant test environment, *ABCA4* mRNA was analyzed from photoreceptor progenitor cells (PPCs), derived from c.5461-10T>C harboring patients' fibroblasts via induced pluripotent stem cells (iPSC). Patient-derived PPCs showed complete skipping of exon 39 or exons 39 and 40 in the mRNA, correlating with the severe early onset cone–rod dystrophy phenotype in STGD1 patients homozygous for the c.5461-10T>C variant.[93]

When using small minigenes, *in vitro* results did not completely correspond with splice defects observed in patient-derived cell cultures, and partially abnormal splicing was registered with clones carrying ABCA4 WT sequences as well.[93] A novel, 'midigene' approach was employed, using bacterial artificial chromosomes as splice vectors, capable of carrying longer inserts including both flanking exons around the variant of interest, thereby, providing more of the natural genomic context and producing more accurate results. Analysis of all 44 reported *ABCA4* non-canonical splice site variants in midigenes from HEK293 cells, revealed skipping of multiple exons, as well as exon elongation and intron retention as aberrant splicing events. Based on the residual fraction of correctly spliced mRNA, the non-canonical splice site variants were classified as 'severe', 'moderate', and 'mild'. From phenotype correlations, it was concluded that variants resulting with >80% correctly spliced mRNA are likely benign and not contributing to the STGD1 disease phenotype.[94]

2.4 Analysis of the *ABCA4* gene by next-generation sequencing

Next-generation sequencing (NGS) enables high-throughput massively parallel sequencing of large numbers of samples and is the current routine method for determining genetic variation in patient cohorts and control groups, in research as well as in clinical diagnostics.

The first NGS analysis of the *ABCA4* gene applied a targeted next-generation sequencing protocol (Roche 454; 454 Life Sciences, Branford, CT) to find all possible disease-associated variants in coding sequences of the *ABCA4* gene in a large cohort of 168 patients diagnosed with *ABCA4*-associated diseases, that remained 'unsolved' after screening with the *ABCA4* microarray. 1 out of 2 expected *ABCA4* mutations had been identified in 111 patients and 0 mutations in 57 patients. In NGS, the second disease-associated allele was identified in 49 of 103 (~48%) of patients with one previously reported mutation. Of the 56 samples with no identified mutations, NGS detected both disease-associated alleles in 4 patients (7.1%) and 1 mutation in 10 patients (17.9%). 57 previously unknown, possibly pathogenic, variants were identified, and 55 of them were reported only in one sample.[12]

Determining the pathogenicity of rare missense (and also splicing-affecting and synonymous) variants presents a significant challenge, especially for the *ABCA4* gene,

where now close to 1000 possibly disease-associated variants have been identified.[11] Since more than half of these have been detected only once, the unequivocal classification for disease-association, which is usually accomplished by segregation and functional analyses, presents an exhausting task. Since *ABCA4*-associated diseases are recessive, any given patient often represents the only affected individual in a family, frequently with no siblings. Tissue samples from patients for direct analysis of *ABCA4* RNA are not possible due to the photoreceptor-specific gene expression. Many frequent ABCA4 protein variants have been analyzed in indirect assays, such as their effect on the protein yield, folding, and ATP-binding and ATPase activity assays in *in vitro* systems[69, 104] and *X. laevis* models.[109] Performing these experiments for hundreds of rare variants is unrealistic, leaving the variant interpretation mostly based on combining evidence from variant frequency comparisons between patients and controls, *in silico* pathogenicity prediction, and family segregation analyses where possible.

Even after complete sequencing of the *ABCA4* coding region in patients with definitive clinical diagnosis of STGD1, approximately 15-20% of patients generally remain 'unsolved' with one identified pathogenic *ABCA4* variant, and in ~15% patients no pathogenic *ABCA4* variants are found. The possible explanations for this include:

1) large deletions or insertions of exons and chromosomal segments, i.e. copy number variants (CNVs), although rare in the *ABCA4* locus, escape detection by PCR-based methods[111],

2) significant fraction of disease-causing variants locate outside the *ABCA4* coding sequences,

3) patients with no *ABCA4* disease-associated variants identified represent phenocopies of *ABCA4*-associated disease, with causal genes other than *ABCA4.*

CNVs are very rare in the *ABCA4* locus (~1% of all patients), however, for complete mutational scanning, a CNV analysis with a comparative genomic hybridization array (aCGH), or with Multiplex Ligation-dependent Probe Amplification (MLPA), could be included.

Many pathogenic mutations are likely located outside of the *ABCA4* coding sequences since the second mutation, required for complete genetic diagnosis of STGD1, was not found in about half of the STGD1 patients with 1 mutation.[12] Although some of these patients could be carrying the *ABCA4* variant by chance due to high population frequency of *ABCA4* variants (estimated at 1:20), it is highly unlikely that a patient clinically diagnosed with STGD1 and carrying one confirmed *ABCA4* mutation does not have the 2nd pathogenic variant in the *ABCA4* locus. Detection of disease-associated variants outside of the *ABCA4* coding sequences requires sequencing of the entire 130kb *ABCA4* genomic locus, more importantly in patients with 1 identified mutation from sequencing of the *ABCA4* coding regions and splice sites.

Finally, in cases with no disease-associated variants in *ABCA4,* whole-exome or –genome sequencing approaches can be applied to determine the genes and mutations causing STGD1-mimicking phenotypes. The fraction of phenocopies (i.e., clinically misdiagnosed patients) depends primarily on the depth of clinical analyses performed at a given eye clinic. Although, even exceptionally extensive clinical data are often not enough for pinpointing the possible genetic cause. Depending on the severity of *ABCA4* mutations and the stage of the disease progress at which a patient is diagnosed, *ABCA4*-associated pathology presents in a wide range of phenotypes from mild Fundus Flavimaculatus to CRD and even RP-like phenotypes. The latter two phenotypes are caused by tens of other distinct genes.

Given the substantially overlapping phenotypes and several treatment options currently in late stages of preclinical development or in clinical trials, the correct and comprehensive molecular diagnosis of *ABCA4*-associated diseases is crucial. The NGS technology offers a reliable time- and cost-efficient platform to analyze large and variable genes in large cohorts. Expanding knowledge on the *ABCA4* gene variation in large patient cohorts allows, together with detailed clinical analysis, improved genotype-phenotype correlations facilitating more accurate disease prognosis and advancing the overall understanding of the *ABCA4* disease etiology.

2.4.1 Leiden Open Variant Database (LOVD)

In an effort to register all described *ABCA4* variants into one database, information of all *ABCA4* variants published before 2016 that were associated with STGD1, arCRD and arRP was gathered with the description of the phenotype, including age, phenotype of onset, segregation information, and uploaded onto the Leiden Open source Variant Database (LOVD; https://databases.lovd.nl/shared/genes/ABCA4). The pathogenicity of all variants was assessed as follows: 1) based on their predicted functional effect for stop-gain and frameshift variants; 2) based on *in silico* predictions for synonymous and non-synonymous variants, and canonical and non-canonical splice variants; 3) based on variant frequency comparison in the collected ABCA4 disease patient dataset versus non-Finnish European general population (The Exome Aggregation Consortium; ExAC). In addition, variants were analyzed based on their homozygous occurrence combined with age of disease onset, their presence in patients with early onset, and their occurrence with a more frequent mild variant. Based on this information and the American College of Medical Genetics and Genomics (ACMG) guidelines, all variants in the ABCA4-LOVD database were classified according to their pathogenicity as either (1) pathogenic, (2) likely pathogenic, (3) uncertain significance, (4) likely benign or (5) benign.[11]

In total, 5,962 *ABCA4* alleles were reported in 3,928 patients, collected from 148 papers. Of the 913 different reported *ABCA4* variants, the majority comprised missense variants (51.6%), followed by protein-truncating variants (34.4%). 316 variants were found to be significantly enriched in the patient cohort. Five variants - p.[G863A,G863del], p.G1961E, p.A1038V, c.5714+5G>A, and p.R2030Q were determined to have a mild pathogenic effect in most cases based on their low or absent homozygous occurrence, while having a relatively high allele frequency in patients.

Whereas ABCA4 LOVD combines all published *ABCA4* variants in STGD1 patients, there are several caveats to the database that should not be ignored. Both, the diagnostic methods and variant screening technologies have improved tremendously over time, so patients have been diagnosed using different criteria, and screened for variants using different technologies, causing bias in the data. Variants considered benign and/or in linkage disequilibrium with the pathogenic variant are often not reported, and conversely, several reported pathogenic variants from analyzing small patient cohorts were classified as benign in the analysis of the larger collective data. Sharing patient genetic variation data will increase statistical power and would allow classifying of additional pathogenic and benign variants. Recently, scientific journals are requiring the uploading of disease-associated variants in publicly available databases.[11]

Currently, mostly STGD1 patients of European origin have been analyzed, and often the ethnicity of patients is not described in publications. Collecting accurate data of patients of all ethnicities will give a better understanding of the ABCA4 disease worldwide.

2.4.2 Databases for genetic variation in general populations

In addition to enabling sequencing of large patient cohorts, another major outcome of NGS technologies is the creation of large sets of detailed human genetic variation data in different populations for free use for biomedical community as reference in disease-specific and population genetic studies.

The Exome Variant Server (EVS, http://evs.gs.washington.edu/EVS/), including whole-exome data of 2203 African American and 4300 European American unrelated individuals, totaling 6503 samples, became available in 2011. Sequencing data of 1092 genomes of various ethnic groups was released in 2012, as a result of the 1000 Genomes Project.[112] The Exome Aggregation Consortium (ExAC) presented exome sequencing data of 60,706 unrelated individuals sequenced as part of various disease-specific and population genetic studies, in 2014.[113] The most comprehensive database, the Genome Aggregation Database (gnomAD; https://gnomad.broadinstitute.org/), currently includes 125,748 exome sequences and 15,708 whole-genome sequences from unrelated individuals.

However, general population databases can contain late-onset STGD1 patients, who were included as healthy controls before presenting with symptoms.

2.4.3 Genetic and clinical analysis of *ABCA4*-associated disease in African American patients (Publication I)

The many different disease-associated variants in the *ABCA4* gene in Stargardt disease (STGD1) are often specific to racial and ethnic groups. The c.2588G>C (p.[G863A,G863del]) variant was described as the founder mutation in Northern European patients,[67] the c.[1622T>C; 3113C>T] (p.[L541P; A1038V]) complex allele in patients of mostly German origin,[6, 42] the c.3386G>T (p.R1129L) variant as the most prevalent STGD1 mutation in Spain,[87] the c.2894A>G (p.N965S) variant in the Danish STGD1 cohort[44], and the c.5318C>T (p.A1773V) variant in Mexico. [114]

Proper genetic diagnosis and interpretation of *ABCA4* alleles in ethnic and racial groups relies on comprehensive knowledge of disease-associated variants in those groups. We investigated *ABCA4* variation and phenotypic expression in a cohort of 44 STGD1 patients of African American descent, a previously under-characterized racial group, where only a small case series had been described.[115]

Patients diagnosed with STGD1 were classified into one of the four disease progress stages as described previously,[2] and screened for mutations in *ABCA4* by NGS, using Illumina TruSeq Custom Amplicon protocol on MiSeq sequencing platform (Illumina, San Diego, CA), and by custom array comparative genome hybridization (aCGH) for large indels.

The possible effect of all identified *ABCA4* variants was assessed using following *in silico* prediction programs: SIFT,[95] Polyphen-2,[96] Align-GVGD,[97] MutationTaster,[98] SpliceSiteFinder-like,[99] MaxEntScan,[100] NNSPLICE,[101] GeneSplicer,[102] and Human Splicing Finder,[103] accessed via Alamut software (http://www.interactive-biosoftware.com). Segregation of the new variants with the disease was analyzed in families if family members were available (Publication I, Table 1). The allele frequencies of all variants were compared to the Exome Variant Server (EVS) dataset of African American and European American general populations.

Analysis of the African American STGD1 cohort

Genetic analysis with a combination of ABCA4 array screening and/or direct sequencing identified both *ABCA4* disease-causing alleles in 27/44 (61.4%) of African American patients, one disease-causing allele in 11 (25%) patients, and no *ABCA4* mutations were

found in 6 (13.6%) patients. Altogether, 65 (73.9%) disease-causing *ABCA4* alleles were identified in the cohort of 44 patients of African American origin presenting with clinically diagnosed *ABCA4*-associated disease. These data correlate well with the comparative numbers/fractions of patients of mostly European descent. In a cohort of 496 patients presenting with likely *ABCA4*-associated diseases of mostly European origin, we identified 698 *ABCA4* mutations (70.4%); while in the cohort of 244 patients of definitely European origin this fraction was 352 (72.1%).

We identified 47 different *ABCA4* variants in the 44 patients (Publication I, Tables 1-3). Eight variants were deemed benign based on similar allele frequencies between patients and the general population of African American descent, and/or by predictive programs. These included frequent variants c.3602T>G (p.L1201R), c.3899G>A (p.R1300Q), c.1927G>A (p.V643M) and c.4925C>A (p.S1642I). Several known frequent missense or silent variants, previously defined as benign, were not included in the analysis. While the allele frequencies of some of these variants are very different between the general populations of European and African descent, these are not different between the affected and unaffected individuals in each ethnic group. Seven *ABCA4* variants were identified in 2 or more patients, 32 *ABCA4* variants were each detected only once (Publication I, Tables 1 and 3). No large indels in the *ABCA4* locus were identified in any patient.

The most frequent *ABCA4* allele in African American STGD1 patients was c.6320G>A (p.R2107H), accounting for 19.3% of all disease-associated alleles in this cohort. For comparison, the most frequent *ABCA4* mutation in patients of European ancestry, p.G1961E, has an average allele frequency of ~11%.[116] The frequency of the p.R2107H variant in the general population of African Americans is 2% (Publication I, Table 3), higher than the suggested and observed frequency for highly penetrant disease-associated alleles in general. It means that 1 in 2500 African Americans are homozygous for the p.R2107H variant, implicating a much higher than the suggested STGD1 prevalence (1:10,000).[17] This has not been recorded, indicating that the p.R2107H variant is not fully penetrant, although the fraction of late-onset STGD1 might be under-estimated. p.R2107H was called 'disease-causing' by all predictive programs (Publication I, Table 2), and reduced protein expression for this variant has been demonstrated *in vitro*.[104]

The patient homozygous for the p.R2107H variant displayed very mild disease phenotype, with the onset of subtle visual symptoms at the age of 60. Fundus examination showed a bull's eye-appearing macular lesion, relative sparing of photoreceptors in the fovea, with a notable absence of fundus flecks, representing disease progression stage I. Best corrected Snellen visual acuity was $20/20^{-3}$.

Fifteen patients were heterozygous for the p.R2107H allele; in four patients this was the only *ABCA4* allele identified, 11 patients were compound heterozygotes with another disease-associated *ABCA4* variant. Patients with clearly deleterious splicing mutations on the other chromosome (c.3523-1G>A, c.768G>T, c.4540-2A>G) demonstrated an earlier disease onset (1^{st}-3^{rd} decade) and rapid progression, while compound heterozygotes with other missense variants showed later disease onset and slower progression. Based on the relatively high allele frequency for the p.R2107H variant in the general population (minor allele frequency (MAF)=0.02), any of the four monoallelic p.R2107H patients might represent carriers by chance, with mutations in another gene responsible for the STGD1-like phenotype, although all of them reported the onset of visual symptoms in their 50-s or later, agreeing with the mild disease phenotype associated with p.R2107H.

It is also possible that modifying variants or combinations of those in *ABCA4* or other genes might play a role in the p.R2107H-associated disease phenotype.

Comparison with patients of European descent

Of the six most frequent disease-associated variants (p.R2107H, p.V989A, p.G991R, c.5461-10T>C, p.R2040Q, and p.P309R) in African American STGD1 cohort, only the intronic c.5461-10T>C variant was identified at approximately similar frequency in patients of European descent, representing 3.41% and 4.71% of possible disease alleles, respectively (Publication I, Table 3). The p.R2107H, p.V989A, and p.G991R variants were identified in a few cases (5, 1 and 2, respectively) in 244 Caucasian patients, and except for p.R2107H (MAF=0.0001), were not detected in European American general population. Their identification in Caucasian patients most likely suggests admixture between ethnic groups. Conversely, one of the most frequent mutations in Caucasian patients, c.4138C>T (p.P1380L), while absent in the African American general population (Publication I, Table 3), was identified in one African American patient. All other most prevalent disease-associated *ABCA4* variants in the STGD1 cohort of European descent, including p.[L541P;A1038V], p.[G863A,G863del], c.5714+5G>A, were not identified in any of the African American STGD1 patients.

The difference of *ABCA4* mutation spectrum between the two ethnic groups reflected in the clinical presentation of the STGD1 disease, with African American patients reporting later onset of symptoms and generally 'milder' disease expression. ~30% of African American patients reported the disease onset in the 5th decade or later, compared to only 12% (p=0.03) from the Caucasian patient cohort.

This study revealed significantly different genetic causality of the same disease between the two compared racial groups. While differences in racial and ethnic groups have been described, these data emphasize the necessity of very careful evaluation of all STGD1 patients depending on their ethnic background. Patient cohorts of mixed ethnicities lead to inaccurate variant frequency calculations and possible misinterpretation of variants.

2.5 Analysis of the *ABCA4* genomic locus in Stargardt disease (Publication II)

Complete sequencing of the *ABCA4* coding and adjacent intronic sequences in STGD1 patients routinely discovers ~80% of possible disease alleles, two expected disease-associated variants are identified in 60-70% patients, one mutation in ~15-20% of patients, and in the remaining ~15% of patients no disease causing variants are identified. In most STGD1 cases with one *ABCA4* disease-associated allele identified, the second allele is expected to reside in the *ABCA4* locus. It can present as a copy number variant (CNV, large deletion or insertion of one exon or more), a synonymous variant in the coding region, or a (deep) intronic variant, which may affect splicing or regulatory sequences like promoter or enhancers.[117] This study was designed to find the missing disease-causing *ABCA4* variation by a combination of next-generation sequencing (NGS), comparative genome hybridization array (aCGH) screening, familial segregation and *in silico* analyses, and it represents the first analysis of the entire *ABCA4* genomic locus in a large STGD1 patient cohort.

2.5.1 Next-generation sequencing of the *ABCA4* genomic locus

A total of 255 STGD1 patients, with one or no *ABCA4* disease-associated variants identified in *ABCA4* coding sequences were recruited and clinically examined during a 10-year period in different centers in the United States, Italy, Spain, and Denmark. Control cohorts included samples from centers in the United States, Spain, and Denmark.

48 patients were sequenced using RainDance microdroplet-PCR target enrichment (RainDance Technologies, Billerica, MA) with subsequent sequencing on Roche 454 platform (454 Life Sciences, Branford, CT). The target included the genomic region chr1:94,453,727-94,595,732 (GRCh37/hg19), covering the *ABCA4* genomic locus, ~9kb of 5' UTR and ~4.6kb of 3'UTR sequences. The design covered 100% of the targeted area via 473 amplicons of 350-500bp.

The rest of cases were analyzed by the Illumina Truseq Custom Amplicon target enrichment strategy and sequencing (Illumina, San Diego, CA). The Illumina design targeted the genomic region chr1:94,456,700-94,591,600, including the *ABCA4* locus, ~4.9kb of 5'UTR, and ~1.7kb of 3'UTR sequences. The region was divided into 9 targets, with seven 500bp and one 1100bp gap introduced into the repeating elements. The cumulative target via Illumina design involved 130,319bp, covered by 94% with 421 amplicons of ~425bp. The next-generation sequencing reads were analyzed against the reference genome GRCh37/hg19, using the variant discovery software NextGENe (SoftGenetics, State College, PA). On average, about 200 variants were called per individual patient.

All variants and their allele frequencies were compared to the 1000 Genomes database,[112] and to the Exome Sequencing Project dataset (http://snp.gs.washington.edu/EVS/; accessed November 2013). New variants that were not recorded in these databases were further analyzed by a combination of predictive *in silico* methods, statistical analyses, segregation with the disease in the families and screening of large patient and control cohorts for population frequencies. Evolutionary conservation of the variants was noted via UCSC Genome Browser (http://genome.ucsc.edu). The Combined Annotation Dependent Depletion (CADD) algorithm (http://cadd.gs.washington.edu/score)[118] was used to estimate combined predicted general deleteriousness of every variant.

2.5.2 Discovery of new disease-associated variants in *ABCA4* non-coding sequences

Sequencing of the entire *ABCA4* genomic locus, at an average depth of coverage of 100X, in 130 patients with *ABCA4*-associated disease harboring one previously known *ABCA4* disease-associated allele, and 6 patients with no known *ABCA4* mutations, resulted in detecting 1745 different variants. After excluding 83 previously known disease-associated or benign variants from coding regions, 695 variants with no statistically significant differences in allele frequencies between the general population and the patient cohort (unless on the same haplotype with the most frequent known *ABCA4* coding mutation, p.G1961E), 526 incorrectly called homopolymer indels, and 300 false-positive single nucleotide variants, 141 new intronic *ABCA4* variants remained in 114 patients. In 22 patients, the second pathogenic *ABCA4* allele was also identified in the coding sequences, due to re-evaluation (in 6/22 cases) of several variants from benign to disease-causing (e.g., p.G991R and p.A1773V), and false-negative results (in 16/22 cases) in the sequencing of the *ABCA4* coding regions. A relatively high A>C/C>A/T>G/G>T false-positive calling rate was experienced with Illumina sequencing. More stringent criteria for variant calling would deliver 'cleaner' results, however, with the risk of

filtering out some real variants. The analysis of all following discussed *ABCA4* intronic variants is presented in Publication II, Table 1.

The most frequent new deep-intronic candidate variant c.[4539+2064C>T;5461-1389C>A] was identified as a complex allele in 4 patients of Spanish or Italian descent. The complex allele segregated with the disease in all three STGD1 families from Spain, and was not identified in 100 matched Spanish control samples. Neither of these variants was predicted to have any effect on splicing, whether on existing cryptic splice sites or on creating new site. Since the site of c.5461-1389C>A is evolutionarily less conserved, and the c.4539+2064C>T variant is adjacent to the previously reported c.4539+2028C>T and c.4539+2001G>A variants[117] from a conserved area, the c.4539+2064C>T variant was called as disease-associated. Later, functional splice assays via midigene analysis of c.4539+2064C>T revealed partial missplicing with several aberrant transcripts containing a previously reported pseudoexon.[119, 120]

c.5196+1056A>G and c.6006-609T>A, were detected in 3/114 patients each and were absent in 368 matched control samples. The c.5196+1056A>G variant segregated with the disease in two families, is predicted to strengthen a cryptic splice donor (Publication II, Supplement Figure 1, A), and was identified, in addition, in 1/119 patients from our replication cohort of STGD1 patients with one known *ABCA4* mutation. Splicing analysis from the patient's cultured keratinocytes revealed inclusion of a 177-bp segment into the ABCA4 transcript.[117] All 5 STGD1 patients harboring the c.5196+1056A>G variant in our cohort, are compound heterozygotes with a missense, and not a deleterious variant, two patients carry the mild p.G1961E variant on the opposite allele. This aggregate evidence suggests that the c.5196+1056A>G variant is a rare, rather severe, deep intronic disease-associated allele.[10]

The c.6006-609T>A variant was additionally identified in 6/119 STGD1 samples with one known *ABCA4* mutation in the replication cohort. Two of these samples originated from Denmark, and screening of Danish control samples identified the variant in 2/182 controls. While the frequency of c.6006-609T>A is 10X elevated in STGD1 patients as compared to all controls (1.9% vs. 0.18%), and 3.5X when compared to Danish controls (0.55%), the variant is not predicted to affect splicing, so there was not enough evidence for disease association at that time. Further studies identified the c.6006-609T>A variant on the same chromosome with the previously reported candidate variant c.4253+43G>A, [12, 121] discussed in more detail in chapter 2.6 Frequent hypomorphic *ABCA4* variants and late-onset disease.

The previously reported c.4539+2001G>A and c.4539+2028C>T variants were both found in 2/114 samples in our primary cohort. The c.4539+2028C>T change was also detected in 2/119 patients of the replication cohort, and comprised altogether 0.86% of *ABCA4* alleles in 233 unsolved STGD1 patients, a frequency significantly lower than the reported 4.6%.[117] Analysis of *ABCA4* mono-allelic patients from Belgium identified the c.4539+2001G>A variant in 18/70 (26%) patients, as the most prevalent variant in the Belgian cohort.[122] c.4539+2001G>A and c.4539+2028C>T variants are not predicted to have an effect on splicing and no aberrantly spliced products were detected in the patients' fibroblasts. However, RNA analysis from patient-derived photoreceptor precursor cells (PPC) revealed fractions of ~25% and ~15% aberrantly spliced transcripts for c.4539+2001G>A and c.4539+2028C>T variants, respectively, resulting in a retina-specific inclusion of a 345bp pseudoexon.[119] c.4539+2001G>A is the first described recurrent disease-associated deep intronic founder variant in *ABCA4*. Depending on the

ABCA4 mutation in *trans*, c.4539+2001G>A was associated with moderate-to-severe impact on the STGD1 phenotype.[122]

Three variants, c.570+1798A>G, c.2161-8G>A, and c.859-9T>C were each detected in 2/114 different patients in our primary *ABCA4* locus screening cohort, while absent in 368 control samples and also in 119 additional STGD1 samples from the replication cohort. According to predictions, the c.570+1798A>G variant creates a very strong new donor site (Publication II, Supplement Figure 1, B), the c.2161-8G>A variant reduces the existing acceptor by 50%, and the c.859-9T>C variant by 14% (Publication II, Supplement Figure 1, C, D). The c.2161-8G>A and c.859-9T>C variants are adjacent to *ABCA4* coding sequences, yet neither of these variants was detected in the Exome Sequencing Project, containing 4300 individuals of European American descent and 2203 individuals of African American descent. Because of strong predicted splicing effect, and absence in control cohort (and in general population in case of c.2161-8G>A), c.570+1798A>G and c.2161-8G>A were considered strong candidates for *ABCA4* disease-associated variants.

Due to the relatively small predicted splicing defect, c.859-9T>C was not considered a strong candidate variant at first. c.859-9T>C was additionally identified homozygously in two STGD1 patients with no other disease-associated *ABCA4* candidate variants, and heterozygously in 3 more STGD1 patients with one known *ABCA4* mutation. Segregation analysis was possible in one family, confirming that c.859-9T>C was not on the same chromosome with the proband's other mutation, c.5917del (Publication II, Figure 1, E). Review of the ethnic origin of all these patients revealed Asian Indian descent originating from Pakistan, India, or Bangladesh. Following study including 38 STGD1 patients of South Asian origin, identified c.859-9T>C in 9.2% of all possible *ABCA4* alleles in this ethnic STGD1 patient cohort. The minor allele frequency of c.859-9T>C in South Asian general population is 0.07%, according to the gnomAD database.[14] These data suggest, that c.859-9T>C is a frequent disease-associated *ABCA4* allele in STGD1 patients of South Asian origin. Although c.859-9T>C was classified as a mild variant, after midigene *in vitro* splicing assay determined aberrant splicing in ~24% of transcripts,[94] it is associated with an early, elevated accumulation of lipofuscin disease phenotype, compared to other *ABCA4* genotypes.[14]

Seventeen more new variants were each detected in 2/114 patients from the locus screening cohort. Eight of these variants were found allelic with the previously known *ABCA4* exon variants p.V931M, p.R212C, or p.[L541P;A1038V], and were eliminated from the pool of possible mutation candidates. Other variants were either in the same patients who carried a stronger intronic mutation candidate, not conserved in non-human primates, or were found with similar frequency in controls. None of them were predicted to affect splicing.

The remaining 117 new *ABCA4* intronic variants were only detected once each in 62 different patients with one previously known *ABCA4* mutation. Twelve of these variants were predicted to have an effect on splicing. The c.5836-3C>A and c.2654-8T>G variants are predicted to reduce the adjacent canonical splice acceptors on average by 25% and 35%, respectively (Publication II, Supplement Figure 1; E, F). Neither of these two variants was detected by Exome Sequencing Project (ESP), nor in our entire STGD1 patient cohort (780 patients) where all *ABCA4* coding regions and adjacent splice sites have been sequenced. c.5836-3C>A variant revealed a strong splicing defect in midigene *in vitro* splicing assay, with no correctly spliced transcript detected.[94]

The variants c.67-2023T>G, c.1938-619A>G, c.3050+370C>T, c.4352+61G>A, and c.2919-826T>A, are all predicted to strengthen the existing cryptic splice donors

(Publication II, Supplement Figure 1; G, H). None of these variants were found in 368 control samples, nor in the replication cohort of 119 additional STGD1 patients with one *ABCA4* disease-associated allele. The c.1938-619A>G variant segregated with the disease in one family (Publication II, Figure 1, F). The c.67-2023T>G and c.2919-826T>A variants are not conserved in non-human primates and were therefore, not considered strong candidates for disease-association in humans. At the same time, these two patients do not harbor any other possible *ABCA4* candidate variants. c.1938-619A>G, c.3050+370C>T, c.4352+61G>A, and c.2919-826T>A confirmed the predicted splicing defects to variable extents in *in vitro* splicing assay (Dr. F. Cremers, personal communication), and are considered as strong candidates for *ABCA4* disease-associated variants. The variants c.2160+584A>G and c.4539+1729G>T are predicted to create new strong splice donors (Publication II, Supplement Figure 1; I, J). The variants were absent in 368 control samples and in the replication cohort of 119 STGD1 patients. Since these positions are highly conserved among species we suggest that the c.2160+584A>G and c.4539+1729G>T variants are strong candidates for intronic *ABCA4* mutations.

The variants c.5196+1137G>A, c.6148-471C>T and c.6730-541T>C have a predicted effect of strengthening the existing cryptic splice acceptors (Publication II, Supplement Figure 1; K-M), and were each detected in 1/114 patient in the primary locus screening cohort. The c.6148-471C>T and c.6730-541T>C variants were not detected in 368 control samples, nor in 119 additional STGD1 samples. The c.6730-541T>C variant was identified on the same chromosome with the probands other known *ABCA4* mutation, and was therefore excluded from possible candidates list.

The recently reported c.5196+1137G>A variant (5.3% of possible *ABCA4* alleles in unexplained STGD1 patients) [117] was additionally found in 9/119 STGD1 samples with one previously known *ABCA4* variant. Three of these patients were from Denmark, three were of African American origin, and three of unknown ethnicity. Screening matched control cohorts revealed no carriers for this allele in 180 Danish control samples, but it was detected in 8/200 (MAF=2%) general population controls of African American descent, strongly suggesting non-pathogenicity. Comparing the sequence in this region in other species revealed the human variant nucleotide A as the reference sequence nucleotide in *Macaca mulatta*. Sequencing identified the G/A variance in c.5196+1137 in *M. mulatta* samples as well, only in reverse, the human major allele c.5196+1137G is the minor allele in *M.mulatta* (MAF=0.24, in 84 samples), and the suggested human mutant allele A is the major allele in macaques. The c.5196+1137G>A variant caused inclusion of pseudoexons in partially aberrant splicing in patient's cultured keratinocytes,[117] however RNA analysis from macaque retina homozygous for the A allele at the position c.5196+1137 clearly showed no splicing defect (Publication II, Figure 2). Splicing may be affected by the local sequence context, and be less conserved, but the 200bp DNA sequence around c.5196+1137 in humans and macaques is 96% identical. PCR RFLP was used for genotyping the c.5196+1137G>A in controls and additional samples, and the high number of identified cases compared to the initial sequencing cohort with only one c.5196+1137G>A identification prompted us to sequence the c.5196+1137G>A samples identified with PCR RFLP. Sequencing revealed another variant c.5196+1136C>A, MAF=0.04 in African general populations, in all 8/200 African American controls and the three African American STGD1 patients, instead of c.5196+1137G>A. PCR RFLP was not distinguishing between those two adjacent variants. c.5196+1137G>A was found allelic with another deep intronic variant c.1555-2745A>G in 5/7 cases. c.1555-2745A>G is not predicted to effect any cryptic splice sites or create new sites. In conclusion,

c.5196+1137G>A was not identified in general population samples, but testing its effect on RNA splicing in patient-derived PPCs should clarify its disease-association.

2.5.3 Analysis of regulatory sequences

To assess the regulatory potential of the new *ABCA4* intronic variants we compared their chromosome coordinates against the predicted regulatory regions from two ENCODE datasets: 1) Combined DNaseI hypersensitivity clusters from 125 cell types and 2) ChIP-seq clustered regions for 161 transcription factors in 91 cell types. Since regulatory regions, in particular promoters, tend to be DNase sensitive, variants that fall within such regions may affect regulatory potential. Also, variants that are located in transcription factor binding sites may potentially have an effect on protein binding. Unfortunately these datasets did not contain eye-specific cell types or transcription factors. The defined regions were assigned normalized scores in the range of 0-1000, with higher scores indicating stronger signal strength. Twenty four of the 141 new *ABCA4* intronic variants are located in regions with both DNaseI hypersensitivity and transcription factor binding scores of various strength (Publication II, Supplementary Table 1). Family members were available for segregation analyses in three cases. Variants c.570+1499C>A and c.67-3166C>T were on the same chromosome with the probands' other mutation, while c.571-1707C>T and c.5018+289T>C were both detected in one patient and on the different chromosome than this patient's other mutation, suggesting possible pathogenicity. Twenty one variants fall within regions with only DNaseI hypersensitivity score, and 12 variants within regions of transcription factor binding consensus sequences.

The new *ABCA4* intronic variants were also subjected to the Combined Annotation Dependent Depletion (CADD) algorithm (https://cadd.gs.washington.edu/snv)[118]. The CADD algorithm combines a diverse array of annotations, including conservation metrics, functional genomic data like DNase hypersensitivity and transcription factor binding, distance to exon-intron boundaries, expression levels in commonly studied cell lines, and protein-level scores like Grantham, SIFT, and PolyPhen into one metric (C-score) for each variant, ranking a variant relative to all possible substitutions of the human genome. A C-score of greater than 10 indicates that the variant is among the top 10% of most deleterious substitutions in the human genome. C-scores for the new *ABCA4* intronic variants range from 0.004 to 15.14 (Publication II, Supplementary Table 1), with nine variants resulting in a C-score greater than 10. Four of those variants were already classified as possibly disease-associated due to their strong predicted effect on splicing (Publication II, Table 1). The other five variants with higher C-scores included c.4539+1971T>C, c.67-3779T>G, c.303-906A>G, c.859-41T>C, and c.1357-221T>A.

In summary, we found a very strong or a probable intronic mutation candidate in 27/114 (23.7%) patients with one existing definite *ABCA4* mutation (Publication II, Table 1). No immediately plausible candidates for intronic mutations were found in 36/114 (31.6%) patients. The remaining 51 (44.7%) patients possessed one or more new intronic variants that were only detected once, had no effect on splicing according to prediction programs and, therefore, are very difficult to confirm or refute as disease-associated alleles with the available methods and the impossibility to analyze patient RNA. However, it is highly likely that a fraction of these are associated with STGD1.

2.5.4 Analysis of the copy number variants

Some disease-associated *ABCA4* alleles can present as large deletions or insertions of exons and chromosomal segments, i. e. copy number variants (CNV), that evade detection in PCR-based methods. We performed whole genome sequencing (WGS, Macrogen Inc, Rockville, MD) in 16 unexplained STGD1 patient samples and 166 STGD1 patient samples were analyzed with custom Comparative Genome Hybridization arrays (aCGH), to identify possible CNVs. A DNA sample with a previously reported 1030 bp heterozygous deletion of exon 18 in *ABCA4*, was used as positive control in aCGH analysis.[111] Experimental procedures of aCGH were performed as described previously with minor modifications.[123] Custom aCGH arrays (Agilent Technologies, Santa Clara, CA), were designed with ultra high-density probes in *ABCA4* locus and eight other known genes causing macular disease *(ELOVL4, PRPH2, BEST1, RS1, CNGB3*), and major AMD-associated loci *(CFH, ARMS2, C2/CFB*), throughout the entire genomic length of the genes plus the 5' promoter regions and 3' downstream region. The details of the array design are described in the Publication II, Supplementary Table 2. Agilent Genomic Workbench version 7.0 software was used for data analysis, and PCR validations were performed for the plausible true-positive CNVs.

Four large deletions were discovered in the analysis of 182 samples, representing 1% of all possible *ABCA4* alleles in those samples, and matching with previously reported frequency assessment for CNVs in genetically unsolved STGD1 cases.[111] However, a much higher fraction of CNVs, 4.5% of all alleles, was recently reported analyzing 67 *ABCA4* monoallelic STGD1 patients from Netherlands, identifying 6 structural variants.[120]

We identified a ~8.4 kb deletion including exon 6, 1095 bp deletion of exon 18, ~2.3 kb deletion of exon 2, and a ~4.8 kb deletion variant, spanning exons 45-47 and integrating a small insertion. The 1095 bp deletion of exon 18 is almost identical with the previously reported 1030 bp deletion of exon 18 that was found in 2/308 STGD1 subjects (0.32%), but was absent in a control group consisting of 96 individuals over age 60 with normal eye examinations.[111] The other 3 large deletions are novel.

The ~4.8 kb deletion of exons 45-47 was identified in a large unusual two-generation family segregating two distinct macular disease phenotypes in a pseudo-dominant inheritance pattern, caused by four distinct *ABCA4* pathogenic alleles. Sequencing of the *ABCA4* coding regions discovered two known missense mutations, p.C54Y and p.G1961E, sequencing of the *ABCA4* genomic locus uncovered the disease-associated deep intronic variant, c.4539+2028C>T, and CNV analysis revealed a large novel deletion combined with a small insertion, c.6148-698_c.6670del/insTGTGCACCTCCCTAG. This family illustrates the clinical and genetic complexity of *ABCA4*-associated diseases, segregating disease-associated variants from all classes of mutations, coding region, deep intronic, and copy number variants, combinations of which accounted for different disease phenotypes. Examples like this highlight the need for thorough genetic analysis in *ABCA4*-associated disease, including complete sequencing of the *ABCA4* genomic locus, and CNV analysis.[8]

2.5.5 Discussion

The *ABCA4* gene has been subject of intensive genetic research along with following major challenges: 1) the size of the gene, 2) the exceptional genetic heterogeneity, 3) the impossibility of *in vivo* RNA analyses from patients due to gene expression restricted to photoreceptors (and possibly RPE) only.

In the current study of non-coding genetic variation on the largest STGD1 patient cohort analyzed to date, we were able to (almost) unequivocally assign pathogenicity to 12/141 variants in intronic sequences of *ABCA4*. Analysis of the pathogenicity of intronic

ABCA4 variants, potentially affecting splicing and regulatory regions influencing the gene expression, remains complicated. With the lack of patient RNA for direct structural and expression studies, the analysis is limited to the comparisons of variant frequencies in STGD1 patients and in the matched general populations, to *in silico* suggestions by predictive softwares, and to segregation analyses in families. Most of the candidate variants which occurred only once in the *ABCA4* locus remain of uncertain significance, especially those with no predicted effect *in silico.*

Analyses of 'illegitimate transcripts' from non-retinal cell cultures or 'minigene' strategies have been employed to assess the effect of certain variants on splicing, but these studies have serious limitations and the results, while suggestive, cannot be considered unequivocal.[42, 117] Recently proposed 'midigene' approach proved very effective for variants in canonical and non-canonical splice consensus sequences.[94]

The challenges of assessing the variants' effect on splicing *in vitro* are best exemplified by the most frequent non-coding disease-associated variant in the *ABCA4* gene, the c.5461-10T>C variant in intron 38, found in ~7% of STGD1 patients of European descent.[12] The functional consequences of this variant remained unknown until recently. The first minigene assay did not show splicing defects, although these are not predicted by algorhitms either.[42] The c.5461-10T>C variant was called pathogenic based on its segregation with the severe disease phenotype in families, and the 150X higher allele frequency in patients compared to general population (0.07 vs 0.00045). No other possibly disease-causing variants were found on the same chromosome with the c.5461-10T>C variant. The functional effect of c.5461-10T>C was elucidated in *ABCA4* mRNA analysis from patient-derived photoreceptor progenitor cells (PPCs), the closest possible biologically relevant cultured cell type. Complete skipping of exon 39 or exons 39 and 40 in the mRNA, due to the c.5461-10T>C variant was shown, explaining the severe cone–rod dystrophy phenotype in STGD1 patients homozygous for the c.5461-10T>C variant.[93]

Other intronic rare variants near canonical splice sites should be regarded as strong candidate variants for pathogenicity as well, even without strong *in silico* effect prediction, until there is more evidence to refute or confirm their disease-association.

Assessment of the variants' effect on splicing addresses just one possible mechanism for pathogenicity in non-coding sequences. A subset of non-coding variants may contribute to the pathogenesis of ABCA4 disease by disrupting gene regulatory elements, like proximal promoters or more distal enhancers, and splicing enhancers. Variants in regulatory regions have been shown to contribute to visual impairment in humans as well as to other human diseases,[124, 125] but have not yet identified for *ABCA4* gene, since enhancer elements around the ABCA4 gene locus had not been characterized. Recently, promoter and candidate enhancer regions surrounding and within the introns of the *ABCA4* locus were identified by analysis of chromatin accessibility, as determined by assay for transposase-accessible chromatin (ATAC-Seq) and dimethylation of histone protein H3 at lysine 4 (H3K4me2) and acetylation of H3 at lysine 27 (H3K27ac), determined by chromatin immunoprecipitation and high throughput sequencing (ChIP-Seq)(unpublished data: Cherry et al., doi: https://doi.org/10.1101/412361). These features correlate with the ability of the underlying DNA sequences to activate reporter expression in *ex vivo* reporter assays in the intact mouse retina suggesting that they are indeed active regulatory elements in a biologically relevant tissue.[126] Some of those promoter and enhancer regions were already included in the *ABCA4* locus sequencing design where we identified ~10 non-coding candidate variants. Sequencing

of all the candidate enhancers and the promoter in unexplained STGD1 patients will determine more prospective regulatory variants to be tested in established *ex vivo* luciferase reporter assays[127] and identify regulatory variants contributing to *ABCA4*-associated retinal disease.

Similar to the coding sequences, there is extensive genetic variability in the non-coding sequences of the *ABCA4* locus. The non-coding sequences do not harbor frequent disease-associated variants in STGD1 patients of European-American descent. Except for the deep intronic founder mutation in Belgian STGD1 cohort, all definitely or likely pathogenic variants are individually rare in the populations of European descent.[110, 120, 121]

Functional analysis of *ABCA4* deep intronic variants remains a laborious endeavor either through midigenes or patient iPSC-derived photoreceptor progenitor cells.[119] The initial analysis to select the strongest candidate pathogenic variants can be performed in large, familial cohorts of STGD1 patients by a combination of computational approaches including the assessment of variant frequencies in STGD1 patients and in the matched general populations, suggestions by predictive software programs, segregation analyses in families and by integrating clinical and genetic analysis. As demonstrated in this study, we have been able to assign pathogenicity using the above criteria, to several rare deep intronic variants, that were confirmed as pathogenic by demonstrating aberrant splicing in subsequent *in vitro* cell culture analyses.

2.6 Frequent hypomorphic *ABCA4* variants and late-onset disease (Publication III)

Despite significant improvement in sequencing technologies and in understanding of genetic variation in the *ABCA4* locus, genetic causality still remains unclear in ~30% of all STGD1 cases. Recent efforts have been directed towards two goals - to find the missing *ABCA4* disease-causing alleles in non-coding sequences, and to determine the causal genes in patients with STGD1 diagnosis but no *ABCA4* alleles identified. Substantial progress has occurred regarding both objectives; several deep intronic *ABCA4* variants have been proven to affect splicing[119-121] and preliminary data indicate that such is also true of rare variants affecting ABCA4 expression. We have also shown that large CNVs, which elude detection by sequencing, are very rare in the *ABCA4* locus and account for only a small fraction of 'missing' alleles. Whole exome sequencing (WES) in STGD1 patients with no *ABCA4* mutations has uncovered causal genes in ~60% of cases of those STGD1 phenocopies, including known retinal disease genes (CRB1, CRX, PROM1, ROM1, etc)[128-130] and new genes (RAB28, RDH11).[131, 132] However, despite all these advances, the second causative mutation in most patients with 1 *ABCA4* allele remained unidentified following complete *ABCA4* locus sequencing.[133] This study demonstrates that some known, frequent *ABCA4* variants, considered benign due to the high minor allele frequency (MAF) in the general population, are in fact hypomorphic alleles, causing disease expression under certain conditions.

The study cohort consisted of 643 STGD1 patients of (mostly Eastern) European descent. Of these, 2 *ABCA4* mutations were identified in 437 cases (68%), 1 mutation in 117 cases (18%) and 0 mutations in 89 cases (14%) (Publication III, Supplemental Table 1), leaving ~23% of disease-associated alleles in 32% of patients yet to be identified. Almost all patients with no *ABCA4* mutations and ~50% of patients with 1 mutation were screened by WES to determine if variants in other genes were causal in these cases.

All cases, where disease-associated variants in other genes were detected, were excluded from this cohort.

2.6.1 p.N1868I variant

The *ABCA4* gene harbors many common missense variants with MAFs >5% in the general population. The allele frequency of the c.5603A>T (p.N1868I) variant is ~6.6% in European, 2% in Latino and South Asian, and 1% in African general populations (gnomAD). p.N1868I variant was identified in patients with ABCA4 disease three to four times more frequently than expected (~20% vs 6.6%), an observation which had been also previously reported.[67, 134] This variant was determined to be on the same allele with several more frequent *ABCA4* disease-causing variants, notably c.5461–10T>C (p.[Thr1821Valfs*13,Thr1821Aspfs*6]) and c.2588G>C (p.[G863A,G863del]) (Publication III, Table 1 and Supplemental Table 1). Therefore, it was assumed that the more frequent occurrence of p.N1868I in patients with ABCA4 disease is due to linkage disequilibrium with other truly pathogenic alleles. However, when complex alleles with known pathogenic mutations were excluded from these calculations, the allele frequency of p.N1868I in patients remained significantly higher.

p.N1868I was detected in 197 alleles in 176/643 patients; homozygously in 21 cases and heterozygously in 155. When dividing the cohort into three groups based on the number of detected *ABCA4* pathogenic alleles as described above, the frequency of the p.N1868I allele was 11.6% (101/874) in 2-mutation cases, 33% (77/234) in 1-mutation cases and 10.6% (19/178) in 0-mutation cases. Therefore, the allele frequency of p.N1868I was 2X higher than in the matched general population in cases with 2 and 0 mutations and 5X higher in cases with 1 *ABCA4* mutation (33% vs 6%; $p<0.0001$). More frequent *ABCA4* disease-causing variants that were always on the same allele (in *cis*) with p.N1868I, included c.5461–10T>C (p.[Thr1821Valfs*13,Thr1821Aspfs*6]), c.2588G>C (p.[G863A,G863del]), c.4469G>A (p.C1490Y), and c.4594G>A (p.D1532N). Other, more rare disease-causing *ABCA4* variants found allelic with p.N1868I after determining the phase included c.319C>T (p.R107*), c.1253T>C (p.F418S), c.2552G>A (p.G851D), c.5114G>A (p.R1705Q), c.5318C>T (p.A1773V) and c.5572T>A (p.Y1858N).

After subtracting the known complex alleles with a disease-associated variant and the p.N1868I variant in *cis*, we determined that over half of the cases with 1 *ABCA4* mutation (67/117, 57.3%) carried the p.N1868I variant in *trans*, i.e. on the other chromosome. (Publication III, Table 1). The haplotype containing the c.5603A>T (p.N1868I) allele has been established.[135] In cases where phase determination was impossible due to the absence of family members, the haplotype was identified by subtracting the genotype of the known allele. Further examination revealed that the mutations on the opposite allele from the p.N1868I variant were deleterious, predominantly null mutations (Publication III, Table 1 and 2). The breakdown of mutations in *trans* with p.N1868I included: 9 stop mutations, 7 small indels resulting in a frameshift, 12 severe splicing-altering variants and 33 missense mutations, including well-known severe variant c.2894A>G (p.N965S)[136] and the complex allele c.[1622T>C;3113C>T] (p.[L541P;A1038V]).[42, 43] Among others, seven cases with arginine to tryptophan substitutions — c.52C>T (p.R18W), c.868C>T (p.R290W), c.1804C>T (p.R602W), c.4918C>T (p.R1640W) and c.6229C>T (p.R2077W) were reported.

Sequencing of the entire *ABCA4* genomic locus had been performed for 93/117 monoallelic ABCA4 cases.[133] In many cases with a single pathogenic *ABCA4* variant, another confirmed or highly likely pathogenic *ABCA4* deep intronic variant(s) were identified. However, none of the 61 cases with the p.N1868I allele harbored any

definitely or very likely pathogenic deep intronic variants, suggesting that the disease phenotype was indeed conferred by the p.N1868I variant in *trans* with a deleterious *ABCA4* mutation.

As a supporting evidence for the possible pathogenicity of p.N1868I, the effect of p.N1868I variant on the ABCA4 protein is predicted as disease-causing by several prediction algorhitms like SIFT, Align-GVGD, MutationTaster, and CADD. p.N1868I protein variant also exhibited a small but reproducible reduction in ABCA4 ATPase activity relative to WT in *in vitro* functional assays.[104]

2.6.2 p.G863A and p.N1868I

The allele frequency of another presumed 'mild' c.2588G>C (p.[G863A,G863del]) variant is ~2% in Northern European populations.[67, 135] and ~0.8% in all Europeans (gnomAD). It was proposed that p.[G863A,G863del] is hypomorphic, and disease-causing only when in *trans* with a 'severe' *ABCA4* allele.[67] In our cohort of 643 STGD1 patients (Publication III, Table 2), the *ABCA4* allele in *trans* from p.[G863A,G863del] was deleterious or very likely deleterious in >70% of cases. However, 2 patients were homozygous for p.[G863A,G863del] and in some cases the opposing allele was known to be mild, for example, the missense variants c.5882G>A (p.G1961E), and c.2947A>G (p.T983A) (Publication III, Supplemental Table 1).

The p.[G863A,G863del] variant is always allelic with p.N1868I in STGD1 patients, which lead us initially to consider the latter as a benign variant on the disease-associated p.[G863A,G863del] haplotype.[135] Analysing data from large cohorts of patients with age-related macular degeneration (AMD) and matched controls revealed that ~90% of cases with the p.[G863A,G863del] variant did not carry the p.N1868I variant, corroborating the previous finding of p.[G863A,G863del] variant being allelic with p.N1868I variant only in ~13% of cases in populations of European ancestry.[135]

The frequency of the complex allele, p.[G863A,G863del;N1868I] is ~13% of all cases carrying the p.[G863A,G863del] allele alone - an average allele frequency of ~1% in the general population of European descent. If neither of these alleles would be disease-causing, we would expect 13 cases (0.01x643x2) with the p.[G863A,G863del] variant and 2 cases (0.01x0.13x643x2) with the complex allele p.[G863A,G863del;N1868I] in the cohort of 643 cases. In fact, 35 cases (17X more of the expected at random) of the complex allele ($p<0.0001$) and only 4 cases (3X less of the expected at random) of the simplex allele ($p<0.0001$) were observed. The 2 homozygous patients carried the p.[G863A,G863del;N1868I] complex allele. The random occurrence for the homozygous complex allele (MAF~0.1%) is 1:1 000 000. These data support the pathogenicity of the complex p.[G863A,G863del;N1868I] allele and not of the p.[G863A,G863del] alone.

2.6.3 p.G1961E variant

The c.5882G>A (p.G1961E) variant is the most common *ABCA4* disease-associated variant in many STGD1 patient cohorts of European descent, regularly comprising 10-20% of all possible *ABCA4* disease alleles.[42, 71, 137, 138] The highest fraction of p.G1961E in a study cohort was reported in analysis of STGD1 patients of South Asian descent where p.G1961E was identified in half of the patients (in 25% of all *ABCA4* alleles).[14] The frequency of p.G1961E in the general population varies widely across ethnic groups, from approximately 0.4% in populations of European origin and 1.4% in South Asians (gnomAD), to approximately 10% in East African (e.g., Somali) populations.[116, 139] Interestingly, the variant is absent in West Africa and, consequently, very rare in African

Americans.[15] The few African American STGD1 cases carrying the p.G1961E allele are most likely due to population admixture with individuals from other ethnic groups. Because of the high MAF in a general population it was suggested that p.G1961E was not pathogenic, at least not in homozygosity. Allele frequency of 10% means that 1/100 Somalis are homozygous for p.G1961E, at the same time the prevalence of STGD1, especially for the late-onset disease, is not known in East Africa.[116]

Assessing the retinal phenotypes in patients homozygous for p.G1961E variant revealed a range in retinal pathology, in general, with a later onset of visual symptoms (in 3rd-5th decade of life) than typically seen in STGD1. The latest age of onset was recorded at 64 years, in a patient initially diagnosed with age-related macular degeneration (AMD), showing that p.G1961E homozygotes in some cases might not express visual symptoms until later in life and that the prevalence of late-onset Stargardt disease may be underestimated and the milder atypical disease phenotype confused with AMD. In patients homozygous for p.G1961E but presenting with earlier disease onset in 1st-2nd decade, the more severe disease phenotypes were linked to additional pathogenic variants found in *ABCA4*.[116]

However, it is plausible that the variable age in disease onset between patients carrying the same *ABCA4* pathogenic variants is influenced by variants, or combinations of these, in additional modifying genes, and as a result, not all homozygous p.G1961E results in disease phenotype. It is also possible that these 'protective' alleles might be more frequent in East African populations. The haplotype for p.G1961E is the same across all populations, suggesting that the mutation occurred only once and in East Africa.

2.6.4 Genotype-phenotype correlations

ABCA4 disease invariably begins as a maculopathy with an enlarging lesion of outer retinal atrophy and accumulation of yellowish flecks, defined as lipofuscin deposits, at the level of the retinal pigment epithelium (RPE). Classical STGD1 patients exhibit progressively severe fleck patterns from very few in the macula to a stage of 'absolute confluence' across the posterior pole between the ages 30 and 40 years (Publication III, Figure 1).

Conversely, patients associated with p.G1961E or p.N1868I variants consistently exhibit milder fleck patterns even at advanced age, low or normal range lipofuscin accumulation and never progress to the absolute confluence stage as illustrated by fundus autofluorescence imaging which detects lipofuscin accumulation (Publication III, Figure 1). Instead of numerous dark atrophy lesions and more advanced fleck stages as in the typical STGD1, p.G1961E or p.N1868I exhibit a well-defined, unifocal lesion with lesion-centric flecks (Publication III, Supplemental Figure 1).

Despite sharing a similar phenotypic trajectory with respect to fleck formation, lipofuscin accumulation, and lesion progression, patients harboring the disease-associated p.N1868I or p.G1961E variant also exhibit very distinct clinical characteristics compared to each other. STGD1 is predominantly juvenile-onset as patients (including those with p.G1961E) report visual symptoms within the first and second decades of life; however, p.N1868I patients report a significantly delayed onset within the fourth decade of life (mean age=36.3 years; $p<0.0001$) (Publication III, Figure 2A). Foveal sparing phenotype, defined as the structural and function preservation of outer retinal layers in the fovea despite the progressing atrophy of the rest of macula, allowing for the retention of central vision, was observed in 84.7% of p.N1868I patients as compared to approximately in one-third of patients with other *ABCA4* alleles (Publication III, Figure 2B). On the contrary to foveal sparing in p.N1868I-associated disease, p.G1961E variant has been

associated with confined macular disease characterized by rapid degeneration of foveal cone photoreceptors, resulting in 'optical gap' and 'bulls eye maculopathy' phenotypes, in either compound heterozygous or homozygous state, and regardless of the mutation on the other allele.[33, 116, 140]

As a complex allele, p.[G863A,G863del;N1868I] resulted in variable phenotypes, depending on the variant on the opposite allele, as expected. Intermediate phenotypes were associated with missense variants, for example, c.2947A>G (p.T983A) and c.1964T>G (p.F655C), while advanced cases were associated with deletions and stop variants, for example, c.5778_5779del (p.R1927fs) and c.6088C>T (p.R2030*). The two patients homozygous for the complex p.[G863A,G863del;N1868I] allele presented with later onset and milder disease and foveal sparing in one of the two cases.

2.6.5 Late-onset ABCA4 disease and AMD

There is an ongoing discussion about the prevalence and role of heterozygous *ABCA4* alleles in AMD, a late-onset, common disease.[42, 77, 78, 82] While some studies demonstrate statistically significantly elevated *ABCA4* alleles in AMD and some do not, it has been suggested that a fraction of AMD cases are actually misdiagnosed late-onset STGD1 patients.[19, 82] It has also been proposed that patients with late-onset ABCA4 disease remain monoallelic after complete *ABCA4* sequencing much more often than those with early onset disease.[19] Phenotype data on true heterozygous carriers of *ABCA4* pathogenic variants, such as parents of STGD1 patients, are conflicting as well, with reports of no detectable disease phenotype or increased lipofuscin accumulation, at least up to 60 years of age,[34] to frequently found disease phenotypes.[141]

Before the discovery of the p.N1868I association with late-onset STGD1, our cohort included 27.6% monoallelic vs. 7.7% biallelic late-onset (≥45 years of age) STGD1 cases. However, after p.N1868I was identified as the second causal variant in ~80% of these monoallelic late-onset cases, a similar fraction of the late-onset phenotype (~10%) was observed in both *ABCA4* monoallelic and biallelic cases, contradicting any association of *ABCA4* monoallelic cases with late-onset STGD1 phenotype.

Genetic analysis including the p.N1868I variant as a causal hypomorph, allows separating true *ABCA4* monoallelic carriers affected with AMD, from biallelic late-onset STGD1 disease.

2.6.6 Hypomorphic intronic *ABCA4* allele

The remaining missing alleles in late-onset STGD1 patients with only one *ABCA4* pathogenic allele could be other, not yet identified as disease-associated, hypomorphic variants either from coding or non-coding sequences of the *ABCA4* locus.

The intronic variant, c.4253+43G>A in intron 28 of the *ABCA4* gene was identified in the first next-generation sequencing analysis of the *ABCA4* gene, statistically significantly (p=0.0003) more often in STGD1 patients, and also two times more frequently in cases with AMD than in matched general population controls. It was suggested as a strong candidate for a disease-associated allele, but there was not enough evidence to make an unequivocal call of pathogenicity.[12]

Extended analysis of our STGD1 patient cohort determined that 29/1155 (2.5%) of STGD1 patients harbor the c.4253+43G>A variant, with statistically significantly higher frequency in patients with only one pathogenic *ABCA4* allele, 23/160 (14.4%), MAF=0.072, compared to MAF=0.013 in all STGD1 cases and MAF=0.006 in the matching general population. It was also identified in 0.41% (3/725) of biallelic patients and in 1.1% (3/270) of cases with no other disease-causing *ABCA4* alleles. Therefore, the c.4253+43G>A

variant is substantially enriched in STGD1 cases, specifically in those with one known pathogenic allele ($p<1\times10^{-7}$; OR=12.87 95% CI [8.37;19.79]).[10] In another patient cohort, testing of 412 genetically unexplained STGD1 cases revealed 29 (7%) cases harboring the c.4253+43G>A variant.[121] Analysis of available family members always resulted in phasing the c.4253+43G>A allele in *trans* with the known *ABCA4* mutation. Almost all variants on the opposite allele from the c.4253+43G>A variant presented as severe, mostly confirmed loss-of-function alleles, including several stop and frameshift variants, c.5461-10T>C, p.[L541P;A1038V], p.C54Y; exactly matching the criteria of pathogenicity for the previously described frequent hypomorphic missense variant p.N1868I.[20] c.4253+43G>A was identified as the first extremely hypomorphic variant in *ABCA4* noncoding sequences, causal only with a deleterious variant on the opposite allele.[10]

In 70% of cases, c.4253+43G>A forms a complex allele with c.6006-609T>A, a deep intronic variant identified in the *ABCA4* genomic locus sequencing analysis.[133] The c.6006-609T>A variant has only been identified as complex allele c.[4253+43G>A;6006-609T>A] in STGD1 patients. According to the gnomAD database, the c.4253+43G>A variant is 1.2–2 times more frequent than c.6006-609T>A in every population except for Finnish Europeans, where the frequency of both variants is the same. Both variants are absent in East Asian population.[10]

Neither of these two variants are predicted to have any effect on splicing, whether on existing cryptic splice sites nor creating new sites; however, the c.4253+43G>A variant does affect several splice enhancer or silencer motifs according to the Alamut prediction tool. *In vitro* splicing assay using midigene constructs confirmed no splicing defect for the c.6006-609T>A variant, but the c.4253+43G>A variant caused exon skipping in ~40% of transcripts.[121]

The clinical phenotype of patients with c.4253+43G>A variant is largely consistent with the described profile of p.N1868I-associated STGD1 patients.[142] As compared to the more prevalent adolescent-onset form of STGD1,[3] patients harboring c.4253+43G>A or p.N1868I progress through milder disease trajectories characterized by macula-confined disease changes, slow accumulation of flecks, and a significantly later onset of disease symptoms (4^{th}-6^{th} decade) due to foveal sparing subphenotype in both the c.4253+43G>A (83.3%) and p.N1868I (84.7%) groups.[10]

2.6.7 Discussion

Some disease-associated *ABCA4* variants present with unusually high allele frequency in general populations. The minor allele frequency of the most common pathogenic variant c.5882G>A (p.G1961E) is as high as 10% in Somalis.[139] While several p.G1961E compound heterozygous and even homozygous STGD1 patients of Somalian ancestry have been described,[116] the penetrance of p.G1961E in this population remains unknown since very few STGD1 cases of Somalian ancestry have been analyzed. However, it is a penetrant hypomorphic allele in all other populations where significant numbers of patients have been screened. For example, the frequency of p.G1961E in South Asian general populations is 1.5% and it was identified as causal variant in ~50% of STGD1 patients (~24% of alleles) from India.[14] Another recently described frequent variant is the c.6320G>A (p.R2107H) allele, with MAF ~2% in African-Americans and MAF >19% in STGD1 patients of AA descent.[15]

The c.5603A>T (p.N1868I) variant is, pending unequivocal functional evidence, (one of) the most frequent causal variant (MAF >6.5% in the matched general population) in Mendelian diseases. It was rightfully considered benign because of the high allele frequency in the general population. The common criteria for pathogenic variants

suggest general population MAFs not higher than 0.5-1%. This study characterized the p.N1868I variant as an 'extreme' hypomorph, which is phenotypically expressed only when in *trans* with a deleterious mutation. The p.N1868I-associated disease phenotype is distinguishable from other STGD1 phenotypes by late-onset and frequent (~85% of cases) association with foveal sparing and, consequently, extended preservation of central vision. Disease-associated p.N1868I compound heterozygous cases account for at least 10% of all known ABCA4 disease and refine another 10% as the p.[G863A,G863del;N1868I] complex allele. Since p.N1868I only results in a mild disease phenotype in combination with loss-of-function alleles, we concluded that many pathogenic missense variants can be considered deleterious even without functional evidence since they are in *trans* with p.N1868I. In addition to the missense variants already proven by functional assays or statistical analyses in large patient cohorts (p.[L541P;A1038V] p.C1490Y, p.N965S, p.C54Y, etc)[43, 104, 136, 143], several other *ABCA4* missense variants, c.3261A>C (p.E1087D), c.3259G>A (p.E1087K), c.5395A>G (p.N1799D), c.1921T>C (p.C641R), c.262G>A (p.G88R) and c.5923G>C (p.G1975R), can be considered deleterious.

Interestingly, while this study demonstrated that the c.2588G>C (p.[G863A,G863del]) variant is most likely not pathogenic without p.N1868I, it presents with a strong effect in *in vitro* assays with reduced protein expression, ATP binding and hydrolysis[104] and reduced retinal transfer.[106] c.2588G>C also has dual effect on protein level, it causes both glycine to alanine substitution and, through aberrant splicing, the deletion of p.G863.[67] The *in vitro* assays likely do not adequately recapitulate *in vivo* conditions, showing discrepant effects, possibly due to detrimental effects of detergent solubilisation on the stability of certain ABCA4 protein variants. Future functional studies of knock-in mouse models, carrying both variants p.N1868I and, p.G863A, individually and as the complex allele, will bypass the issues of *in vitro* assays.

According to structural model for the ABCA4 protein (Publication III, Figure 3) built by homology modelling similar to that of previously described for the RS1 protein,[144] both variants, p.N1868I and p.G863A, are located on the same, lumenal side of the photoreceptor outer segment disc membrane and both are predicted to affect the protein folding by the unfolding mutation screen.[145, 146] It is very plausible, that two missense variants on the same allele can exert an additive impact on the protein function, specifically in case of variants that individually cause milder effects. For example, in STGD1 patients homozygous for the hypomorphic p.G1961E variant but presenting with early onset severe disease phenotypes, additional pathogenic *ABCA4* variants were identified on the same alleles with p.G1961E.[116]

It is known that not all combinations of *ABCA4* disease-associated alleles result in disease phenotype, explaining the estimated STGD1 prevalence of 1 in 8,000 to 10,000[17, 18], even though the carrier frequency for *ABCA4* pathogenic variants in the general populations is as high as 1:20.[67] It has been debated if several mild and more frequent in general populations pathogenic *ABCA4* variants are penetrant in homozygosity. We have analyzed STGD1 patients homozygous for p.G1961E,[116] p.R2107H,[15] and p.[G863A,G863del] variants, and demonstrated that the latter is not disease causing independently, but rather requires p.N1868I variant in *cis* for pathogenicity. We described p.N1868I variant as an extreme hypomorph that is by definition not disease causing in homozygosity, nor in compound heterozygosity with a 'moderate' variant. Based on theoretical calculations on general population *ABCA4* allele frequency data, and a couple of asymptomatic siblings of p.N1868I patients, a less than 5% penetrance

was estimated for the p.N1868I variant when in *trans* with a deleterious variant.[147] While this conclusion was drawn from several assumptions,[18] the generally observed variability in age of onset and phenotypes among individuals carrying the same combinations of *ABCA4* variants is suggestive of possible *cis*- and/or *trans*-acting genetic modifiers which may influence the penetrance of *ABCA4* alleles and, consequently, the overall clinical presentation of STGD1. The modifying genetic variation contributing to the extreme heterogeneity of ABCA4 disease is mostly unknown, leading to possible misdiagnoses, mischaracterized prognostic outcomes and continued challenges in clinical trial design. Determining *cis*-acting modifiers in the *ABCA4* locus in phenotype groups with the same (class of) mutations presenting with different disease phenotype, and *trans*-acting modifiers of ABCA4 disease by whole exome sequencing will elucidate the role of modifiers in the very heterogeneous ABCA4 disease expressivity.

Identification of the p.N1868I variant as a conditional hypomorph explained ~50% of previously unsolved 'one mutation' cases of ABCA4 disease, an important contribution to STGD1 causality that was corroborated in analyses of unsolved STGD1 cases in other cohorts[120] (JZ, unpublished data). This included a large fraction (~80%) of late-onset cases, which were hypothesized to be caused by monoallelic *ABCA4* mutations.[19] These are in fact biallelic STGD1 patients harboring the extremely hypomorphic p.N1868I variant.

These data not only suggest a substantial revision of genetic screening and assessment of pathogenic variants in ABCA4 disease but, more importantly has broad implications for other Mendelian diseases, where careful consideration should be given to seemingly benign and frequent variants, and to specific combinations of these.

CONCLUSIONS

Since the *ABCA4* gene was cloned over 20 years ago,[1] the *ABCA4* locus continues to be the most heterogeneous of monogenic eye diseases in terms of both genetic and phenotypic expression.[148]

Genetic and clinical characterization of African American STGD1 patients revealed a significantly different mutation spectrum compared to the most studied STGD1 patients of European origin. c.6320G>A (p.R2107H) was identified as the most frequent disease-associated variant in African American STGD1 patients, comprising 19.3% of all *ABCA4* disease alleles in this cohort. The allele frequency of p.R2107H in the general population of African Americans is 2%; higher than the generally considered MAF=0.5-1% treshold for pathogenic variants. Except for a couple of patients harboring the c.5461-10T>C and p.G1961E mutations, all other most prevalent disease-associated *ABCA4* variants in the STGD1 cohort of European descent, including p.[L541P;A1038V], p.[G863A,G863del], c.5714+5G>A, were not identified in any of the African American STGD1 patients. p.R2107H was associated with mild and later onset disease phenotype, ~30% of African American STGD1 patients reported the disease onset in the 5th decade or later, compared to only 12% (p=0.03) from the European American patient cohort. Patients' ethnicities have to be recorded and the identified variants compared to matched populations for accurate variant frequency calculations and variant interpretation.

Sequencing of the *ABCA4* coding regions routinely identifies 2 pathogenic *ABCA4* variants in ~70% of STGD1 patients. In 15-20% of patients only 1 pathogenic variant is identified, and in 10-15% no *ABCA4* disease-associated variants are found. The second causal variant in monoallelic patients is expected to reside in *ABCA4* non-coding sequences. Sequencing of the entire 130kb *ABCA4* locus in 233 *ABCA4* monoallelic samples identified many rare and unique intronic variants that can have a pathogenic effect via 2 mechanisms: 1) causing aberrant splicing by enhancing cryptic splice sites or creating new sites; 2) and/or affecting enhancer sequences that influence protein expression. No frequent pathogenic variants in non-coding sequences were identified in an ethnically diverse patient cohort of European descent. The most frequent deep intronic variant c.[4539+2064C>T;5461-1389C>A] was identified in 4/233 samples. The pathogenicity of non-coding variants was assessed by *in silico* predictions, variant frequency comparisons to general population, evolutionary conservation, and by segregation analyses in the families. Based on these analyses, we found a very strong or a probable intronic mutation candidate in 23.7% patients with one existing definite *ABCA4* mutation. The functional effect of several of these candidate variants has already been confirmed in subsequent *in vitro* splicing assays, validating our criteria for pathogenic candidate variant selection.

We determined that c.5603A>T (p.N1868I) variant, previously considered benign due to high allele frequency in general populations (MAF=6.6% in non-Finnish Europeans, gnomAD) is an extremely hypomorphic variant, causing very mild and late onset STGD1 phenotype only as compound heterozygote with a loss-of-function *ABCA4* variant on the other chromosome. The specific disease phenotype, observed in ~85% of p.N1868I-associated patients is characterized by preservation of foveal cones, and therefore, central vision, despite the progressing atrophy of the rest of macula. Identification of the p.N1868I variant as a conditional hypomorph has an impact on several aspects of the ABCA4 disease:

1) To our knowledge, this is a causal variant with the highest reported MAF=6.6% in Mendelian disease, suggesting increasing the variant frequency threshold for analysis of pathogenicity.
2) ~50% of previously unsolved monoallelic STGD1 cases (~10% of all STGD1 patients of European origin) are biallelic compound heterozygotes harboring the p.N1868I variant and a deleterious *ABCA4* allele in *trans*.
3) ~80% of late-onset cases, which were hypothesized to be caused by monoallelic *ABCA4* mutations, are biallelic p.N1868I compound heterozygotes, thereby differentiating late-onset STGD1 from phenotypically similar AMD.
4) The frequent and considered mild pathogenic founder variant in Northern Europe c.2588G>C (p.[G863A,G863del];MAF~2%) forms a complex allele with p.N1868I in 10% of cases in the general population, but in 92% of cases in STGD1 patients, suggesting the highly penetrant pathogenicity of the complex allele p.[G863A,G863del;N1868I], and not the p.[G863A,G863del] alone.

Discovery of frequent hypomorphic variants as substantial contributors to the ABCA4 disease, particularly to the late-onset phenotype, suggests a significant revision of diagnostic screening and assessment of *ABCA4* variation. Altogether, the genetic analysis of the *ABCA4* locus caused a paradigm shift and, consequently, serves as a model for analyzing the genetic causality in not only retinal diseases but all monogenic diseases in general.

MATERIALS AND METHODS

STGD1 patients were, after written informed consent, recruited and clinically examined during a 15-year period in different centers in the United States, Italy, Spain, and Denmark.

ABCA4 genotyping array screening – publication I

Next-generation sequencing of *ABCA4* coding sequences – publications I-III

Comparative genome hybridization array (aCGH) – publications I-II

Next-generation sequencing of *ABCA4* genomic locus – publication II

RT-PCR – publication II

TaqMan genotyping – publications II-III

In silico variant interpretations – publications I-III

REFERENCES

1. Allikmets R, Singh N, Sun H, et al. A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. *Nat Genet* 1997;15:236-246.
2. Fishman GA, Stone EM, Grover S, Derlacki DJ, Haines HL, Hockey RR. Variation of clinical expression in patients with Stargardt dystrophy and sequence variations in the ABCR gene. *Arch Ophthalmol* 1999;117:504-510.
3. Tanaka K, Lee W, Zernant J, et al. The Rapid-Onset Chorioretinopathy Phenotype of ABCA4 Disease. *Ophthalmology* 2018;125:89-99.
4. Cremers FP, van de Pol DJ, van Driel M, 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;7:355-362.
5. Martinez-Mir A, Paloma E, Allikmets R, et al. Retinitis pigmentosa caused by a homozygous mutation in the Stargardt disease gene ABCR. *Nat Genet* 1998;18:11-12.
6. Maugeri A, Klevering BJ, Rohrschneider K, et al. Mutations in the ABCA4 (ABCR) gene are the major cause of autosomal recessive cone-rod dystrophy. *Am J Hum Genet* 2000;67:960-966.
7. Shroyer NF, Lewis RA, Yatsenko AN, Lupski JR. Null missense ABCR (ABCA4) mutations in a family with stargardt disease and retinitis pigmentosa. *Invest Ophthalmol Vis Sci* 2001;42:2757-2761.
8. Lee W, Xie Y, Zernant J, et al. Complex inheritance of ABCA4 disease: four mutations in a family with multiple macular phenotypes. *Hum Genet* 2016;135:9-19.
9. Ciccone L, Lee W, Zernant J, et al. Hyperreflective Deposition in the Background of Advanced Stargardt Disease. *Retina* 2017.
10. Zernant J, Lee W, Nagasaki T, 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.
11. Cornelis SS, Bax NM, Zernant J, et al. In Silico Functional Meta-Analysis of 5,962 ABCA4 Variants in 3,928 Retinal Dystrophy Cases. *Hum Mutat* 2017;38:400-408.
12. Zernant J, Schubert C, Im KM, et al. Analysis of the ABCA4 gene by next-generation sequencing. *Invest Ophthalmol Vis Sci* 2011;52:8479-8487.
13. Burke TR, Tsang SH. Allelic and phenotypic heterogeneity in ABCA4 mutations. *Ophthalmic Genet* 2011;32:165-174.
14. Lee W, Schuerch K, Zernant J, et al. Genotypic spectrum and phenotype correlations of ABCA4-associated disease in patients of south Asian descent. *Eur J Hum Genet* 2017;25:735-743.
15. Zernant J, Collison FT, Lee W, et al. Genetic and clinical analysis of ABCA4-associated disease in African American patients. *Hum Mutat* 2014;35:1187-1194.
16. Nash BM, Wright DC, Grigg JR, Bennetts B, Jamieson RV. Retinal dystrophies, genomic applications in diagnosis and prospects for therapy. *Transl Pediatr* 2015;4: 139-163.
17. Blacharsky P. Fundus flavimaculatus. In: Newsome DA (ed), *Retinal Dystrophies and Degenerations.* New York: Raven Press; 1988:135-159.
18. Allikmets R, Zernant J, Lee W. Penetrance of the ABCA4 p.Asn1868Ile Allele in Stargardt Disease. *Invest Ophthalmol Vis Sci* 2018;59:5564-5565.

19. Westeneng-van Haaften SC, Boon CJ, Cremers FP, Hoefsloot LH, den Hollander AI, Hoyng CB. Clinical and genetic characteristics of late-onset Stargardt's disease. *Ophthalmology* 2012;119:1199-1210.
20. Zernant J, Lee W, Collison FT, 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;54:404-412.
21. Fish G, Grey R, Sehmi KS, Bird AC. The dark choroid in posterior retinal dystrophies. *Br J Ophthalmol* 1981;65:359-363.
22. Burke TR, Duncker T, Woods RL, et al. Quantitative fundus autofluorescence in recessive Stargardt disease. *Invest Ophthalmol Vis Sci* 2014;55:2841-2852.
23. Fishman GA. Fundus flavimaculatus. A clinical classification. *Arch Ophthalmol* 1976;94:2061-2067.
24. Lois N, Holder GE, Bunce C, Fitzke FW, Bird AC. Phenotypic subtypes of Stargardt macular dystrophy-fundus flavimaculatus. *Arch Ophthalmol* 2001;119:359-369.
25. Hussain RM, Ciulla TA, Berrocal AM, Gregori NZ, Flynn HW, Jr., Lam BL. Stargardt macular dystrophy and evolving therapies. *Expert Opin Biol Ther* 2018;18:1049-1059.
26. Molday LL, Rabin AR, Molday RS. ABCR expression in foveal cone photoreceptors and its role in Stargardt macular dystrophy. *Nat Genet* 2000;25:257-258.
27. Sun H, Nathans J. Stargardt's ABCR is localized to the disc membrane of retinal rod outer segments. *Nat Genet* 1997;17:15-16.
28. Lenis TL, Hu J, Ng SY, et al. Expression of ABCA4 in the retinal pigment epithelium and its implications for Stargardt macular degeneration. *Proc Natl Acad Sci U S A* 2018;115:E11120-E11127.
29. Molday RS, Zhang K. Defective lipid transport and biosynthesis in recessive and dominant Stargardt macular degeneration. *Prog Lipid Res* 2010;49:476-492.
30. Zhou J, Kim SR, Westlund BS, Sparrow JR. Complement activation by bisretinoid constituents of RPE lipofuscin. *Invest Ophthalmol Vis Sci* 2009;50:1392-1399.
31. Boyer NP, Higbee D, Currin MB, et al. Lipofuscin and N-retinylidene-N-retinylethanolamine (A2E) accumulate in retinal pigment epithelium in absence of light exposure: their origin is 11-cis-retinal. *J Biol Chem* 2012;287:22276-22286.
32. Biswas-Fiss EE, Affet S, Ha M, Biswas SB. Retinoid binding properties of nucleotide binding domain 1 of the Stargardt disease-associated ATP binding cassette (ABC) transporter, ABCA4. *J Biol Chem* 2012;287:44097-44107.
33. Noupuu K, Lee W, Zernant J, Tsang SH, Allikmets R. Structural and genetic assessment of the ABCA4-associated optical gap phenotype. *Invest Ophthalmol Vis Sci* 2014;55:7217-7226.
34. Duncker T, Stein GE, Lee W, et al. Quantitative Fundus Autofluorescence and Optical Coherence Tomography in ABCA4 Carriers. *Invest Ophthalmol Vis Sci* 2015;56:7274-7285.
35. Muller PL, Gliem M, Mangold E, et al. Monoallelic ABCA4 Mutations Appear Insufficient to Cause Retinopathy: A Quantitative Autofluorescence Study. *Invest Ophthalmol Vis Sci* 2015;56:8179-8186.
36. Mata NL, Weng J, Travis GH. Biosynthesis of a major lipofuscin fluorophore in mice and humans with ABCR-mediated retinal and macular degeneration. *Proc Natl Acad Sci U S A* 2000;97:7154-7159.
37. Sparrow JR, Blonska A, Flynn E, 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* 2013;54:2812-2820.

38. Kong J, Kim SR, Binley K, et al. Correction of the disease phenotype in the mouse model of Stargardt disease by lentiviral gene therapy. *Gene Ther* 2008;15:1311-1320.
39. Dobri N, Qin Q, Kong J, et al. A1120, a nonretinoid RBP4 antagonist, inhibits formation of cytotoxic bisretinoids in the animal model of enhanced retinal lipofuscinogenesis. *Invest Ophthalmol Vis Sci* 2013;54:85-95.
40. Radu RA, Yuan Q, Hu J, et al. Accelerated accumulation of lipofuscin pigments in the RPE of a mouse model for ABCA4-mediated retinal dystrophies following Vitamin A supplementation. *Invest Ophthalmol Vis Sci* 2008;49:3821-3829.
41. Ma L, Kaufman Y, Zhang J, Washington I. C20-D3-vitamin A slows lipofuscin accumulation and electrophysiological retinal degeneration in a mouse model of Stargardt disease. *J Biol Chem* 2011;286:7966-7974.
42. Rivera A, White K, Stohr H, 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* 2000;67:800-813.
43. Zhang N, Tsybovsky Y, Kolesnikov AV, et al. Protein misfolding and the pathogenesis of ABCA4-associated retinal degenerations. *Hum Mol Genet* 2015;24: 3220-3237.
44. Rosenberg T, Klie F, Garred P, Schwartz M. N965S is a common ABCA4 variant in Stargardt-related retinopathies in the Danish population. *Mol Vis* 2007;13:1962-1969.
45. Molday LL, Wahl D, Sarunic MV, Molday RS. Localization and functional characterization of the p.Asn965Ser (N965S) ABCA4 variant in mice reveal pathogenic mechanisms underlying Stargardt macular degeneration. *Hum Mol Genet* 2018;27: 295-306.
46. Makelainen S, Godia M, Hellsand M, et al. An ABCA4 loss-of-function mutation causes a canine form of Stargardt disease. *PLoS Genet* 2019;15:e1007873.
47. Occelli LM, Schon C, Seeliger MW, et al. Gene Supplementation Rescues Rod Function and Preserves Photoreceptor and Retinal Morphology in Dogs, Leading the Way Towards Treating Human PDE6A-Retinitis Pigmentosa. *Hum Gene Ther* 2017.
48. Mowat FM, Breuwer AR, Bartoe JT, et al. RPE65 gene therapy slows cone loss in Rpe65-deficient dogs. *Gene Ther* 2013;20:545-555.
49. Pichard V, Provost N, Mendes-Madeira A, et al. AAV-mediated Gene Therapy Halts Retinal Degeneration in PDE6beta-deficient Dogs. *Mol Ther* 2016;24:867-876.
50. Narfstrom K, Katz ML, Bragadottir R, et al. Functional and structural recovery of the retina after gene therapy in the RPE65 null mutation dog. *Invest Ophthalmol Vis Sci* 2003;44:1663-1672.
51. Kaufman Y, Ma L, Washington I. Deuterium enrichment of vitamin A at the C20 position slows the formation of detrimental vitamin A dimers in wild-type rodents. *J Biol Chem* 2011;286:7958-7965.
52. Radu RA, Mata NL, Nusinowitz S, Liu X, Sieving PA, Travis GH. Treatment with isotretinoin inhibits lipofuscin accumulation in a mouse model of recessive Stargardt's macular degeneration. *Proc Natl Acad Sci U S A* 2003;100:4742-4747.
53. Vallerand IA, Lewinson RT, Farris MS, et al. Efficacy and adverse events of oral isotretinoin for acne: a systematic review. *Br J Dermatol* 2018;178:76-85.
54. Gao S, Parmar T, Palczewska G, et al. Protective Effect of a Locked Retinal Chromophore Analog against Light-Induced Retinal Degeneration. *Mol Pharmacol* 2018;94:1132-1144.
55. Maeda A, Golczak M, Chen Y, et al. Primary amines protect against retinal degeneration in mouse models of retinopathies. *Nat Chem Biol* 2011;8:170-178.

56. Radu RA, Han Y, Bui TV, et al. Reductions in serum vitamin A arrest accumulation of toxic retinal fluorophores: a potential therapy for treatment of lipofuscin-based retinal diseases. *Invest Ophthalmol Vis Sci* 2005;46:4393-4401.
57. Berchuck JE, Yang P, Toimil BA, Ma Z, Baciu P, Jaffe GJ. All-trans-retinal sensitizes human RPE cells to alternative complement pathway-induced cell death. *Invest Ophthalmol Vis Sci* 2013;54:2669-2677.
58. Binley K, Widdowson P, Loader J, et al. Transduction of photoreceptors with equine infectious anemia virus lentiviral vectors: safety and biodistribution of StarGen for Stargardt disease. *Invest Ophthalmol Vis Sci* 2013;54:4061-4071.
59. McClements ME, Barnard AR, Singh MS, et al. An AAV Dual Vector Strategy Ameliorates the Stargardt Phenotype in Adult Abca4(-/-) Mice. *Hum Gene Ther* 2018.
60. Han Z, Conley SM, Makkia RS, Cooper MJ, Naash MI. DNA nanoparticle-mediated ABCA4 delivery rescues Stargardt dystrophy in mice. *J Clin Invest* 2012;122:3221-3226.
61. Lu B, Malcuit C, Wang S, et al. Long-term safety and function of RPE from human embryonic stem cells in preclinical models of macular degeneration. *Stem Cells* 2009;27:2126-2135.
62. Schwartz SD, Tan G, Hosseini H, Nagiel A. Subretinal Transplantation of Embryonic Stem Cell-Derived Retinal Pigment Epithelium for the Treatment of Macular Degeneration: An Assessment at 4 Years. *Invest Ophthalmol Vis Sci* 2016;57:ORSFc1-9.
63. Li Y, Tsai YT, Hsu CW, et al. Long-term safety and efficacy of human-induced pluripotent stem cell (iPS) grafts in a preclinical model of retinitis pigmentosa. *Mol Med* 2012;18:1312-1319.
64. Wahlin KJ, Maruotti JA, Sripathi SR, et al. Photoreceptor Outer Segment-like Structures in Long-Term 3D Retinas from Human Pluripotent Stem Cells. *Sci Rep* 2017;7:766.
65. Allikmets R, Wasserman WW, Hutchinson A, et al. Organization of the ABCR gene: analysis of promoter and splice junction sequences. *Gene* 1998;215:111-122.
66. Rozet JM, Gerber S, Souied E, et al. Spectrum of ABCR gene mutations in autosomal recessive macular dystrophies. *Eur J Hum Genet* 1998;6:291-295.
67. Maugeri A, van Driel MA, van de Pol DJ, 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;64:1024-1035.
68. Fumagalli A, Ferrari M, Soriani N, et al. Mutational scanning of the ABCR gene with double-gradient denaturing-gradient gel electrophoresis (DG-DGGE) in Italian Stargardt disease patients. *Hum Genet* 2001;109:326-338.
69. Shroyer NF, Lewis RA, Yatsenko AN, Wensel TG, Lupski JR. Cosegregation and functional analysis of mutant ABCR (ABCA4) alleles in families that manifest both Stargardt disease and age-related macular degeneration. *Hum Mol Genet* 2001;10: 2671-2678.
70. Rozet JM, Gerber S, Ghazi I, 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* 1999;36:447-451.
71. Lewis RA, Shroyer NF, Singh N, et al. Genotype/Phenotype analysis of a photoreceptor-specific ATP-binding cassette transporter gene, ABCR, in Stargardt disease. *Am J Hum Genet* 1999;64:422-434.

72. van Driel MA, Maugeri A, Klevering BJ, Hoyng CB, Cremers FP. ABCR unites what ophthalmologists divide(s). *Ophthalmic Genet* 1998;19:117-122.
73. Fujinami K, Zernant J, Chana RK, et al. Clinical and molecular characteristics of childhood-onset Stargardt disease. *Ophthalmology* 2015;122:326-334.
74. Klevering BJ, Maugeri A, Wagner A, et al. Three families displaying the combination of Stargardt's disease with cone-rod dystrophy or retinitis pigmentosa. *Ophthalmology* 2004;111:546-553.
75. Simonelli F, Testa F, Zernant J, et al. Association of a homozygous nonsense mutation in the ABCA4 (ABCR) gene with cone-rod dystrophy phenotype in an Italian family. *Ophthalmic Res* 2004;36:82-88.
76. Yatsenko AN, Shroyer NF, Lewis RA, Lupski JR. Late-onset Stargardt disease is associated with missense mutations that map outside known functional regions of ABCR (ABCA4). *Hum Genet* 2001;108:346-355.
77. Allikmets R, Shroyer NF, Singh N, et al. Mutation of the Stargardt disease gene (ABCR) in age-related macular degeneration. *Science* 1997;277:1805-1807.
78. Allikmets R. Further evidence for an association of ABCR alleles with age-related macular degeneration. The International ABCR Screening Consortium. *Am J Hum Genet* 2000;67:487-491.
79. Shroyer NF, Lewis RA, Allikmets R, et al. The rod photoreceptor ATP-binding cassette transporter gene, ABCR, and retinal disease: from monogenic to multifactorial. *Vision Res* 1999;39:2537-2544.
80. Simonelli F, Testa F, de Crecchio G, et al. New ABCR mutations and clinical phenotype in Italian patients with Stargardt disease. *Invest Ophthalmol Vis Sci* 2000;41:892-897.
81. Stone EM, Webster AR, Vandenburgh K, et al. Allelic variation in ABCR associated with Stargardt disease but not age-related macular degeneration. *Nat Genet* 1998;20:328-329.
82. Fritsche LG, Fleckenstein M, Fiebig BS, et al. A subgroup of age-related macular degeneration is associated with mono-allelic sequence variants in the ABCA4 gene. *Invest Ophthalmol Vis Sci* 2012;53:2112-2118.
83. Hageman GS, Anderson DH, Johnson LV, et al. A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc Natl Acad Sci U S A* 2005;102:7227-7232.
84. Jakobsdottir J, Conley YP, Weeks DE, Mah TS, Ferrell RE, Gorin MB. Susceptibility genes for age-related maculopathy on chromosome 10q26. *Am J Hum Genet* 2005;77:389-407.
85. Jaakson K, Zernant J, Kulm M, et al. Genotyping microarray (gene chip) for the ABCR (ABCA4) gene. *Hum Mutat* 2003;22:395-403.
86. Pastinen T, Kurg A, Metspalu A, Peltonen L, Syvanen AC. Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *Genome Res* 1997;7:606-614.
87. Valverde D, Riveiro-Alvarez R, Bernal S, et al. Microarray-based mutation analysis of the ABCA4 gene in Spanish patients with Stargardt disease: evidence of a prevalent mutated allele. *Mol Vis* 2006;12:902-908.
88. Aguirre-Lamban J, Riveiro-Alvarez R, Garcia-Hoyos M, et al. Comparison of high-resolution melting analysis with denaturing high-performance liquid chromatography for mutation scanning in the ABCA4 gene. *Invest Ophthalmol Vis Sci* 2010;51:2615-2619.

89. Ernest PJ, Boon CJ, Klevering BJ, Hoefsloot LH, Hoyng CB. Outcome of ABCA4 microarray screening in routine clinical practice. *Mol Vis* 2009;15:2841-2847.
90. Richards S, Aziz N, Bale 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* 2015;17:405-424.
91. Ghosh R, Oak N, Plon SE. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biol* 2017;18:225.
92. Jian X, Boerwinkle E, Liu X. In silico tools for splicing defect prediction: a survey from the viewpoint of end users. *Genet Med* 2014;16:497-503.
93. Sangermano R, Bax NM, Bauwens M, et al. Photoreceptor Progenitor mRNA Analysis Reveals Exon Skipping Resulting from the ABCA4 c.5461-10T-->C Mutation in Stargardt Disease. *Ophthalmology* 2016;123:1375-1385.
94. Sangermano R, Khan M, Cornelis SS, et al. ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Res* 2018;28:100-110.
95. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res* 2001;11:863-874.
96. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248-249.
97. Tavtigian SV, Deffenbaugh AM, Yin L, et al. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet* 2006;43:295-305.
98. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 2010;7:575-576.
99. Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* 1987;15:7155-7174.
100. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004;11:377-394.
101. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol* 1997;4:311-323.
102. Pertea M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res* 2001;29:1185-1190.
103. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009;37:e67.
104. Sun H, Smallwood PM, Nathans J. Biochemical defects in ABCR protein variants associated with human retinopathies. *Nat Genet* 2000;26:242-246.
105. Tsybovsky Y, Wang B, Quazi F, Molday RS, Palczewski K. Posttranslational modifications of the photoreceptor-specific ABC transporter ABCA4. *Biochemistry* 2011;50:6855-6866.
106. Quazi F, Lenevich S, Molday RS. ABCA4 is an N-retinylidene-phosphatidylethanolamine and phosphatidylethanolamine importer. *Nat Commun* 2012;3:925.
107. Biswas-Fiss EE, Kurpad DS, Joshi K, Biswas SB. Interaction of extracellular domain 2 of the human retina-specific ATP-binding cassette transporter (ABCA4) with all-trans-retinal. *J Biol Chem* 2010;285:19372-19383.

108. Garces F, Jiang K, Molday LL, et al. Correlating the Expression and Functional Activity of ABCA4 Disease Variants With the Phenotype of Patients With Stargardt Disease. *Invest Ophthalmol Vis Sci* 2018;59:2305-2315.
109. Wiszniewski W, Zaremba CM, Yatsenko AN, et al. ABCA4 mutations causing mislocalization are found frequently in patients with severe retinal dystrophies. *Hum Mol Genet* 2005;14:2769-2778.
110. Schulz HL, Grassmann F, Kellner U, 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* 2017;58:394-403.
111. Yatsenko AN, Shroyer NF, Lewis RA, Lupski JR. An ABCA4 genomic deletion in patients with Stargardt disease. *Hum Mutat* 2003;21:636-644.
112. Genomes Project C, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491:56-65.
113. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285-291.
114. Chacon-Camacho OF, Granillo-Alvarez M, Ayala-Ramirez R, Zenteno JC. 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* 2013;109:77-82.
115. Utz VM, Chappelow AV, Marino MJ, et al. Identification of three ABCA4 sequence variations exclusive to African American patients in a cohort of patients with Stargardt disease. *Am J Ophthalmol* 2013;156:1220-1227 e1222.
116. Burke TR, Fishman GA, Zernant J, et al. Retinal phenotypes in patients homozygous for the G1961E mutation in the ABCA4 gene. *Invest Ophthalmol Vis Sci* 2012;53:4458-4467.
117. Braun TA, Mullins RF, Wagner AH, et al. Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. *Hum Mol Genet* 2013;22: 5136-5145.
118. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310-315.
119. Albert S, Garanto A, Sangermano R, et al. Identification and Rescue of Splice Defects Caused by Two Neighboring Deep-Intronic ABCA4 Mutations Underlying Stargardt Disease. *Am J Hum Genet* 2018;102:517-527.
120. Bauwens M, Garanto A, Sangermano R, 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* 2019.
121. Sangermano R, Garanto A, Khan M, et al. Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. *Genet Med* 2019.
122. Bauwens M, De Zaeytijd J, Weisschuh N, et al. An augmented ABCA4 screen targeting noncoding regions reveals a deep intronic founder variant in Belgian Stargardt patients. *Hum Mutat* 2015;36:39-42.
123. Gonzaga-Jauregui C, Zhang F, Towne CF, Batish SD, Lupski JR. GJB1/Connexin 32 whole gene deletions in patients with X-linked Charcot-Marie-Tooth disease. *Neurogenetics* 2010;11:465-470.

124. Ghiasvand NM, Rudolph DD, Mashayekhi M, Brzezinski JAt, Goldman D, Glaser T. Deletion of a remote enhancer near ATOH7 disrupts retinal neurogenesis, causing NCRNA disease. *Nat Neurosci* 2011;14:578-586.
125. Gordon CT, Lyonnet S. Enhancer mutations and phenotype modularity. *Nat Genet* 2014;46:3-4.
126. Andzelm MM, Cherry TJ, Harmin DA, et al. MEF2D drives photoreceptor development through a genome-wide competition for tissue-specific enhancers. *Neuron* 2015;86:247-263.
127. Matsuda T, Cepko CL. Controlled expression of transgenes introduced by in vivo electroporation. *Proc Natl Acad Sci U S A* 2007;104:1027-1032.
128. Ma CJ, Lee W, Stong N, 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* 2019.
129. Tsang SH, Burke T, Oll M, et al. Whole exome sequencing identifies CRB1 defect in an unusual maculopathy phenotype. *Ophthalmology* 2014;121:1773-1782.
130. Yamamoto S, Jaiswal M, Charng WL, et al. A drosophila genetic resource of mutants to study mechanisms underlying human genetic diseases. *Cell* 2014;159: 200-214.
131. Riveiro-Alvarez R, Xie YA, Lopez-Martinez MA, et al. New mutations in the RAB28 gene in 2 Spanish families with cone-rod dystrophy. *JAMA Ophthalmol* 2015;133: 133-139.
132. Xie YA, Lee W, Cai C, et al. New syndrome with retinitis pigmentosa is caused by nonsense mutations in retinol dehydrogenase RDH11. *Hum Mol Genet* 2014;23: 5774-5780.
133. Zernant J, Xie YA, Ayuso C, et al. Analysis of the ABCA4 genomic locus in Stargardt disease. *Hum Mol Genet* 2014;23:6797-6806.
134. Aguirre-Lamban J, Gonzalez-Aguilera JJ, Riveiro-Alvarez R, 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* 2011;52:6206-6212.
135. Maugeri A, Flothmann K, Hemmrich N, et al. The ABCA4 2588G>C Stargardt mutation: single origin and increasing frequency from South-West to North-East Europe. *Eur J Hum Genet* 2002;10:197-203.
136. Bertelsen M, Zernant J, Larsen M, Duno M, Allikmets R, Rosenberg T. Generalized choriocapillaris dystrophy, a distinct phenotype in the spectrum of ABCA4-associated retinopathies. *Invest Ophthalmol Vis Sci* 2014;55:2766-2776.
137. Hargitai J, Zernant J, Somfai GM, et al. Correlation of clinical and genetic findings in Hungarian patients with Stargardt disease. *Invest Ophthalmol Vis Sci* 2005;46:4402-4408.
138. Passerini I, Sodi A, Giambene B, Mariottini A, Menchini U, Torricelli F. Novel mutations in of the ABCR gene in Italian patients with Stargardt disease. *Eye (Lond)* 2010;24:158-164.
139. Guymer RH, Heon E, Lotery AJ, et al. Variation of codons 1961 and 2177 of the Stargardt disease gene is not associated with age-related macular degeneration. *Arch Ophthalmol* 2001;119:745-751.
140. Cella W, Greenstein VC, Zernant-Rajang J, et al. G1961E mutant allele in the Stargardt disease gene ABCA4 causes bull's eye maculopathy. *Exp Eye Res* 2009;89: 16-24.

141. Kjellstrom U. Reduced macular function in ABCA4 carriers. *Mol Vis* 2015;21: 767-782.
142. Collison FT, Lee W, Fishman GA, et al. CLINICAL CHARACTERIZATION OF STARGARDT DISEASE PATIENTS WITH THE p.N1868I ABCA4 MUTATION. *Retina* 2018.
143. Cideciyan AV, Swider M, Aleman TS, et al. ABCA4 disease progression and a proposed strategy for gene therapy. *Hum Mol Genet* 2009;18:931-941.
144. Sergeev YV, Caruso RC, Meltzer MR, Smaoui N, MacDonald IM, Sieving PA. Molecular modeling of retinoschisin with functional analysis of pathogenic mutations from human X-linked retinoschisis. *Hum Mol Genet* 2010;19:1302-1313.
145. McCafferty CL, Sergeev YV. In silico Mapping of Protein Unfolding Mutations for Inherited Disease. *Sci Rep* 2016;6:37298.
146. McCafferty CL, Sergeev YV. Dataset of eye disease-related proteins analyzed using the unfolding mutation screen. *Sci Data* 2016;3:160112.
147. Runhart EH, Sangermano R, Cornelis SS, 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* 2018;59:3220-3231.
148. Allikmets R. Simple and complex ABCR: genetic predisposition to retinal disease. *Am J Hum Genet* 2000;67:793-799.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my thesis supervisor Professor Rando Allikmets. Dr. Allikmets employed me in his lab in 2001, and through his continuous guidance over all those years I have learned and grown tremendously as a research scientist. I would also like to thank him especially for his patience and support during my thesis writing process.

I am deeply grateful for my colleagues, Winston Lee, for all his excellent work on phenotyping as a substantial contribution bringing our genetic research to the current level, Dr. Takayuki Nagasaki for all the support with computing and big data, and all the current and former fellow lab members for big and little everyday things and their wonderful companionship throughout the years.

I wish to thank Laboratory of Personalized Genomic Medicine for the technical help and collaboration on the next-generation sequencing projects, and my former colleagues from Asper Biotech in Estonia, for our collective effort in developing the microarray genotyping technology, which brought me to Columbia University.

I thank all participating patients and families, and their attending physicians for their willing and continued participation in these studies.

I would like to thank everyone in the Department of Chemistry and Biotechnology at Tallinn University of Technology who helped arranging my thesis defense at Tallinn University of Technology, especially Professors Vello Tõugu and Andres Veske, and Mrs. Maarja Märss.

Finally, I would like to give my gratitude to my parents for their life-long love and support, and to my little family here in America, for always being there.

This work was supported, in part, by grants from the National Eye Institute/NIH, EY021163, EY019861, EY021237 and EY019007 (Core Support for Vision Research); The Pangere Family Foundation, Chicago Lighthouse; Foundation Fighting Blindness (Owings Mills, MD); and unrestricted funds from Research to Prevent Blindness (New York, NY) to the Department of Ophthalmology, Columbia University.

This work is dedicated to the patients with ABCA4 disease and their families.

ABSTRACT
ABCA4 Locus as a Model for Complex Genetics in Monogenic Diseases

Inherited retinal dystrophies (IRD) comprise a genetically and clinically diverse group of disorders characterized by progressive and irreversible vision loss over time. To date, DNA variants in more than 300 genes have been associated with various forms of IRDs.

Genetic variation in the *ABCA4* gene has emerged as the most frequent cause of inherited macular dystrophy, causing the phenotypically heterogeneous autosomal recessive Stargardt disease (STGD1, MIM #248200). Phenotypes caused by variants in the *ABCA4* gene range from early onset, rapidly progressing disease resulting in panretinal degeneration to late-onset, milder phenotype with limited visual deterioration that can phenotypically overlap with the more prevalent AMD. Phenotypic variability is reflected in extensive genetic variation in the *ABCA4* gene, with ~1000 disease-causing variants described in the *ABCA4* locus. The general population frequency of *ABCA4* mutations is 5% - 1:20 people carry a potentially pathogenic *ABCA4* allele. However, the estimated prevalence of Stargardt disease is 1:8,000-10,000, since not all combinations of variants result in disease expression.

Analysis of the Stargardt disease and the underlying *ABCA4* gene is complicated by both, genetic and phenotypic heterogeneity. The three most frequent disease-causing variants in STGD1 cohorts of European origin, p.G1961E, p.[L541P;A1038V], and p.[G863A;N1868I] account for ~20% of all possible disease alleles, while most reported variants are very rare. Many *ABCA4* variants are common in patients with specific racial and ethnic backgrounds, while absent in others.

Due to variable disease manifestation and phenotypic overlap with other IRDs, no pathogenic *ABCA4* variants are found in 10-15% of patients with diagnosis compatible with ABCA4 disease, representing STGD1-like phenocopies caused by mutations in other genes. Both disease-causing *ABCA4* alleles, as required for the unequivocal genetic diagnosis for recessive STGD1, were, until recently, identified in 60-70% of STGD1 patients, and one *ABCA4* mutation in 15-20% of patients.

The emphasis of my research has been to comprehensively identify the complete genetic causality of ABCA4 disease and, more recently, focusing on the genetically unsolved fraction of STGD1 patients with only one identified *ABCA4* pathogenic variant.

Sequencing of the entire 130kb *ABCA4* genomic locus in 233 monoallelic samples to find the second causal variant in *ABCA4* non-coding sequences, identified many rare and unique intronic variants that can potentially exert a pathogenic effect either by causing aberrant splicing in *ABCA4* transcripts or by changing enhancer sequences that influence ABCA4 protein expression. However, no frequent pathogenic variants in non-coding sequences were identified in an ethnically diverse patient cohort of European descent. Based on *in silico* predictions, variant frequency comparisons to general population, evolutionary conservation, and by segregation analyses in the families, we identified 20 strong intronic mutation candidates in 30 patients with one existing definite *ABCA4* mutation. The functional defect of several of these deep intronic candidate variants has already been proven in subsequent *in vitro* splicing assays, confirming our criteria for pathogenic candidate variant selection.

Some disease-associated *ABCA4* variants present with unusually high minor allele frequency (MAF) in general populations. We identified the p.R2107H variant, with MAF~2% in African Americans, as the most frequent (20% of alleles) pathogenic variant

in African American STGD1 patients. The criteria commonly used in defining pathogenic variants suggest general population MAFs not higher than 0.5-1%.

The common missense variant p.N1868I, identified in patients with ABCA4 disease three to four times more frequently than expected, but since this variant is on the same allele with several most frequent *ABCA4* disease-causing variants, and due to its high allele frequency in general populations (MAF 6.6% in non-Finnish Europeans), it was always considered benign. However, we showed that p.N1868I variant was specifically enriched (5X) in unsolved monoallelic STGD1 patients and is a hypomorphic variant, associated with very mild and late onset (mean age=36.3 years) foveal sparing STGD1 phenotype, only in compound heterozygotes with a loss-of-function *ABCA4* variant on the other chromosome. Discovery of the contribution of the frequent p.N1868I hypomorphic variant to ABCA4 disease explained ~50% of previously unsolved monoallelic STGD1 cases and suggests a significant revision of diagnostic screening and assessment of genetic variation in the *ABCA4* gene.

The p.N1868I variant is, to our knowledge, the most frequent in the general population (MAF>6.5%) disease-causing variant described in Mendelian disease, warranting a change in the paradigm of allele frequency-based pathogenicity assessments and careful consideration of seemingly benign and frequent variants in all monogenic diseases, especially for late-onset diseases.

These studies have deciphered the genetic complexity of Stargardt disease and established the *ABCA4* locus as a model for analyzing the genetic causality in not only retinal diseases but all monogenic diseases in general.

KOKKUVÕTE

ABCA4 lookus kui monogeensete haiguste kompleksse geneetika mudel

Pärilikud silma võrkkesta (reetina) düstroofiad koondavad geneetiliselt ja kliiniliselt heterogeense grupi haigusi, mida iseloomustab progressiivne ja pöördumatu nägemisekaotus. Tõestatud on enam kui 300 geeni DNA mutatsioonide põhjuslik seos erinevate pärilike reetina taandarengutega.

Mutatsioonid *ABCA4* geenis põhjustavad kõige sagedamini esinevat pärilikku kollatähni taandarengut, varieeruva haigusfenotüübiga retsessiivset Stargardti haigust (STGD1, MIM #248200). *ABCA4* geenivariantide tagajärjel tekkinud haiguspilt varieerub varajase algusega kiirelt progresseeruvast ja kogu reetinat hõlmavast düstroofiast hilise alguse ja leebete sümptomitega piiratud nägemisekaotuseni, mille haigustunnused võivad teatud määral kattuda enam levinud vanadusega seotud kollatähni taandarenguga (ingl. k. age-related macular degeneration). Fenotüübiline varieeruvus tuleneb ulatuslikust heterogeensusest *ABCA4* geenis, mille eksonites ja splaiss-saitides on kirjeldatud ~1000 erinevat haigust põhjustavat DNA varianti. *ABCA4* mutatsioonide koondsagedus populatsioonis on 5% - 1 inimene 20-st on *ABCA4* võimaliku haigusalleeli kandja. Stargardti haiguse hinnanguline esinemissagedus on siiski 1:8000–1:10 000, kuna kõik mutatsioonide kombinatsioonid ei vii haiguse kujunemiseni.

Suur geneetiline ja fenotüübiline heterogeensus raskendavad Stargardti haiguse ja *ABCA4* geeni analüüsi. Kolm kõige sagedasemat mutatsiooni Euroopa päritolu Stargardti patsientides - p.G1961E, p.[L541P;A1038V], ja p.[G863A;N1868I] hõlmavad kokku ~20% kõigist võimalikest haigusalleelidest Euroopa STGD1 kohordis, samas kui enamik kirjeldatud mutatsioone on väga harvaesinevad. Paljud *ABCA4* variandid on sagedased teatud kindla rassilise ja etnilise kuuluvusega patsientides, samal ajal puuduvad teise päritoluga kohortides.

Varieeruva haiguspildi ja fenotüüpide kattuvuse tõttu teiste reetina düstroofiatega ei leita *ABCA4* mutatsioone 10-15% Stargardti diagnoosiga patsientidel, kelle STGD1-laadne haiguspilt on põhjustatud mutatsioonidest teistes geenides. Mõlemad *ABCA4* haigusalleelid, mida eeldab retsessiivse STGD1 geneetiline diagnoos, tuvastati kuni hiljuti 60-70% Stargardti patsientidel, ja üks *ABCA4* mutatsioon 15-20% patsientidel.

Minu uurimistöö peamine eesmärk on Stargardti haiguse põhjuste täielik välja selgitamine, keskendudes eriti STGD1 patsientidele, kel on *ABCA4* geeni sekveneerimisel leitud vaid üks *ABCA4* haigusalleel. Uurimistöö ühe osana iseloomustasime ka *ABCA4* mutatsioonispektrit ja haiguspildi avaldumist varem vähe uuritud afroameerika STGD1 patsientide kohordis.

Et leida haigust põhjustavaid DNA variante *ABCA4* geeni mittekodeerivates järjestustes, sekveneerisime 130 kb *ABCA4* genoomse lookuse 233-l geneetiliselt 'lahendamata' ühe *ABCA4* mutatsiooniga STGD1 patsiendil, ja registreerisime palju harvu ja unikaalseid mittekodeerivaid DNA variante, mis võivad olla haigust tekitavad põhjustades splaissingu vigu *ABCA4* transkriptsioonil või muutes enhancer elementide järjestusi mõjudes seeläbi ABCA4 valgu ekspressioonile. Sagedasi mittekodeerivaid mutatsioone etniliselt mitmekesises Euroopa päritolu STGD1 kohordis ei tuvastatud. Kombineeritud analüüsi tulemusel, kasutades funktsionaalset efekti ennustavat tarkvara, võrreldes variantide esinemissagedusi patsientide ja üldpopulatsiooni vahel, hinnates DNA järjestuse evolutsioonilist konserveerumist, ja variantide perekondlikul

segregeerumisel haigusfenotüübiga, identifitseeriti 20 tugevat *ABCA4* introni mutatsiooni kandidaati 30-s patsiendis. Mitme kirjeldatud introni variandi funktsionaalne defekt on juba leidnud tõestust *in vitro* splaissing testides, kinnitades kandidaatvariantide selektsioonil rakendatud kriteeriumide õigsust.

Osade *ABCA4* haigus-seoseliste variantide sagedus üldpopulatsioonis on tavatult kõrge. Aafrika päritolu Ameerika STGD1 patsientidel kõige sagedamini esineva (20% haigusalleelidest) *ABCA4* mutatsiooni p.R2107H alleelisagedus on afroameerika üldpopulatsioonis 2%, samas kui mutatsioonide defineerimisel rakendatakse tavaliselt üldpopulatsiooni alleelisageduse piirmäärana 0.5-1%.

Variandi p.N1868I esinemissagedus STGD1 patsientides on 3 korda kõrgem kui 6.6% Euroopa üldpopulatsioonis. Kuna p.N1868I asub samal alleelil mitme sagedase *ABCA4* mutatsiooniga, ja tema kõrge alleelisageduse tõttu üldpopulatsioonis on seda varianti alati loetud polümorfismiks. Näitasime, et p.N1868I variant esines üldpopulatsiooniga võrreldes 5 korda sagedamini ühe *ABCA4* mutatsiooniga patsientides, ja et p.N1868I variandi liitheterosügoot *ABCA4* funktsionaalse null-alleeliga vastaskromosoomil põhjustab hilise algusega (keskmine vanus 36.3 aastat) leebete sümptomitega fooveat säilitavat Stargardti haigust. Sagedase hüpomorfse p.N1868I variandi rolli avastamine Stargardti haiguse kujunemisel selgitas haiguse põhjuse ~50%-l seni geneetiliselt 'lahendamata' ühe *ABCA4* mutatsiooniga STGD1 patsiendil, ja tingib *ABCA4* diagnostilise skriinimise ja variantide interpretatsiooni kriteeriumide olulise ümber hindamise.

p.N1868I variant *ABCA4* geenis on teadaolevalt kõrgeima üldpopulatsiooni alleelisagedusega (>6.5%) tõestatud haigus-seoseline variant monogeensetes haigustes, revideerides alleelisagedusel põhinevate kriteeriumide kehtivat paradigmat variantide patogeensuse hindamisel, eriti haiguste puhul, mis algavad vanemas eas.

Käesoleva uurimistöö tulemused tõestavad Stargardti haiguse geneetilist kompleksust ja *ABCA4* lookuse analüüsi kui mudelit mitte ainult pärilike reetina haiguste, vaid laiemalt monogeensete haiguste geneetilisel analüüsil.

APPENDIX

Publication I

Zernant J, Collison FT, Lee W, Fishman GA, Noupuu K, Yuan B, Cai C, Lupski JR, Yannuzzi LA, Tsang SH, Allikmets R (2014). Genetic and Clinical Analysis of *ABCA4*-Associated Disease in African American Patients. Hum Mutat. 35(10):1187-1194.

RESEARCH ARTICLE

Human Mutation

OFFICIAL JOURNAL
HGVS
HUMAN GENOME VARIATION SOCIETY
www.hgvs.org

Genetic and Clinical Analysis of *ABCA4*-Associated Disease in African American Patients

Jana Zernant,[1] Frederick T. Collison,[2] Winston Lee,[1] Gerald A. Fishman,[2,3] Kalev Noupuu,[1] Bo Yuan,[4] Carolyn Cai,[1] James R. Lupski,[4,5] Lawrence A. Yannuzzi,[6] Stephen H. Tsang,[1,7] and Rando Allikmets[1,7]*

[1]*Department of Ophthalmology, Columbia University, New York, New York;* [2]*The Pangere Center for Hereditary Retinal Diseases, The Chicago Lighthouse for People Who Are Blind or Visually Impaired, Chicago, Illinois;* [3]*Department of Ophthalmology, University of Illinois at Chicago, Chicago, Illinois;* [4]*Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas;* [5]*Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas;* [6]*Vitreous-Retina-Macula Consultants of New York, New York, New York;* [7]*Department of Pathology & Cell Biology, Columbia University, New York, New York*

Communicated by Mark H. Paalman

Received 2 May 2014; accepted revised manuscript 25 June 2014.
Published online 27 July 2014 in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.22626

ABSTRACT: Autosomal recessive Stargardt disease (STGD1) is caused by hundreds of mutations in the *ABCA4* gene, which are often specific to racial and ethnic groups. Here, we investigated the *ABCA4* variation and their phenotypic expression in a cohort of 44 patients of African American descent, a previously under-characterized racial group. Patients were screened for mutations in *ABCA4* by next-generation sequencing and array-comparative genomic hybridization (aCGH), followed by analyses for pathogenicity by in silico programs. Thorough ophthalmic examination was performed on all patients. At least two (expected) disease-causing alleles in the *ABCA4* gene were identified in 27 (61.4%) patients, one allele in 11 (25%) patients, and no *ABCA4* mutations were found in six (13.6%) patients. Altogether, 39 different disease-causing *ABCA4* variants, including seven new, were identified on 65 (74%) chromosomes, most of which were unique for this racial group. The most frequent *ABCA4* mutation in this cohort was c.6320G>A (p.(R2107H)), representing 19.3% of all disease-associated alleles. No large copy number variants were identified in any patient. Most patients reported later onset of symptoms. In summary, the *ABCA4* mutation spectrum in patients of West African descent differs significantly from that in patients of European descent, resulting in a later onset and "milder" disease.

Hum Mutat 35:1187–1194, 2014. Published 2014 Wiley Periodicals, Inc.**

KEY WORDS: ABCA4; Stargardt disease; next-generation sequencing; allelic heterogeneity; African American

Introduction

Mutations in the *ABCA4* gene are responsible for autosomal recessive Stargardt disease (STGD1; MIM #248200) [Allikmets et al., 1997], cone-rod dystrophy (CRD) [Cremers et al., 1998; Maugeri et al., 2000] and retinitis pigmentosa phenotypes [Cremers et al., 1998; Martinez-Mir et al., 1998; Shroyer et al., 2001]. *ABCA4*-associated diseases present with an extensive clinical and genetic heterogeneity, the current count of disease-associated *ABCA4* variants exceeds 800 [Allikmets, 2007] (RA and JZ, unpublished data). The most frequent disease-associated *ABCA4* variants have each been described in only ~10% of STGD1 patients of European descent [Burke et al., 2012]. However, many variants are more common in patients with specific geographic and ethnic backgrounds, such as the c.2588G>C (p.[G863A,G863del]) founder mutation in Northern European patients, [Maugeri et al., 1999] the c.[1622T>C; 3113C>T] (p.[L541P; A1038V]) complex allele in patients of mostly German origin [Maugeri et al., 2000; Rivera et al., 2000], the c.3386G>T (p.R1129L)founder mutation in Spain [Valverde et al., 2006], the c.2894A>G (p.N965S) variant in the Danish population [Rosenberg et al., 2007], and the c.5318C>T (p.A1773V) variant in Mexico [Chacon-Camacho et al., 2013].

Complete sequencing of the *ABCA4* coding and adjacent intronic sequences in patients with STGD1 routinely discovers ~75%–80% of mutations with the fraction of patients harboring the expected two disease-associated alleles comprising 65%–70%, with one mutation 15%–20%, and with no mutations in the remaining ~15% [Zernant et al., 2011]. These fractions depend on many variables, most importantly the quality of the clinical diagnosis and the ethnic composition of the cohort.

It is therefore clear that proper genetic diagnosis and interpretation of *ABCA4* alleles in ethnic and racial groups relies on many factors starting with a comprehensive database of disease-associated variants. Interestingly, there has been limited number of studies on STGD1 patients of African American descent. Only one case report [Huynh et al., 2014] and a small case series [Utz et al., 2013] have been described. The current study was designed to comprehensively address this issue by analyzing the clinical findings and, especially, the genetic composition of *ABCA4*-associated disease in African Americans.

Materials and Methods

Patients

Patients (44) affected with STGD1 (41), including seven with bull's eye phenotype, and CRD (three) were, after consenting in writing, recruited and clinically examined over a 10-year period at the Department of Ophthalmology, Columbia University, at the

*Correspondence to: Rando Allikmets, Columbia University, Eye Institute Research, Rm. 202, 160 Fort Washington Avenue, New York, NY 10032. E-mail: rla22@columbia.edu

Contract grant sponsors: National Eye Institute/NIH (EY021163, EY019861, and EY019007); Core Support for Vision Research; Foundation Fighting Blindness (Owings Mills, Maryland); Research to Prevent Blindness (New York, NY); Pangere Family Foundation, Chicago Lighthouse.

University of Illinois at Chicago, and at the Pangere Center for Hereditary Retinal Diseases, The Chicago Lighthouse for People Who Are Blind or Visually Impaired. Onset decade was defined by the age at which symptoms were first reported. Visual acuity was measured using the Early Treatment Diabetic Retinopathy Study (EDTRS) Chart 1 or a Snellen acuity chart. Diagnosis was based on patient history and ophthalmic examination. Fundus photography, fundus autofluorescence (AF), spectral domain-optical coherence tomography (SD-OCT) (Heidelberg Spectralis, Heidelberg, Germany, or OPTOS Instrumentation, Marlborough, MA), and full-field electroretinography (ERG) were performed as needed using standard acquisition protocols following pupil dilation with tropicamide 1% and phenylephrine 2.5%.

Patients diagnosed with STGD1 were classified into one of the four stages as described previously [Fishman et al., 1999]. Stage I disease was characterized by parafoveal or perifoveal flecks with frequent atrophy of the central macula. Stage II disease was characterized by flecks that were more numerous and extended anterior to the vascular arcades and/or nasal to the optic disc, Stage III was defined by resorption of flecks. Choriocapillaris atrophy within the macula was often observed in stage III patients. Stage IV disease, characterized by widespread RPE and chorioretinal atrophy throughout the fundus, was not observed in our cohort of Stargardt disease patients [Fishman et al., 1999].

All genetic research was carried out with the approval of the Institutional Review Boards of Columbia University and the University of Illinois at Chicago, and in accordance with the Declaration of Helsinki.

Array Screening

Screening with the *ABCA4* array was performed on most study subjects followed by direct Sanger sequencing to confirm identified changes, as previously described [Jaakson et al., 2003]. Since the array screening had been performed over many years, different versions of the *ABCA4* chip had been used, from the least representative (~300 mutations) to the recent version of the array (>600 variants).

Sequencing and Data Analysis

All 50 exons and exon–intron boundaries of the *ABCA4* gene were amplified using Illumina TruSeq Custom Amplicon protocol (Illumina, San Diego, CA), followed by sequencing on Illumina MiSeq platform. The next-generation sequencing reads were analyzed and compared to the *ABCA4* reference sequence NG_009073.1, using the variant discovery software NextGENe (SoftGenetics LLC, State College, PA). All detected possibly disease-associated variants were confirmed by Sanger sequencing. Nucleotide positions and protein translation correspond to CCDS747.1 and NP_000341.2, respectively. Nucleotide numbering uses the A of the ATG translation initiation start site as nucleotide 1.

The allele frequencies of all variants were compared to the Exome Variant Server (EVS) dataset, NHLBI Exome Sequencing Project, Seattle, WA (http://snp.gs.washington.edu/EVS/; accessed March 2014). All *ABCA4* variants reported in this manuscript were submitted to the ABCA4 LSDB (http://www.lovd.nl/ABCA4) at the Leiden Open Variation Database 3.0 (http://www.lovd.nl/3.0/home).

Segregation of the new variants with the disease was analyzed in families if family members were available (Table 1).

The possible effect of all *ABCA4* variants was assessed using the combination of following prediction programs: Polyphen-2 [Adzhubei et al., 2010], Align-GVGD [Tavtigian et al., 2006], SIFT [Ng and Henikoff, 2001], MutationTaster [Schwarz et al., 2010], SpliceSiteFinder [Zhang, 1998], MaxEntScan [Yeo and Burge, 2004], NNSPLICE [Reese et al., 1997], GeneSplicer [Pertea et al., 2001], and Human Splicing Finder [Desmet et al., 2009]. All of these algorithms except for Polyphen-2 were accessed via Alamut software version 2.2 (Interactive Biosoftware, Rouen, France; http://www.interactive-biosoftware.com), using automated computation of this software version. Polyphen-2 results were retrieved from the single entry Web form (http://genetics.bwh.harvard.edu/pph2/) with the HumDiv-model and version 2.2.2 of the software. Predictions are included as supporting data to the conclusions that are based on different frequencies of the variants between the patients and general population.

Array-Comparative Genome Hybridization

Custom array-comparative genome hybridization (aCGH) (Agilent Technologies, CA, United States) with high-density probes tiling the critical genetic loci of ABCA4-associated disease was designed to assess for copy number variations (CNVs) involving these genes. Each array slide was in 8 × 60K format, which tested eight separate samples with approximate 60,000 probes for each sample. Nine "primary" loci, including the *ABCA4* locus, other known genes often mimicking ABCA4-associated phenotypes (*ELOVL4*, *PRPH2*, *BEST1*, *RS1*, and *CNGB3*), and major age-related macular degeneration (AMD)-associated loci (*CFH*, *ARMS2*, and *C2/CFB*), were probed with ultra-high density (~20 probes per kilobase [kb] of DNA) throughout the entire genomic length of the genes plus the 5′ promoter regions and 3′ downstream regions, as well as at slightly lower density (~10 probes per kb) for flanking 5′ and 3′ conserved regions. Eighteen "secondary" AMD loci (*C3*, *APOE*, *CFI*, *LIPC*, *SYN3/TIMP3*, *CETP*, *COL8A1*, *BBX*, *PLD1*, *SPEF2*, *ADAM19*, *VEGFA*, *FRK*, *MEPCE*, *CHMP7/LOXL2*, *TGFBR1*, *NPS*, and *PICK1*) were probed with ultra-high density in exons (~8 probes per exon) and with lower resolution (~3 probes per kb) throughout each gene plus the flanking 5′ promoter regions and 3′ downstream regions.

DNA from 22 of the 44 African American patients diagnosed with *ABCA4*-associated disease, five with one mutation and 17 with two mutations in ABCA4, were used for aCGH analysis. For a positive control the DNA carrying a known, previously reported 1,030 bp heterozygous deletion of exon 18 in ABCA4, was used [Yatsenko et al., 2003]. Experimental procedures of aCGH were performed as described previously [Gonzaga-Jauregui et al., 2010] with minor modifications, and results were analyzed using Agilent Genomic Workbench version 7.0 software (Agilent Technologies, CA, United States). The called CNVs were filtered against several criteria and the plausible true-positive CNVs were tested by PCR for molecular validation.

Comparison with Patients of European Ancestry

The genetic and clinical data from patients of West African descent were compared with a Columbia in-house database of *ABCA4*-associated patients of European descent, which includes 244 patients. Comparisons were made for allele frequencies across patient and general populations, predicted or tested functional effect of selected mutations, disease onset, Fishman stage, and so on [Fishman et al., 1999].

Results

ABCA4 Mutations in African American Patients

After genetic analysis with a combination of *ABCA4* array screening and/or direct sequencing both *ABCA4* disease-causing alleles

Table 1. Pathogenic *ABCA4* Variants in Patients of African American Descent

Patient	Clinical Assessment			BCVA		CNV	
ID	*Diagnosis*	*Stage*	*Onset decade*	*OD*	*OS*	analysis	ABCA4 mutations
1	STGD1	2	8th	20/25	20/25		p.[(P1380L)];[?]
2	STGD1	2	5th	10/100	10/160	X	p.[(R220C(;)R2107H)]
3	STGD1	1	4th	20/30	20/200		p.[(G1961E(;)D403V)]
4	STGD1	2–3	3rd	20/400	20/100	X	p.[(W1408L(;)R2107H)]
5	STGD1	0	5th	20/200	20/50	X	p.[(P62S)];[?]
6	STGD1	1	7th	20/25-2	20/25-3	X	p.[(R2107H)];[?]
7	STGD1	N/A	4th	20/70+2	20/25-2	X	p.[(P309R(;)R681∗)]
8	STGD1	1	4th	20/15-3	20/20+1	X	p.[(G991R(;)L1126P)]
9	STGD1	N/A	2nd	20/400	20/400	X	c.5289del(;)p.(V989A)
10	STGD1	N/A	N/A	N/A	N/A		p.[(R2077G(;)R2107H)]
11	STGD1	2	3rd	20/200	20/200	X	c.3523-1G>A(;)p.(R2107H)
12	STGD1	2	4th	20/20	20/20	X	p.[(G991R)];[?]
13	STGD1	1	4th	20/70+2	20/25-2		no mutations found
14	STGD1	3	1st	20/400	20/400	X	c.768G>T(;)p.(R2107H)
15	STGD1	1	5th	20/50-2	20/40	X	p.[(V989A(;)R2107H(;)P870S)]
16	STGD1	2–3	1st	20/400	20/300	X	p.[(G607R(;)R2040Q)]
17[a]	STGD1	1	2nd	10/100-1	10/60-1	X	p.[(P309R(;)R537C(;)R2107H)]
18	STGD1	2	1st	20/400	20/400	X	c.5461-10T>C(;)p.(R2040Q)
19	STGD1	2	2nd	10/120+1	10/100-1		c.5461-10T>C(;)p.(E531G)
20	STGD1	1	3rd	20/80-2	20/125+1		No mutations found
21[a]	STGD1 (BEM)	1	2nd	20/100	20/150	X	p.[(G991R)];[(L1138P)]
22[a]	STGD1	2	3rd	20/400	20/50-2		c.4537dup;p.(V1686M)
23	STGD1	2	2nd	20/125	20/125		p.[(R1640W)];[?]
24	CRD	3	1st	CF	CF		p.[(T983A(;)L1729P)]
25	STGD1	1	4th	20/30+2	20/40-2	X	p.[(V989A)];[(V989A)]
26	STGD1	3	2nd	20/150	20/25		p.[(N965S(;)R2040Q)]
27[a]	CRD	3	1st	CF	CF	X	c.4540-2A>G;p.(R2107H)
28	STGD1	2	6th	20/30	20/40-1	X	c.5461-10T>C;[?]
29	STGD1	1	6th	20/20	20/20	X	p.[(F1015L)];[?]
30	STGD1 (BEM)	1	5th	20/50	20/30		No mutations found
31	STGD1 (BEM)	1	6th	20/20	20/20		p.[(R2107H)];[?]
32	STGD1 (BEM)	1	3rd	20/20	20/20	X	p.[(R1300∗(;)R2106C)]
33	STGD1 (BEM)	1	5th	20/100	20/125		p.[(R2107H)];[?]
34	STGD1	3	3rd	LP	LP		p.[(R2107H)];[?]
35	STGD1	2	5th	20/40+1	20/40	X	p.[(W339G(;)R2107H)]
36	STGD1	3	1st	20/30	20/40		No mutations found
37	STGD1	1	5th	20/25	20/25-1		p.[(I975M(;)K1978E)]
38	STGD1	2	2nd	20/200	20/200		p.[(V989A)];[?]
39	STGD1	1	3rd	20/100-1	20/200+1		c.302+1G>A(;)p.(R2107H)
40	STGD1 (BEM)	1	7th	20/20-3	20/20-3		p.[(R2107H)];[(R2107H)]
41	STGD1	1	3rd	30/30-2	20/100		No mutations found
42	STGD1 (BEM)	1	5th	20/40-2	20/50-2		No mutations found
43	CRD	N/A	3rd	5/180+1	5/160+1	X	p.[(V989A)];[?]
44[a]	STGD1	2	N/A	N/A	N/A		c.4253+4C>T(;)p.[(G1961E(;)R2107H)]

Nucleotide positions and protein translation correspond to CCDS747.1 and NP_000341.2, respectively. Nucleotide numbering uses the A of the ATG translation initiation start site as nucleotide 1.
BCVA: best corrected visual acuity; CNV: copy number variation; N/A: not available; CF: counting fingers; LP: light perception.
[a]The variants are confirmed on different chromosomes.

were identified in 27/44 (61.4%) of African American patients, one disease-causing allele in 11 (25%) patients, and no *ABCA4* mutations were found in six (13.6%) patients. Altogether, 65 (73.9%) disease-causing *ABCA4* alleles were identified in the cohort of 44 patients of African American origin presenting with clinically diagnosed *ABCA4*-associated disease. These data correlate well with the comparative numbers/fractions of patients of mostly, or exclusively, European descent. In a cohort of 496 patients presenting with likely *ABCA4*-associated diseases of mostly European origin, we identified 698 *ABCA4* mutations (70.4%), whereas in the cohort of 244 patients of definitely European origin, this fraction was 352 (72.1%).

The combination of array screening and sequencing identified 47 different *ABCA4* variants in the 44 patients (Tables 1–3). Eight variants (Tables 2 and 3) were deemed benign based on similar allele frequencies between patients and the general population of African American descent (data derived from screening of 2203 individuals on the EVS), and/or by predictive programs. These include frequent variants c.3602T>G (p.L1201R), c.3899G>A (p.R1300Q), c.1927G>A (p.V643M), and c.4925C>A

Table 2. Analysis of the *ABCA4* (GenBank Reference Sequence NG_009073.1) Variants by Predictive Programs

Exon/ intron	Nucleotide change	Protein variant	Polyphen-2 prediction	AGVGD class	SIFT prediction	TASTER prediction	Predicted effect on splicing
3	c.184C>T	p.P62S	Probably damaging	C65	Deleterious	Disease causing	
IVS3	c.302+1G>A						Eliminates the donor
6	c.658C>T	p.R220C	Possibly damaging	C25	Deleterious	Polymorphism	
6	c.768G>T						Donor decrease by 47%
8	c.926C>G	p.P309R	Benign	C65	Deleterious	Disease causing	
8	c.1015T>G	p.W339G	Possibly damaging	C65	Deleterious	Disease causing	
9	**c.1208A>T**	**p.D403V**	Benign	C65	Deleterious	Disease causing	
11	c.1538T>C	p.V513A[a]	Benign	C0	Tolerated	Polymorphism	
12	**c.1592A>G**	**p.E531G**	Probably damaging	C65	Deleterious	Disease causing	
12	c.1609C>T	p.R537C	Probably damaging	C45	Deleterious	Disease causing	
13	c.1819G>A	p.G607R	Probably damaging	C65	Deleterious	Disease causing	
13	c.1927G>A	p.V643M[a]	Probably damaging	C15	Deleterious	Disease causing	
14	c.2041C>T	p.R681*					
16	c.2546T>C	p.V849A	Benign	C25	Deleterious	Polymorphism	
17	**c.2608C>T**	**p.P870S**	Possibly damaging	C65	Deleterious	Disease causing	
19	c.2791G>A	p.V931M[a]	Possibly damaging	C0	Deleterious	Disease causing	
19	c.2894A>G	p.N965S	Probably damaging	C45	Deleterious	Disease causing	New acceptor site
20	**c.2925C>G**	**p.I975M**	Possibly damaging	C0	Deleterious	Disease causing	
20	c.2947A>G	p.T983A	Probably damaging	C55	Deleterious	Disease causing	New acceptor site
20	c.2966T>C	p.V989A	Benign	C25	Deleterious	Disease causing	
20	c.2971G>C	p.G991R	Probably damaging	C65	Deleterious	Disease causing	
20	**c.3043T>C**	**p.F1015L**	Possibly damaging	C15	Deleterious	Disease causing	
23	c.3377T>C	p.L1126P	Probably damaging	C65	Deleterious	Disease causing	
23	c.3413T>C	p.L1138P	Probably damaging	C65	Deleterious	Disease causing	
23	c.3416T>C	p.Y1139C[a]	Benign	C55	Deleterious	Disease causing	
IVS23	c.3523-1G>A						Eliminates the acceptor
24	c.3602T>G	p.L1201R[a]	Benign	C65	Deleterious	Disease causing	
27	c.3898C>T	p.R1300*					
27	c.3899G>A	p.R1300Q[a]	Benign	C0	Tolerated	Polymorphism	
28	c.4139C>T	p.P1380L	Probably damaging	C65	Deleterious	Disease causing	
28	c.4223G>T	p.W1408L	Probably damaging	C55	Deleterious	Disease causing	
IVS28	c.4253+4C>T						Weakens the donor by 14%
30	c.4537dup	p.Q1513fs					
IVS30	c.4540-2A>G						Eliminates the acceptor
35	c.4918C>T	p.R1640W	Probably damaging	C65	Deleterious	Disease causing	
35	c.4925C>A	p.S1642I[a]	Benign	C35	Deleterious	Polymorphism	
36	c.5056G>A	p.V1686M	Probably damaging	C15	Deleterious	Disease causing	
36	c.5077G>A	p.V1693I[a]	Benign	C0	Tolerated	Polymorphism	
36	c.5186T>C	p.L1729P	Possibly damaging	C0	Deleterious	Disease causing	
37	**c.5289del**	**p.V1764fs**					
IVS38	c.5461-10T>C						Effect unknown
42	c.5882G>A	p.G1961E	Probably damaging	C65	Deleterious	Disease causing	
43	**c.5932A>G**	**p.K1978E**	Probably damaging	C55	Deleterious	Disease causing	
44	c.6119G>A	p.R2040Q	Probably damaging	C35	Deleterious	Disease causing	
45	c.6229C>G	p.R2077G	Probably damaging	C65	Deleterious	Disease causing	
46	c.6316C>T	p.R2106C	Probably damaging	C65	Deleterious	Disease causing	
46	c.6320G>A	p.R2107H	Probably damaging	C25	Deleterious	Disease causing	

Nucleotide positions and protein translation correspond to CCDS747.1 and NP_000341.2, respectively. Nucleotide numbering uses the A of the ATG translation initiation start site as nucleotide 1.
Novel *ABCA4* variants are indicated **in bold**.
Prediction programs (except for Polyphen-2) were accessed via Alamut software version 2.2 (http://www.interactive-biosoftware.com), using automated computation of this software version. Polyphen-2 results were retrieved from the single entry Web form (http://genetics.bwh.harvard.edu/pph2/) with the HumDiv-model and version 2.2.2 of the software.
[a]Benign variants.

(p.S1642I). Several known frequent missense or silent variants, including c.635G>A (p.R212H), c.1268A>G (p.H423R), c.1269C>T, c.2828G>A (p.R943Q), c.4203C>A, c.5603A>T (p.N1868I), c.5814A>G, c.5843C>T (p.P1948L), c.5844A>G, c.6069C>T, c.6249C>T, c.6285T>C, c.6732G>A, and c.6764G>T (p.S2255I), which have been previously defined as benign, were not included in the analysis. Although the allele frequencies of some of these variants are very different between the general populations of European and African descent, these are not different between the affected and unaffected individuals in each ethnic group. Seven *ABCA4* variants, c.6320G>A (p.(R2107H)), c.2966T>C (p.(V989A)), c.2971G>C (p.(G991R)), c.5461–10T>C, c.6119G>A (p.(R2040Q)), c.926C>G (p.(P309R)), and c.5882G>A (p.(G1961E)), were identified in two or more patients, 32 *ABCA4* variants were each detected only once (Tables 1 and 3).

Five mutations, c.1208A>T (p.(D403V)), c.2608C>T (p.(P870S)), c.3043T>C (p.(F1015L)), c.52089del (p.(V1764fs)), and c.5932A>G (p.(K1978E)), had not been detected before, and two mutations, c.1592A>G (p.(E531G)) and c.2925C>G (p.(I975M)), had not been described as associated with the disease. Interestingly, no large (i.e., those undetectable with PCR-based sequencing methods, usually one exon or larger) copy number variants in the *ABCA4* locus were identified in any patient.

Table 3. ABCA4 Variants Found in Patients of African American Origin, Ranked by Allele Frequency and Compared with the Frequency in the Patients of European Descent

	African origin cohort (44)			European origin cohort (244)		
Variant	Alleles	Allele frequency	Frequency by EVS	Alleles	Allele frequency	Frequency by EVS
p.R2107H	17	19.32	2.04	5	1.02	0.01
p.L1201R[a]	12	13.64	9.35	1	0.20	0.05
p.V989A	6	6.82	0.25	1	0.20	
p.R1300Q[a]	5	5.68	6.17	2	0.41	0.05
p.V643M[a]	3	3.41	1.82			
p.G991R	3	3.41	0.64	2	0.41	
p.S1642I[a]	3	3.41	1.23			
p.R2040Q	3	3.41	0.05			0.01
c.5461-10T>C	3	3.41		23	4.71	0.03
p.P309R	2	2.27	0.16			
p.G1961E	2	2.27	0.11	55	11.27	0.42
p.P62S	1	1.14				
c.302+1G>A	1	1.14				
c.768G>T	1	1.14				
p.R220C	1	1.14				
p.W339G	1	1.14				
p.D403V	1	1.14				
p.V513A[a]	1	1.14				
p.E531G	1	1.14	0.09			
p.R537C	1	1.14	0.02			
p.G607R	1	1.14	0.02			
p.R681*	1	1.14				
p.V849A[a]	1	1.14	1.23			
p.P870S	1	1.14				
p.N965S	1	1.14		3	0.61	0.01
p.V931M[a]	1	1.14	0.41			
p.F1015L	1	1.14				
p.I975M	1	1.14	0.02			
p.T983A	1	1.14		1	0.20	
p.L1126P	1	1.14				
p.L1138P	1	1.14				
p.Y1139C[a]	1	1.14	0.09			
c.3523-1G>A	1	1.14				
p.R1300*	1	1.14				
p.P1380L	1	1.14		26	5.33	0.02
p.W1408L	1	1.14				
c.4253+4C>T	1	1.14		1	0.20	
c.4537dup	1	1.14		1	0.20	
c.4540-2A>G	1	1.14		3	0.61	
p.R1640W	1	1.14		3	0.61	
p.L1729P	1	1.14				
p.V1686M	1	1.14	0.07	1	0.20	0.03
p.V1693I[a]	1	1.14	0.18			
c.5289del	1	1.14				
p.K1978E	1	1.14				
p.R2077G	1	1.14		1	0.20	
p.R2106C	1	1.14	0.05	1	0.20	

Frequency by EVS—frequency of the variant in the general populations of, respectively, African American and European descent as determined in the Exome Variant Server (http://snp.gs.washington.edu/EVS/; accessed March 2014).
[a]Benign variants.

Analysis of Frequent Variants

From more frequent (i.e., represented in three or more patients) variants from both ethnic groups, the intronic disease-associated variant with yet unknown function, c.5461–10T>C, was the only variant detected in both groups with similar allele frequencies, 3.4% in the African American and 4.71% in the European patient group (Table 3). It was also the only frequent variant not seen in the general population of West African descent. The other six frequent variants in African American patients—p.R2107H, p.V989A, p.G991R, p.R2040Q, p.P309R, and p.G1961E—were seen in the matched general population at frequencies between 0.05% and 2% (Table 3). However, their respective frequencies in the patient cohort were at least 10× higher (Tables 2 and 3), therefore strongly suggesting their association with the disease.

The most frequent *ABCA4* mutation in patients of West African descent was p.R2107H, accounting for 19.3% of all disease-associated alleles in this cohort, which is much higher than for any other known *ABCA4* disease-associated variant. By comparison, the most frequent *ABCA4* mutation in patients of European ancestry—p.G1961E—has an average allele frequency of ~11% [Burke et al., 2012]. Another, somewhat surprising, observation is that the frequency of the p.R2107H in the general population of African Americans is 2% (Table 3), which is much higher than the suggested and observed frequency for highly penetrant disease-associated alleles in general. It also suggests that one in 2,500 African Americans

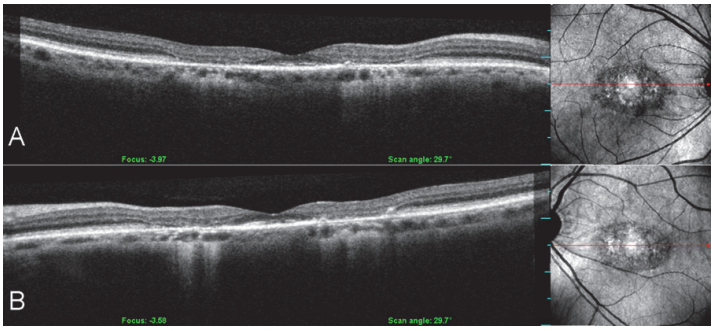


Figure 1. Clinical presentation of STGD1 in patient 40. Spectral-domain optical coherence tomography horizontal line scans through the fovea of the right **(A)** and left **(B)** eyes of patient 40, who was homozygous for the p.R2107H mutation. The inner-segment ellipsoid and outer nuclear layers are absent in areas of the parafovea, but both layers are present in the fovea. Hyper-reflectance seen in the parafovea on the infrared scanning laser ophthalmoscopy images (right side of **A** and **B**) corresponds to increased reflectance from the choroid and sclera on SD-OCT, which is suggestive of RPE and choriocapillaris atrophy in those areas.

is homozygous for the p.R2107H variant, implicating a much higher than the suggested prevalence (1:10,000) of *ABCA4*-associated disease [Blacharski, 1988]. Furthermore, if the variant is fully penetrant, 4% of STGD patients of African descent should be homozygous for the variant; one out of 44 (2.3%) patients was in our cohort. Finally, the p.R2107H variant was called "disease-causing" by all predictive programs (Table 2) and has been previously shown to alter ATP hydrolysis in vitro [Sun et al., 2000]. Altogether, these data suggest that the p.R2107H variant is a highly penetrant disease-associated allele in patients of West African descent presenting with *ABCA4*-associated diseases.

Analysis of Rare Variants

Among 36 variants detected only once, 32 were classified as disease-associated by predictive programs and/or segregation analyses (Table 2). Variants c.1538T>C (p.(V513A)), c.2546T>C (p.(V849A)), c.2791G>A (p.(V931M)), c.5077G>A (p.(V1693I)), and possibly, c.3416T>C (p.(Y1139C)) were deemed benign. Interestingly, of the 32 variants only 10 had been seen in large patient cohorts of European descent suggesting that 22 (more than 2/3) variants are specific for patients of West African descent. The clinical expression of the disease in patients with rare mutations was very heterogeneous, as expected. The onset, expression and progression depend, as in patients of European descent, on the specific combinations of severe and mild *ABCA4* disease-associated alleles. Since most mutations were detected only once in this group, no specific genotype/phenotype correlations were possible. Seven patients carried definitely deleterious mutation (~8% from all alleles), which is two times less than, on average, seen in cohorts of European descent (~15%). This may also explain the milder, later onset disease in African American patients, as discussed below.

Patient Phenotypes

Patient #40 was homozygous for the p.R2107 variant. He was seen at the age of 61 years, at which point he had noted subtle color vision changes for the past year. Best corrected Snellen visual acuity was $20/20^{-3}$ in each eye. Fundus examination showed a bull's eye-appearing macular lesion, with a notable absence of fundus flecks and relative foveal sparing. SD-OCT (Fig. 1) showed loss of the inner-segment ellipsoid layer (ISe) and outer nuclear layer (ONL) in the parafoveal area, with relative sparing of those layers in the fovea.

Patient #25 was homozygous for the p.V989A variant. When seen at the age of 39 years, she reported blurred central vision, mild color vision changes, and some difficulty seeing in dim illumination. Her best corrected visual acuities at the most recent examination, when she was 42 years of age, were $20/30^{+2}$ OD and $20/40^{-2}$ OS. On fundus examination, she had bull's eye-appearing macular lesions with a ring of flecks confined to the macula (stage 1) and relative foveal sparing (Fig. 2) in both eyes. SD-OCT showed ISe and ONL loss in the parafoveal area, with relative sparing of those layers in the fovea (Fig. 2). Short wavelength FAF showed a ring of hypo-AF, a few highly hypo-autofluorescent areas of atrophy, and hyper-autofluorescent flecks in the macula. Full-field ERG scotopic and photopic amplitudes were within normal limits.

Fifteen patients were heterozygous for the p.R2107H allele; in four patients, this was the only *ABCA4* allele found, 11 patients were compound heterozygous for another disease-associated variant. The clinical presentation of the disease varied widely in these patients most likely depending on the severity of the other allele. For example, patients with clearly deleterious splicing mutations on the other chromosome (c.3523–1G>A in patient #11, c.768G>T in patient #14, and c.4540–2A>G in patient #27) had an earlier disease onset and rapid progression, whereas patients with a "milder" second mutation had later onset and slower disease progression. As an example from this group, patient #15 was compound heterozygous for p.R2107H and p.V989A mutations and also had a third variant, p.P870S. She was seen at the age of 49 years, at which time she reported decreased central vision for the past year, as well as difficulty with night driving for the past 5 years. Her best corrected visual acuity was $20/50^{-2}$ OD and 20/40 OS. Fundus examination showed bull's eye-appearing lesions in both eyes, with a ring of flecks in the macula, and mild temporal optic disc pallor, but otherwise normal optic nerves and retinal vessels.

Comparison with Patients of European Descent

Of the six most frequent disease-associated variants (p.R2107H, p.V989A, p.G991R, c.5461–10T>C, p.R2040Q, and p.P309R) in African American patients, only the intronic c.5461–10T>C

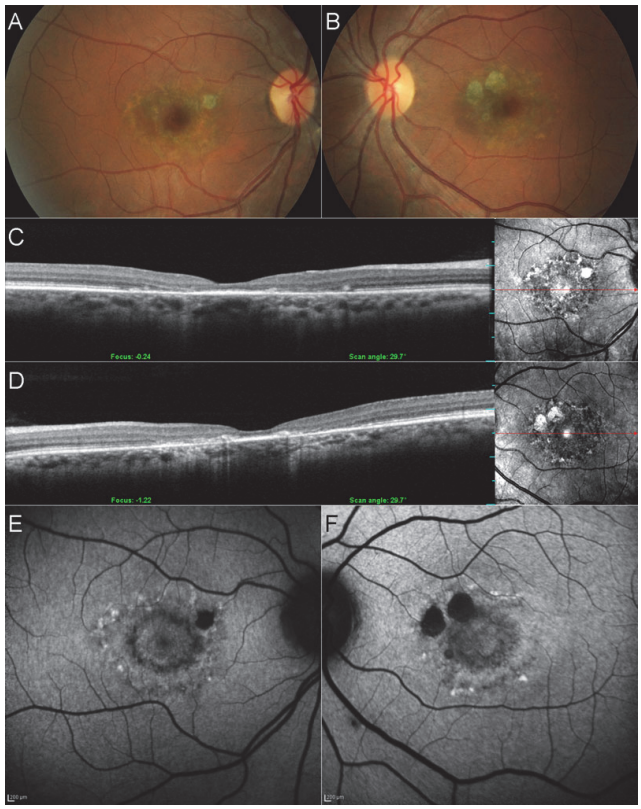


Figure 2. Clinical presentation of STGD1 in patient 25. Fundus photos of the right **(A)** and left **(B)** eyes of patient 25 (homozygous for p.V989A mutations), show bull's eye-appearing macular lesions with macular flecks and well-demarcated atrophic areas in both eyes. Horizontal spectral-domain optical coherence tomography **(C and D)** line scans show loss of inner-segment ellipsoid and outer nuclear layers in the parafovea of the each eye. Corresponding infrared scanning laser ophthalmoscopy images (right side of **C and D**) shows some hypo-reflectance in the parafoveal areas, with hyper-reflectance primarily in the areas of the demarcated atrophic lesions. Short-wavelength fundus autofluorescence **(E and F)** highlights bull's eye-appearing macular lesions in each eye. Macular flecks are hyper-autofluorescent, while localized loss of **RPE** cells are evidenced by the well demarcated hypo-autofluorescent regions.

variant was seen at equal frequency in patients of European descent (Table 3). The p.R2107H, p.V989A, and p.G991R variants were seen in Caucasian patients at much lower frequencies; in fact, the allele frequencies for these variants were lower in patients of European descent than in the general population of West African descent (Table 3). Although these variants were proven to represent disease-associated mutations, their presence in Caucasian patients most likely suggests admixture between ethnic groups. The same is, conversely, true for the two most frequent mutations in Caucasian patients, c.5882G>A (p.(G1961E)) and c.4138C>T (p.(P1380L)), which are (almost) absent in the general population of West African origin (Table 3). These mutations were detected in two and one African American patients, respectively, suggesting admixture. All other most prevalent disease-associated *ABCA4* variants in the cohort of European descent, including c.[1622T>C; 3113C>T] (p.[L541P;A1038V]), c.2588G>C (p.G863A), c.5714+5G>A, c.3322C>T (p.R1108C), c.6079C>T (p.L2027F), and c.161G>A (p.C54Y), were not seen in African American patients at all.

The clinical presentation of the disease differed significantly between the two ethnic groups with patients of West African descent having later onset of the disease. When compared with randomly selected 100 STGD1/CRD patients of Caucasian descent from our database, only 14% of patients of West African descent had the disease onset in the 1st decade, compared with 25% of white patients. Most interestingly, close to 30% of African American patients had the disease onset in the 5th decade or later, whereas this fraction in the Caucasian cohort was only 12% ($P = 0.03$). Detailed analysis of the disease progression was available in only a limited number of patients precluding statistical analysis.

Discussion

Genetic analysis of 44 African American patients revealed that *ABCA4* mutations in patients of this ethnic background differ to a large extent from patients with *ABCA4*-associated diseases of European descent. The mutation spectrum was different on

several levels: First, most mutations in patients of West African descent were unique, or more specific, for this ethnic group. Very different distribution of *ABCA4* disease-associated alleles was true for both frequent and rare variants. Second, several *ABCA4* variants that had been previously classified as disease-causing in patients of European descent were deemed benign since they represent ancestral alleles as determined by evolutionary analyses, and by comparison of allele frequencies in African American patients with *ABCA4*-associated diseases and the general population. Conversely, several variants that were previously considered not associated with the disease, since these were very frequent (up to 2%) in the general population of African American origin, were proven to be causal in *ABCA4*-associated diseases since the allele frequencies for these variants were at least 10× higher in affected individuals than in the general population. This observation was further confirmed by segregation analyses in families and predictive programs.

The differences in genetic causality were also clearly reflected in the clinical phenotypes. The disease in African American patients had a statistically significantly later onset and a trend toward milder expression. Further studies are necessary to establish possible differences in the rate of disease progression. In summary, the spectrum of *ABCA4* mutations in patients of West African descent differs significantly from that in patients of European origin. In addition, the phenotypic presentation of *ABCA4* alleles in African Americans is also different from that in patients of European ancestry. This study confirms extensive genetic and a degree of clinical heterogeneity of *ABCA4*-associated diseases in patients of West African descent and offers guidelines for interpretation of both clinical and, especially, genetic analyses to correctly diagnose African American patients with *ABCA4*-associated diseases.

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. Nat Methods 7(4):248–249.

Allikmets R. 2007. Stargardt disease: from gene discovery to therapy. In: Tombran-Tink J, Barnstable CJ, editors. *Retinal degenerations: biology, diagnostics and therapeutics.* Totowa, NJ: Humana Press. p 105–118.

Allikmets R, Singh N, Sun H, Shroyer NF, Hutchinson A, Chidambaram A, Gerrard B, Baird L, Stauffer D, Peiffer A, Rattner A, Smallwood P, et al. 1997. A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. Nat Genet 15(3): 236–246.

Blacharski P. 1988. In: Newsome DA, editor. *Retinal dystophies and degenerations.* New York: Raven Press. p 135–159.

Burke TR, Fishman GA, Zernant J, Schubert C, Tsang SH, Smith RT, Ayyagari R, Koenekoop RK, Umfress A, Ciccarelli ML, Baldi A, Iannaccone A, et al. 2012. Retinal phenotypes in patients homozygous for the G1961E mutation in the ABCA4 gene. Invest Ophthalmol Vis Sci 53(8):4458–4467.

Chacon-Camacho OF, Granillo-Alvarez M, Ayala-Ramirez R, Zenteno JC. 2013. 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.

Cremers FP, van de Pol DJ, van Driel M, den Hollander AI, van Haren FJ, Knoers NV, Tijmes N, Bergen AA, Rohrschneider K, Blankenagel A, Pinckers AJ, Deutman AF, et al. 1998. Autosomal recessive retinitis pigmentosa and cone-rod dystrophy caused by splice site mutations in the Stargardt's disease gene ABCR. Hum Mol Genet 7(3):355–362.

Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 37(9):e67.

Fishman GA, Stone EM, Grover S, Derlacki DJ, Haines HL, Hockey RR. 1999. Variation of clinical expression in patients with Stargardt dystrophy and sequence variations in the ABCR gene. Arch Ophthalmol 117(4):504–510.

Gonzaga-Jauregui C, Zhang F, Towne CF, Batish SD, Lupski JR. 2010. *GJB1*/Connexin 32 whole gene deletions in patients with X-linked Charcot-Marie-Tooth disease. Neurogenetics 11(4):465–470.

Huynh N, Jeffrey BG, Turriff A, Sieving PA, Cukras CA. 2014. Sorting out co-occurrence of rare monogenic retinopathies: Stargardt disease co-existing with congenital stationary night blindness. Ophthalmic Genet 35(1):51–56.

Jaakson K, Zernant J, Kulm M, Hutchinson A, Tonisson N, Glavac D, Ravnik-Glavac M, Hawlina M, Meltzer MR, Caruso RC, Testa F, Maugeri A, et al. 2003. Genotyping microarray (gene chip) for the ABCR (ABCA4) gene. Hum Mutat 22(5): 395–403.

Martinez-Mir A, Paloma E, Allikmets R, Ayuso C, del Rio T, Dean M, Vilageliu L, Gonzalez-Duarte R, Balcells S. 1998. Retinitis pigmentosa caused by a homozygous mutation in the Stargardt disease gene ABCR. Nat Genet 18(1): 11–12.

Maugeri A, Klevering BJ, Rohrschneider K, Blankenagel A, Brunner HG, Deutman AF, Hoyng CB, Cremers FP. 2000. Mutations in the ABCA4 (ABCR) gene are the major cause of autosomal recessive cone-rod dystrophy. Am J Hum Genet 67(4):960–966.

Maugeri A, van Driel MA, van de Pol DJ, Klevering BJ, van Haren FJ, Tijmes N, Bergen AA, Rohrschneider K, Blankenagel A, Pinckers AJ, Dahl N, Brunner HG, et al. 1999. 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(4):1024–1035.

Ng PC, Henikoff S. 2001. Predicting deleterious amino acid substitutions. Genome Res 11(5):863–874.

Pertea M, Lin X, Salzberg SL. 2001. GeneSplicer: a new computational method for splice site prediction. Nucleic Acids Res 29(5):1185–1190.

Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. J Comput Biol 4(3):311–323.

Rivera A, White K, Stohr H, Steiner K, Hemmrich N, Grimm T, Jurklies B, Lorenz B, Scholl HP, Apfelstedt-Sylla E, Weber BH. 2000. A comprehensive survey of sequence variation in the ABCA4 (ABCR) gene in Stargardt disease and age-related macular degeneration. Am J Hum Genet 67(4):800–813.

Rosenberg T, Klie F, Garred P, Schwartz M. 2007. N965S is a common ABCA4 variant in Stargardt-related retinopathies in the Danish population. Mol Vis 13:1962–1969.

Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. 2010. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 7(8):575–576.

Shroyer NF, Lewis RA, Yatsenko AN, Lupski JR. 2001. Null missense ABCR (ABCA4) mutations in a family with stargardt disease and retinitis pigmentosa. Invest Ophthalmol Vis Sci 42(12):2757–2761.

Sun H, Smallwood PM, Nathans J. 2000. Biochemical defects in ABCR protein variants associated with human retinopathies. Nat Genet 26(2): 242–246.

Tavtigian SV, Deffenbaugh AM, Yin L, Judkins T, Scholl T, Samollow PB, de Silva D, Zharkikh A, Thomas A. 2006. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. J Med Genet 43(4):295–305.

Utz VM, Chappelow AV, Marino MJ, Beight CD, Sturgill-Short GM, Pauer GJ, Crowe S, Hagstrom SA, Traboulsi EI. 2013. Identification of three ABCA4 sequence variations exclusive to African American patients in a cohort of patients with Stargardt disease. Am J Ophthalmol 156(6):1220–1227 e2.

Valverde D, Riveiro-Alvarez R, Bernal S, Jaakson K, Baiget M, Navarro R, Ayuso C. 2006. Microarray-based mutation analysis of the ABCA4 gene in Spanish patients with Stargardt disease: evidence of a prevalent mutated allele. Mol Vis 12:902–908.

Yatsenko AN, Shroyer NF, Lewis RA, Lupski JR. 2003. An ABCA4 genomic deletion in patients with Stargardt disease. Hum Mutat 21(6):636–644.

Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol 11(2–3):377–394.

Zernant J, Schubert C, Im KM, Burke T, Brown CM, Fishman GA, Tsang SH, Gouras P, Dean M, Allikmets R. 2011. Analysis of the ABCA4 gene by next-generation sequencing. Invest Ophthalmol Vis Sci 52(11): 8479–8487.

Zhang MQ. 1998. Statistical features of human exons and their flanking regions. Hum Mol Genet 7(5):919–932.

Publication II

Zernant J, Xie YA, Ayuso C, Riveiro-Alvarez R, Lopez-Martinez MA, Simonelli F, Testa F, Gorin MB, Strom SP, Bertelsen M, Rosenberg T, Boone PM, Yuan B, Ayyagari R, Nagy PL, Tsang SH, Gouras P, Collison FT, Lupski JR, Fishman GA, Allikmets R (2014). Analysis of the *ABCA4* genomic locus in Stargardt disease. Hum Mol Genet. 23(25):6797-6806.

Human Molecular Genetics, 2014 **1–10**
doi:10.1093/hmg/ddu396

Analysis of the *ABCA4* genomic locus in Stargardt disease

Jana Zernant[1], Yajing (Angela) Xie[1], Carmen Ayuso[3,4], Rosa Riveiro-Alvarez[3,4], Miguel-Angel Lopez-Martinez[3,4], Francesca Simonelli[5], Francesco Testa[5], Michael B. Gorin[6,7], Samuel P. Strom[6,7,8], Mette Bertelsen[9], Thomas Rosenberg[9], Philip M. Boone[10], Bo Yuan[10], Radha Ayyagari[11], Peter L. Nagy[2], Stephen H. Tsang[1,2], Peter Gouras[1], Frederick T. Collison[12], James R. Lupski[10], Gerald A. Fishman[12] and Rando Allikmets[1,2,*]

[1]Department of Ophthalmology and [2]Department of Pathology and Cell Biology, Columbia University, New York, NY, USA, [3]Department of Genetics, Instituto de Investigacion Sanitaria-University Hospital Fundacion Jimenez Diaz, UAM (IIS-FJD), Madrid, Spain, [4]Centro de Investigacion Biomedica en Red (CIBER) de Enfermedades Raras, ISCIII, Madrid, Spain, [5]Eye Clinic, Multidisciplinary Department of Medical, Surgical and Dental Sciences, Second University of Naples, Naples, Italy, [6]Department of Ophthalmology, [7]Department of Human Genetics, Jules Stein Eye Institute and [8]Department of Pathology, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA, [9]Kennedy Center Eye Clinic, Glostrup Hospital, Glostrup, Denmark, [10]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA, [11]Department of Ophthalmology, University of California San Diego, La Jolla, CA, USA and [12]The Pangere Center for Hereditary Retinal Diseases, The Chicago Lighthouse for People Who are Blind or Visually Impaired, Chicago, IL, USA

Received May 29, 2014; Revised May 29, 2014; Accepted July 29, 2014

Autosomal recessive Stargardt disease (STGD1, MIM 248200) is caused by mutations in the *ABCA4* gene. Complete sequencing of *ABCA4* in STGD patients identifies compound heterozygous or homozygous disease-associated alleles in 65–70% of patients and only one mutation in 15–20% of patients. This study was designed to find the missing disease-causing *ABCA4* variation by a combination of next-generation sequencing (NGS), array-Comparative Genome Hybridization (aCGH) screening, familial segregation and *in silico* analyses. The entire 140 kb *ABCA4* genomic locus was sequenced in 114 STGD patients with one known *ABCA4* exonic mutation revealing, on average, 200 intronic variants per sample. Filtering of these data resulted in 141 candidates for new mutations. Two variants were detected in four samples, two in three samples, and 20 variants in two samples, the remaining 117 new variants were detected only once. Multimodal analysis suggested 12 new likely pathogenic intronic *ABCA4* variants, some of which were specific to (isolated) ethnic groups. No copy number variation (large deletions and insertions) was detected in any patient suggesting that it is a very rare event in the *ABCA4* locus. Many variants were excluded since they were not conserved in non-human primates, were frequent in African populations and, therefore, represented ancestral, and not disease-associated, variants. The sequence variability in the *ABCA4* locus is extensive and the non-coding sequences do not harbor frequent mutations in STGD patients of European-American descent. Defining disease-associated alleles in the *ABCA4* locus requires exceptionally well characterized large cohorts and extensive analyses by a combination of various approaches.

INTRODUCTION

Mutations in the *ABCA4* gene are responsible for a wide variety of retinal dystrophy phenotypes from autosomal recessive Stargardt disease (STGD1) (1) to cone–rod dystrophy (CRD) (2,3) and, in some advanced cases, retinitis pigmentosa (RP) (2,4,5). While CRD and RP phenotypes are also caused by mutations in many other genes, *ABCA4* is the only recognized gene

*To whom correspondence should be addressed. Email: rla22@columbia.edu

responsible for STGD1 (MIM 248200), a predominantly juvenile-onset macular dystrophy frequently associated with early-onset central visual impairment, progressive bilateral atrophy of the foveal retinal pigment epithelium, and the presence of yellowish flecks, defined as lipofuscin deposits, around the macula and/or in the central and near-peripheral areas of the retina.

Over 800 disease-associated *ABCA4* variants have been already identified (6) and the most frequent of these have been described in only ~10% of STGD1 patients (7). Several studies have identified frequent 'ethnic group-specific' *ABCA4* alleles, such as the p.G863A/G863del founder mutation in Northern European patients (8), the p.[L541P;A1038V] complex allele in patients of mostly German origin (3,9), the p.R1129L founder mutation in Spain (10), the p.N965S variant in the Danish population (11) and the p.A1773V variant in Mexico (12).

Complete sequencing of the *ABCA4* coding and adjacent intronic sequences in patients with STGD1 routinely discovers ~80% of mutations with the fraction of patients harboring the expected two disease-associated alleles at 65–70%, with one mutation ~15–20%, and with no mutations in the remaining ~15% (13). These fractions depend on many variables, most importantly the quality of the clinical diagnosis and the ethnic composition of the cohort. Most of the cases with no detected *ABCA4* mutations likely represent phenocopies (13); i.e. in those patients mutations in other gene(s) cause a STGD1-like phenotype. However, based on the known carrier frequency of pathogenic *ABCA4* variants, in most cases with one *ABCA4* allele the second allele is expected to reside in the *ABCA4* locus. It can present as a copy number variant (CNV, large deletion or insertion of one exon or more) which eludes detection by PCR-based sequencing techniques, a synonymous variant in the coding region, or a (deep) intronic variant, which may affect splicing or a regulatory region, such as a promoter or an enhancer (14). Very few of these have been identified (13,15,16).

This study was designed to find the missing *ABCA4* mutations by a combination of next-generation sequencing, array-comparative genomic hybridization (aCGH) arrays, *in silico* and RNA analyses, and segregation analyses in families.

RESULTS

Discovery of new disease-associated variants by next-generation sequencing

Sequencing of the entire *ABCA4* genomic locus, at an average depth of coverage of 100×, in 130 patients with *ABCA4*-associated disease harboring one previously known *ABCA4* disease-associated allele, and 6 patients with no known *ABCA4* mutations, resulted in detecting 1745 different variants. Eighty-three of these were previously known disease-associated or benign variants from coding regions and pathogenic splice site variants. Six hundred and ninety-five (695) variants were also detected in 1000 Genomes Project or Exome Sequencing Project, with no statistically significant differences in allele frequencies between the general population and the patient cohort, unless the variants were on the same allele (haplotype) with the frequent known *ABCA4* coding mutation, p.G1961E. Five hundred and twenty-six (526) variants were incorrectly called deletions or insertions from single nucleotide repeat areas (homopolymers) that have proven to be difficult for the NGS approach. We also experienced a relatively high A>C/C>A/T>G/G>T false-positive calling rate with Illumina sequencing. The number of false positives can be reduced by more stringent criteria for variant calling; however, this may also exclude some real variants. After the filtering and verification steps 141 new intronic *ABCA4* variants remained in 114 patients. In 22 patients with one previously known *ABCA4* mutation, the second pathogenic *ABCA4* allele was also found in the coding sequence or adjacent splice sites. In 6/22 cases this was due to re-evaluation of several variants which had been classified as benign, e.g. p.G991R and p.A1773V. The remaining 16 cases represented false-negative results, probably due to technical reasons in the first, sequencing, step of the ABCA4 coding regions.

Of the 141 new possible candidates for disease-associated variants, two variants, c.4539+2064C>T and c.5461-1389C>A, were detected together (one the same chromosome) as a complex allele in four patients of Spanish or Italian descent (Table 1). The c.5461-1389C>A variant is in an evolutionarily less conserved area, the c.4539+2064C>T variant is adjacent to the recently reported c.4539+2028C>T and c.4539+2001G>A variants from a conserved area (14). According to predictive programs, none of these variants have any effect on splicing, whether on existing cryptic splice sites or on creating new sites. The c.4539+2064C>T and c.5461-1389C>A haplotype segregated with the disease in all three STGD1 families from Spain (Fig. 1A–C), and were absent in 100 matched Spanish control samples, making these variants very likely candidates for intronic *ABCA4* mutations.

Two variants, c.5196+1056A>G and c.6006-609T>A, were detected in 3/114 unrelated patients each and were absent in 368 matched control samples. The c.5196+1056A>G variant segregated with the disease in two families; i.e. it was on a different chromosome than the proband's other *ABCA4* mutation (Fig. 1D). This variant is also predicted to strengthen a cryptic splice donor (Supplementary Material, Fig. S1A). The variant was found, in addition, in 1/119 patients from our replication cohort of STGD1 patients with one known *ABCA4* mutation and has been recently reported as a disease-associated allele (14). The aggregate evidence suggests that the c.5196+1056A>G variant is a rare, deep intronic disease-associated allele.

The c.6006-609T>A variant, which does not have a predicted effect on splicing, was detected in 6/119 additional unrelated samples from our replication cohort of STGD1 patients with one known *ABCA4* mutation. Two of these samples were from the Columbia University patient cohort and four from European cohorts including two from Denmark. The variant was not present in 368 European-American control samples, but was detected in 2/182 Danish control samples. While the frequency of this variant is 10× elevated in STGD1 patients as compared with all controls (1.9 versus 0.18%) and 3.5× if compared with Danish controls (0.55%), it is premature to unequivocally call the variant as associated with the disease.

Three variants, c.570+1798A>G, c.2161-8G>A, and c.859–9T>C were each detected in 2/114 different patients in our primary *ABCA4* locus screening cohort, but absent in 368 control samples and also in 119 additional STGD1 samples from the replication cohort (Table 1). The c.570+1798A>G

Table 1. Analysis of the new intronic *ABCA4* variants which were either detected twice or more in the cohort of 114 STGD1 patients, and/or have a predicted effect on splicing

Position on chr1	Variant	Effect on splicing (combined Alamut prediction)	C score	Primary *ABCA4* locus cohort (114)	Validation (replication) cohort (119)	Segregation with disease	Controls[a]	Conservation in primates	Disease association
94 580 645	c.67-2023T>G	Cryptic donor strongly activated	1.888	1	0		0/368	No	Probably not
94 566 773	c.570+1798A>G	New donor site	8.955	2	0		0/368	Yes	Yes
94 563 992	c.768+358C>T	No effect	9.324	2	0		0/368	Yes	Possibly yes
94 550 775	c.769-1778T>C	No effect	4.085	2	0		3/368	Yes	No
94 546 283	c.859-9T>C	Weakens the acceptor by ~14%	4.495	2	0	Yes	0/368, 0/120[b]	Yes	Yes
94 526 934	c.1938-619A>G	Cryptic donor strongly activated	2.211	1	0	Yes	0/368	Yes	Yes
94 525 509	c.2160+584A>G	New donor site	12.53	1	0		0/368	Yes	Yes
94 522 386	c.2161-8G>A	Weakens the acceptor by ~50%	13.25	2	0		0/368	Yes	Yes
94 514 521	c.2654-8T>G	Weakens the acceptor by ~35%	11.3	1	0		0/368	Yes	Yes
94 511 126	c.2919-826T>A	Cryptic donor strongly activated	1.513	1	0		0/368	No	Probably not
94 509 799	c.3050+370C>T	New donor site	1.072	1	0		0/368	Yes	Yes
94 495 923	c.4352+61G>A	Cryptic donor strongly activated	6.914	1	0		0/368	No	Probably not
94 493 272	c.4539+1729G>T	New donor site	1.586	1	0		0/368	Yes	Yes
94 493 000	c.4539+2001G>A[c]	No effect	4.653	2	0		0/368	Yes	Possibly yes
94 492 973	c.4539+2028C>T[c]	No effect	3.531	2	2		NS	Yes	Possibly yes
94 492 937	c.4539+2064C>T	No effect; the most frequent new variant	3.598	4	0	Yes	0/100[d]	Yes	Yes
94 484 082	c.5196+1056A>G[c]	Cryptic donor strongly activated	2.025	3	1	Yes	0/368	Yes	Yes
94 484 001	c.5196+1137G>A[c]	Cryptic acceptor strongly activated	2.681	1	9		0/368, 8/200[e]	No	No
94 478 330	c.5461-1389C>A[f]	No effect; the most frequent new variant	4.805	4	0	Yes	0/368, 0/100[d]	No	No
94 473 856	c.5836-3C>A	Weakens the acceptor by ~26%	11.84	1	0		NS	No	Yes
94 471 747	c.6006-609T>A	No effect, second most frequent variant	7.22	3	6		0/368, 2/180[g]	Yes	Possibly yes
94 468 019	c.6148-471C>T	Cryptic donor strongly activated	8.47	1	0		0/368	Yes	Yes
94 462 292	c.6730-541T>C	Cryptic acceptor strongly activated	5.682	1	0	No	0/368	Yes	No

[a]Control cohort of 368 samples of European ancestry was screened if not indicated differently.
[b]Cohort of 120 control samples of Asian-Indian origin.
[c]Variant is previously reported by Braun *et al*. (2013).
[d]Cohort of 100 controls from Spain.
[e]Cohort of 200 African-American general population controls.
[f]Variant is on the same allele with c.4539+2064C>T.
[g]Cohort of 180 controls from Denmark. NS, not screened.

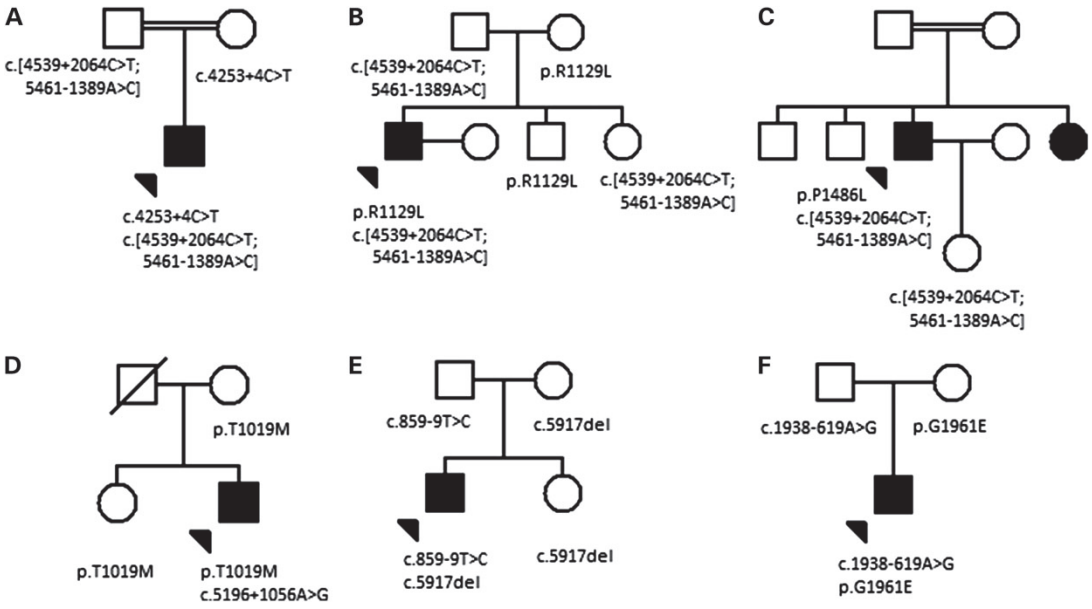


Figure 1. Pedigrees segregating the new *ABCA4* intronic variants with STGD1.

variant creates a very strong new donor splice site according to all predictive software (Supplementary Material, Fig. S1B). The c.2161-8G>A variant weakens the existing acceptor by 50%, and the c.859-9T>C change by 14% (Supplementary Material, Fig. S1C and D). The c.2161-8G>A and c.859-9T>C variants are adjacent to *ABCA4* exon sequences and can be detected also by sequencing of *ABCA4* coding regions. However, neither of these variants has been detected in the Exome Sequencing Project currently containing 4300 individuals of European-American descent and 2203 individuals of African-American descent. In addition to the 2/114 patients from the *ABCA4* locus screening cohort, c.859-9T>C has been detected twice, in homozygosity, in STGD1 patients with no other known *ABCA4* mutations, and in three more STGD1 patients in heterozygous state with one known *ABCA4* mutation. Segregation analysis was possible in one family, and confirmed that the c.859-9T>C was not on the same chromosome with the proband's other mutation (Fig. 1E). All evidence points to these three variants as being intronic disease-associated *ABCA4* variants; the c.859-9T>C variant is discussed in detail below.

Seventeen more new variants were detected twice in 114 patients from the locus screening cohort. Five of these variants were detected together in the same two patients both of whom also carried the most likely benign p.V931M variant, so these variants were eliminated from the pool of possible mutation candidates. Three more variants were on the same chromosome with the previously known *ABCA4* exon variants, (with p.R212C, p.T1253M, and p.[L541P;A1038V]), respectively. Other variants were either in the same patients who already carried a stronger intronic mutation candidate (5), not conserved in non-human primates (NHPs) (1), or were found in controls with similar frequency (1). None of them had a predicted effect on splicing. Among these variants are also the recently reported by Braun *et al*. c.4539+2028C>T and c.4539+2001G>A variants. Both variants were found in 2/114 samples in our primary cohort. The c.4539+2028C>T change was also detected in 2/119 patients of the replication cohort, a frequency significantly lower than reported by Braun *et al*. (14).

The remaining 117 new *ABCA4* intronic variants were only detected once each in 62 different patients with one previously known *ABCA4* mutation. Twelve of these variants were predicted to have an effect on splicing (Table 1). The c.5836-3C>A and c.2654-8T>G weaken the existing splice acceptor sites on average by 25 and 35%, respectively (Supplementary Material, Fig. S1E and F). Neither of these two variants has been detected by Exome Sequencing Project, nor in our entire STGD1 patient cohort (780 patients) where all *ABCA4* coding regions and adjacent splice sites have been sequenced.

The variants c.67-2023T>G, c.1938-619A>G, c.3050+370C>T, c.4352+61G>A and c.2919-826T>A, all strengthen the existing cryptic splice donors sites (Supplementary Material, Fig. S1G and H). None of these variants were found in 368 control samples, nor in the replication cohort of 119 additional STGD1 patients with one *ABCA4* disease-associated allele. The c.1938-619A>G variant segregated with the disease in one family (Fig. 1F). The c.67-2023T>G and c.2919-826T>A variants are not conserved in non-human primates and, therefore, not likely disease-causing in humans. At the same time, these two patients do not carry any other possible *ABCA4* mutant alleles. The variants c.2160+584A>G and c.4539+1729G>T are predicted to create new strong splice donors (Supplementary Material, Fig. S1I and J). The variants are

absent in 368 control samples and in the replication cohort of 119 STGD1 patients (Table 1). Since these positions are highly conserved among species we suggest that the c.2160+584A>G and c.4539+1729G>T variants are very good candidates for intronic *ABCA4* mutations.

The variants c.5196+1137G>A, c.6148-471C>T and c.6730-541T>C have a predicted effect of strengthening the existing cryptic splice acceptors (Supplementary Material, Fig. S1K–M), and were each all detected in one patient in the primary locus screening cohort. The c.6730-541T>C variant was on the same chromosome with the probands other known *ABCA4* mutation, and can therefore be excluded from possible new mutations list. The c.6148-471C>T and c.6730-541T>C variants were not detected in 368 control samples, nor in 119 additional STGD1 samples. The recently reported (14) c.5196+1137G>A variant was additionally found in nine STGD1 samples with one previously known *ABCA4* variant (Table 1). Three of these patients were from Denmark, three were of African-American origin, and three of unknown ethnicity. Screening matched control cohorts revealed no carriers for this allele in 180 Danish control samples, but it was detected in 8/200 (MAF = 2%) general population controls of African-American descent, strongly suggesting non-pathogenicity. Additional support for this variant not being pathogenic comes from evolutionary analysis (see below). The human major allele c.5196+1137G is the minor allele in *Macaca mulatta* and the suggested mutant allele A is the major allele in macaques.

Analysis of regulatory sequences

To assess the effect of the new *ABCA4* intronic variants on putative regulatory regions we compared their location against the chromosome coordinates of the DNaseI hypersensitivity and transcription factor binding regions from the ENCODE project. Since regulatory regions, in particular promoters, tend to be DNase sensitive, variants that fall within such regions may affect regulatory potential. Also, variants that are located in transcription factor binding sites may potentially have an effect on protein binding. The combined DNaseI hypersensitivity data were derived from 125 cell types, and the dataset for transcription factor binding regions involved 161 transcription factors in 91 cell types. Unfortunately the datasets did not contain eye-specific cell types or transcription factors. The defined regions were assigned normalized scores in the range of 0–1000, with higher scores indicating stronger signal strength. Twenty-four of the 141 new *ABCA4* intronic variants are located in regions with both DNaseI hypersensitivity and transcription factor binding scores of various strength (Supplementary Material, Table S1). Family members were available for segregation analyses in three cases. Variants c.570+1499C>A and c.67-3166C>T were on the same chromosome with the probands' other mutation, while c.571-1707C>T and c.5018+289T>C were both detected in one patient and on the different chromosome than this patient's other mutation, suggesting possible pathogenicity. Twenty-one variants fall within regions with only DNaseI hypersensitivity score, and 12 variants within regions of transcription factor binding consensus sequences.

The new *ABCA4* intronic variants were also subjected to the Combined Annotation Dependent Depletion (CADD) algorithm (http://cadd.gs.washington.edu/score) (17). The CADD algorithm combines a diverse array of annotations into one metric (C score) for each variant, ranking a variant relative to all possible substitutions of the human genome (17). C score correlates with allelic diversity, pathogenicity and experimentally measured regulatory effects. A C score of >10 indicates that the variant is among the top 10% of most deleterious substitutions in the human genome. C scores for the new *ABCA4* intronic variants range from 0.004 to 15.14 (Supplementary Material, Table S1), with nine variants resulting in a C score greater than 10. Four of those variants were already classified as possibly disease-associated due to their strong predicted effect on splicing (Table 1). The other five variants with higher C scores included c.4539+1971T>C, c.67-3779T>G, c.303-906A>G, c.859-41T>C and c.1357-221T>A (Supplementary Material, Table S1).

In summary, we found a very strong or a probable intronic mutation candidate in 27/114 (23.7%) patients with one existing definite *ABCA4* mutation (Table 1). No immediately plausible candidates for intronic mutations were found in 36/114 (31.6%) patients. The remaining 51 (44.7%) patients possessed one or more new intronic variants that were only detected once, had no effect on splicing according to prediction programs and, therefore, are very difficult to confirm or refute as disease-associated alleles with the available methods and the impossibility of obtaining the patient RNA. However, it is highly likely that a fraction of these are associated with STGD1.

Analysis of previously reported variants

Recently several *ABCA4* intronic variants were suggested to account for a substantial fraction of pathogenic *ABCA4* alleles (14). Since we have directly sequenced the *ABCA4* gene and flanking intronic sequences in >780 STGD1 patients and the entire *ABCA4* genomic locus in 114 STGD patients with one

Table 2. Frequency of the variants described in Braun *et al*. (2013), and in the current study.

Variant # (Braun *et al*.)	Location on chr1	Variant	Braun *et al*. primary cohort (n=28)	CU locus cohort (n = 114)	CU exon seq cohort (n = 780)
V1	94 484 001	c.5196+1137G>A	4	1	–
V2	94 483 922	c.5196+1216C>A	1	0	–
V3	94 484 082	c.5196+1056A>G	1	3	–
V4	94 493 000	c.4539+2001G>A	1	2	–
V5	94 492 973	c.4539+2028C>T	3	2	–
V6	94 466 602	c.6342G>A/p.V2114V	1	0	1
V7	94 487 399	c.4773+3A>G	3	0	0

mutation, we compared the Braun *et al*. data to ours (Table 2). The silent exonic p.V2114V variant, which we had described in our earlier study as a possibly disease-associated mutation (13), is very rare; it was detected in 1 out of 780 STGD1 patients (Table 2). The 'near-exonic' c.4773+3A>G variant was not detected in any of our 780 patients. We were also unable to detect one of the remaining five 'deep intronic' variants (c.5196+1216C>A) in any patient with one mutation and the other four were seen, in total, in 9/114 (7.9%) STGD1 patients, a statistically significant difference from the Braun *et al*. data 10/28 (17.9%; $P = 0.001$).

We then analyzed the evolutionary and/or ethnic origin of the variants focusing first on the most frequent variant in the Braun *et al*. study, c.5196+1137G>A. Evolutionary conservation of a nucleotide is one of the most important criteria for determining the pathogenicity of a variant, especially in a highly conserved gene such as *ABCA4*. The ABCA4 protein performs a very specialized function in the visual cycle; therefore it is exceptionally conserved in mammals and in all vertebrates with visual cycle. For example, the mouse and human ABCA4 proteins are 88% identical, allowing the human protein to perform the transport function in mouse (18). The conservation extends beyond coding sequences and includes splice sites; in fact, the *ABCA4* gene has the same structure, consisting of exactly the same 50 exons in non-human primates, such as *Pan troglodytes* and *M. mulatta*. The evolutionary conservation extends, in some regions, deep into the introns; for example, the 200 bp sequences surrounding the c.5196+1137G>A variant are 96% identical between human and macaque. Most importantly, *M. mulatta* has the adenosine nucleotide in human position c.5196+1137 (Fig. 2), as a major allele with guanosine as a minor allele; i.e. a situation exactly the reverse to that observed in humans, suggesting that this is an ancestral variant and not a disease-causing mutation. To further prove this assumption, we screened 200 unrelated individuals from the general population of African-American descent and identified eight heterozygotes, resulting in the allele frequency of ~2%. The cohort of STGD1 patients at Columbia University contains 46 African-American patients of whom one carried the c.5196+1137G>A variant, resulting in the allele frequency of ~1%, comparable to that in the general population. No other disease-associated *ABCA4* variant was identified in this patient. In addition, we detected the c.5196+1137G>A variant only once in 114 STGD1 patients (allele frequency 0.4%) of European descent in the Columbia cohort (Tables 1 and 2). However, it was found in 3 out of 24 STGD1 patients (allele frequency 6.25%) in a STGD1 cohort from Denmark (Table 1). Interestingly, all three patients derived from the same region in Denmark, suggesting either admixture, or that this variant may be more frequent also in some (isolated) ethnic groups other than of African descent since the variant was not detected in 180 unaffected Danish individuals from Copenhagen. Finally, RNA analysis from macaque homozygous for the A allele at the human position c.5196+1137 clearly showed no effect on splicing (Fig. 2), which eliminates the c.5196+1137G>A variant as possibly pathogenic in STGD1 patients.

The three remaining deep intronic variants described by Braun *et al*. (14), c.5196+1056A>G, c.4539+2028C>T and c.4539+2001G>A are discussed in the previous section. All of these are very rare and are conserved in NHPs. The c.5196+1056A>G variant, which affects splicing, is likely a

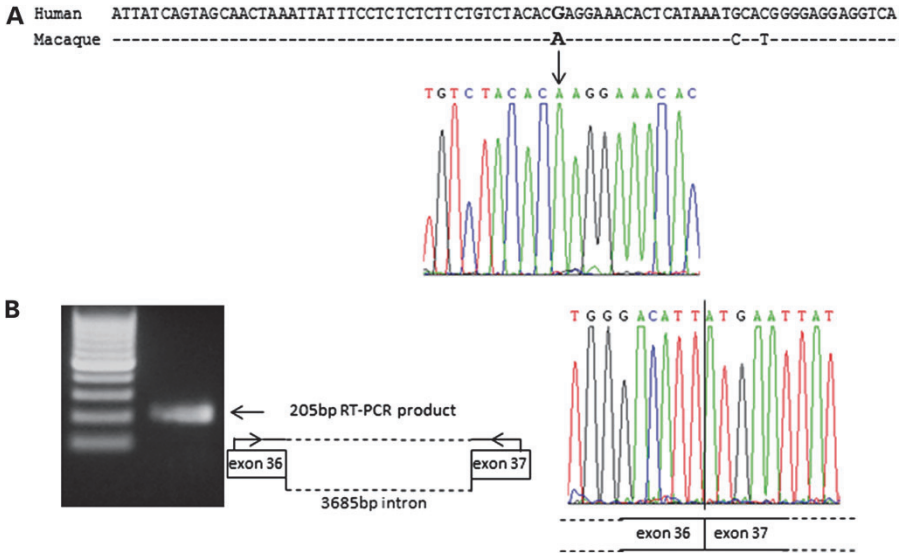


Figure 2. Analysis of the c.5196+1137G>A variant in *M. mulatta*. (**A**) Alignment of the *Homo sapiens* chromosome1: 94 484 046–94 483 967, GRCh37.p13 Primary Assembly, with *M. mulatta* respective sequences. *ABCA4* intronic position c.5196 + 1137G is marked with large font in bold. Differences in macaque sequence compared with human are designated with letters. (**B**) Confirmation of correct splicing of *ABCA4* exons 36 and 37 in a c.5196+1137A macaque retina in RT-PCR analysis. No alternate splicing products were detected.

disease-associated mutation. The other two are not predicted to affect splicing and are too rare to investigate by other means, so the pathogenicity of these variants remains unconfirmed.

New frequent intronic variant in STGD1 patients of Asian Indian descent

As described above, we detected the c.859-9T>C variant in 2/114 patients in our primary *ABCA4* locus screening cohort. Since the c.859-9T>C variant weakens the existing splice acceptor only by 14%, we initially did not consider the variant a strong candidate for disease association. However, subsequently the variant was detected in homozygous state in two STGD1 patients who did not harbor any other exonic or intronic *ABCA4* mutations and, heterozygously, in three more STGD1 patients with one known *ABCA4* mutation. Review of the ethnic origin of all these patients determined that they were all of Asian Indian descent originating from either Pakistan, India or Bangladesh. Segregation analysis was possible in one family, and confirmed that the c.859-9T>C segregated with the disease (Fig. 1E); i.e. it was in *trans* configuration with the second *ABCA4* mutation, c.5917del, in the proband. The c.859-9T>C variant is adjacent to *ABCA4* exon sequences and is, therefore, detected also by exome sequencing, or by direct sequencing of *ABCA4* coding regions. However, it has not been detected in the Exome Sequencing Project currently containing 4300 individuals of European-American descent and 2200 individuals of African-American descent. It was also absent in 368 control samples and also in 119 additional STGD1 samples of European-American descent with one known *ABCA4* mutation (Table 1). Further screening in Asian-Indian population did not detect this variant in 120 subjects, both from the general population (50 individuals), or from the patients affected with Leber congenital amaurosis or Leber hereditary optic neuropathy (70 cases). Altogether, the data suggest that the variant is not frequent in the ethnically matched general population and is very frequent in Indian patients with STGD1 (7 alleles in 15 patients—23.3%) suggesting that the c.859-9T>C variant is a frequent disease-associated *ABCA4* allele in patients of Asian-Indian origin.

Analysis of the copy number variants by aCGH arrays

Few lines of evidence have suggested that some disease-associated *ABCA4* alleles can present as CNVs, mostly in the form of large deletions encompassing one or more exons (15). These reports have been rare, so we reasoned that a small fraction of missing *ABCA4* alleles might present as CNVs. In total of 104 STGD1 patients with one known *ABCA4* allele and 5 patients with no *ABCA4* mutations were screened on the custom CGH arrays, 57 of these were also included in the locus sequencing. aCGH data of the total number of 109 samples was analyzed by using Agilent Genomic Workbench (version 7.0), where CNVs were called by using ADM-2 algorithm with threshold of 4.0. No large (>500 bp) CNVs were identified in the *ABCA4* locus, while ultra-small, seemingly true-positive CNV calls were further validated by PCR, which did not confirm any, reflecting the decreased accuracy of aCGH for the ultra-small CNVs Since the array contained several genes in the *CFH* locus with known frequent CNVs and confirmed in a positive control a heterozygous 1030 bp deletion in the *ABCA4* locus, we can exclude any technical issues and conclude that, despite reports of CNVs (mainly large deletions) in 1–2% of STGD1 patients, these events are likely much more rare and the CNVs do not account for a reasonable fraction of missing *ABCA4* alleles in all populations studied.

DISCUSSION

The *ABCA4* gene was described as the causal gene for STGD1 17 years ago (1) and has been subjected to intensive genetic research; however, the complete understanding of genetic causality has yet to be elucidated. Major factors challenging genetic analysis are: (1) the size of the gene, (2) the exceptional genetic heterogeneity, (3) the expression of the gene in only photoreceptors, rods and cones and, consequently and most importantly, (4) the impossibility of obtaining retinal tissue samples *in vivo* from patients for RNA analyses. Until recently, even the analysis of amino acid-changing variants was stymied by the lack of a direct functional assay; these analyses were limited to indirect *in vitro* studies of the transporters' ability to bind and/or hydrolyze ATP (19,20). This problem has been somewhat alleviated by the recent studies of Molday *et al.* (21,22), where direct transport assays, albeit still *in vitro*, have been proposed and successfully used for several *ABCA4* missense alleles. Large-scale use of the assay for hundreds of documented *ABCA4* variants is still time- and cost-prohibitive. However, the functional analyses of variants that do not affect the protein sequence directly, such as splicing defects and variants in regulatory regions likely modulating levels of expression, remain refractory to experimental verification.

Some studies (9,14) have employed the 'illegitimate transcript' or 'minigene' strategies to assess the effect of certain variants on splicing, but these studies have serious limitations (in part stemming from the word 'illegitimate') and the results, while suggestive, cannot be considered unequivocal, as also demonstrated in the current study. With the lack of patient RNA for direct structural and expression studies, the analysis is limited to the assessment of variant frequencies in STGD1 patients and in the matched general populations, to *in silico* suggestions by predictive software programs, and to segregation analyses in families. The latter approach is also seriously hampered by the fact that most variants in *ABCA4* coding and non-coding regions are extremely rare, most often represented in singleton cases and the available families/pedigrees are usually nuclear, i.e. mostly very small, where the segregation analysis has a very limited power.

In the current study of non-coding genetic variation on the largest STGD1 cohort analyzed to date, we were able to (almost) unequivocally assign pathogenicity to only 12/141 variants in intronic sequences of *ABCA4*. Most of the variants which occurred only once in the *ABCA4* locus cannot be called especially when the predictive programs do not suggest any effect on splicing. Moreover, the assessment based on predictive programs is not unequivocal, several studies have suggested that molecularly confirmed predictions range from 70 to 80% (23). One has to take into account the fact that assessment of the effect on splicing addresses just one possible mechanism for non-coding sequences. Effect on these on other regulatory

elements affecting the gene expression, such as transcription factor binding sites, enhancers, promoters, etc., is still very difficult to assess in most cases (24). The best known example of an intronic variant in the *ABCA4* locus is the c.5461-10T>C variant in intron 38, which is the second or third most frequent variant (found in ~7% of STGD1 patients of European descent) after the p.G1961E and p.[L541P;A1038V] mutations (13). The c.5461-10T>C variant always segregates with the disease phenotype in families, is very rare in the general population (<0.001) and is shown not to affect splicing in a minigene approach (9). Since there is no mutation in the *ABCA4* coding sequence on the same chromosome with the c.5461-10T>C variant, the latter has to be a disease-associated mutation, although its functional consequences remain unknown.

A recent study suggested a few deep intronic variants in the *ABCA4* locus which were associated with STGD1 in a large fraction, up to 50%, of patients with one exonic mutation. Our detailed analysis of these data with a multi-faceted approach on a much larger cohort of European-American descent could not confirm these conclusions. Specifically, the reported variants were much, 10–50 times, rarer than originally claimed (Table 2) and some of them were deemed not associated with STGD1 after evolutionary and RNA analyses employing NHPs (Fig. 2). Concerns also include the ethnic origin and relatedness of patients in cohorts described by Braun *et al*. (14), since the statistically significant difference in allele frequencies between two studies can occur, excluding technical issues, for two main reasons: (1) cohorts include related subjects or, (2) difference in the racial or ethnic origin of the cohorts. For example, recently several missense *ABCA4* variants that were originally considered pathogenic since they were very rare in populations of European descent have now been classified non-pathogenic since they are major alleles (i.e. the 'wild-type' variants) in non-human primates (e.g. p.R1300Q) and/or frequent in some racial and ethnic groups, for example, in African-Americans (e.g. p.L1201R, p.R1300Q and p.V643M) (25).

In summary, the analysis of the entire *ABCA4* locus in large cohorts of STGD1 patients revealed the following:

(1) The genetic variability in the non-coding sequences in the *ABCA4* locus, similar to the coding sequences, is exceptionally vast.
(2) There are no frequent pathogenic variants in the non-coding sequences of the *ABCA4* locus; all definitely or likely disease-associated variants are individually rare in the populations of European descent.
(3) Some variants are more frequent in specific racial and ethnic groups.
(4) Copy number variations in the *ABCA4* locus are very rare.

Analysis of the pathogenicity of specifically intronic *ABCA4* variants, which affect splicing and regulatory regions influencing the gene expression, is complicated due to the impossible task of obtaining RNA from photoreceptor cells from affected individuals *in vivo*. Studying iPS cells obtained from individual patients, which are then directed towards differentiating into photoreceptors with the goal of expressing *ABCA4*, could be a plausible approach.

MATERIALS AND METHODS

Patients

STGD1 patients (255) were, after written informed consent, recruited and clinically examined during a 10-year period in different centers in the USA, Italy, Spain and Denmark. Control cohorts included samples from centers in the USA, Spain and Denmark. In total, 255 patients, and 918 control samples were included in the analyses.

Our primary locus sequencing cohort of 136 samples consisted of 49 STGD1 patients from Columbia University, 22 from the University of Illinois at Chicago and the Pangere Center at the Chicago Lighthouse, 22 from UCLA, 25 patients from Italy and 18 from Spain. The second, validation/replication cohort consisted of 119 STGD1 patients from the same centers. Age of onset was defined as the age at which symptoms were first reported. Visual acuity was measured using the Early Treatment Diabetic Retinopathy Study Chart 1 or a Snellen acuity chart. Clinical examination, fundus photography, fundus autofluorescence and spectral domain-optical coherence tomography (SD-OCT) (Heidelberg Spectralis HRA + OCT) were performed using standard acquisition protocols following pupil dilation with Tropicamide 1% and Neosynephrine 2%. All research was carried out with the approval of the Institutional Review Boards at the respective centers and in accordance with the Declaration of Helsinki.

M. mulatta samples

DNA samples from 86 unrelated rhesus macaques (*M. mulatta*) from NHP colonies at the National Institutes of Health and the Primate Facility at the University of Oregon were genotyped for the c.5196+1137G>A variant.

Sequencing

The first set of 48 patients were sequenced using RainDance microdroplet-PCR target enrichment (RainDance Technologies, Billerica, MA) with subsequent sequencing on Roche 454 platform (454 Life Sciences, Branford, CT). We targeted the genomic region chr1:94 453 727–94 595 732 (GRCh37/hg19 Assembly); including the *ABCA4* genomic locus chr1:94 458 394–94 586 705 and 9027 bp of 5′UTR and 4667 of 3′UTR sequences. The design covered 100% of the targeted area via 473 amplicons of 350–500 bp.

The second set of 95 patient samples were analyzed by the Illumina Truseq Custom Amplicon target enrichment strategy followed by sequencing on Illumina MiSeq platform (Illumina, San Diego, CA). The Illumina design targeted the genomic region 94 456 700–94 591 600, including the *ABCA4* locus, and 4895 bp of 5′UTR and 1694 bp of 3′UTR sequences. The region was divided into nine targets, with seven 500 bp and one 1100 bp gap introduced into the repeating elements. The cumulative target via Illumina design involved 130 319 bp, covered by 94% with 421 amplicons of 425 bp. The next-generation sequencing reads were analyzed using the variant discovery software NextGENe (SoftGenetics, State College, PA). The reads were aligned against the targeted region in the reference genome GRCh37/hg19. For a variant to be called, we required the read containing the variant to match 85% to the

aligned position, the variant to be covered by at least 10 reads and the variant to be present in 20% of all reads aligned to the given position. We also used the overall confidence score of 10 of the NextGENe software as a further filter. We used the previously determined *ABCA4* coding variants as controls to set these filtering criteria. On average, ~200 variants were called per individual patient.

Analysis of the *ABCA4* variants

All variants and their allele frequencies were compared with the 1000 Genomes database (26), and to the Exome Variant Server (EVS) dataset, NHLBI Exome Sequencing Project, Seattle, WA, USA (http://snp.gs.washington.edu/EVS/; accessed November 2013). New variants that were not recorded in these databases were further analyzed by a combination of predictive *in silico* methods and statistical analyses. The possible effect of all new non-coding *ABCA4* variants on splicing was assessed using five different algorithms (SpliceSiteFinder, MaxEntScan, NNSPLICE, GeneSplicer, Human Splicing Finder) via Alamut software (http://www.interactive-biosoftware.com). In order to assess the regulatory potential of the new *ABCA4* intronic variants we compared their chromosome coordinates against the predicted regulatory regions from two ENCODE datasets: (1) Combined DNaseI hypersensitivity clusters from 125 cell types ('Digital DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE' http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=366969521&c=chr1&g=wgEncodeRegDnaseClusteredV2, filename: wgEncodeRegDnaseClusteredV2.bed.gz, last accessed on 23 August 2012); and (2) ChIP-seq clustered regions for 161 transcription factors in 91 cell types ('Transcription Factor ChIP-seq V4 (161 factors) with Factorbook motifs from ENCODE'http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=366969521&c=chr1&g=wgEncodeRegTfbsClusteredV3, filename: wgEncodeRegTfbsClusteredV3.bed.gz, last accessed on 21 July 2013). Evolutionary conservation of the variants was noted via UCSC Genome Browser (http://genome.ucsc.edu). The Combined Annotation Dependent Depletion (CADD) algorithm (http://cadd.gs.washington.edu/score) was used to estimate combined predicted general deleteriousness of every variant. The variants' segregation with the disease in available families was analyzed by Sanger sequencing, and screening of patient and control cohorts for allele frequencies in various populations was performed using TaqMan Genotyping technology (Life Technologies, Carlsbad, CA). For genotyping for the c.5196+1137G>A variant we used PCR RFLP with forward primer 5′GTGGGCCTAGCTCCTTTTAT3′, reverse primer 5′GGAGACCAACACAAATGACC3′ (Life technologies, Carlsbad, CA), and the DNA restriction endonuclease BssSI (New England Biolabs, Ipswich, MA)

Array-comparative genome hybridization (aCGH)

Custom aCGH arrays (Agilent Technologies, CA), in a 8 × 60K format, was designed with high-density probes tiling the critical genetic loci of *ABCA4*-associated disease were designed to assess for CNVs involving these loci. The *ABCA4* locus and eight other known genes causing macular disease *(ELOVL4, PRPH2, BEST1, RS1, CNGB3*), and major age-related macular degeneration-associated loci *(CFH, ARMS2, C2/CFB*), were considered as 'primary loci' and therefore probed with ultra-high density throughout the entire genomic length of the genes plus the 5′ promoter regions and 3′ downstream regions, as well as at slightly lower density for flanking 5′ and 3′ conserved regions. Eighteen genetic loci *(C3, APOE, CFI, LIPC, SYN3/TIMP3, CETP, COL8A1, BBX, PLD1, SPEF2, ADAM19, VEGFA, FRK, MEPCE, CHMP7/LOXL2, TGFBR1, NPS, PICK1)* were considered as 'secondary loci' and probed with ultra-high density in exons and with lower resolution throughout each gene plus the flanking 5′ promoter regions and 3′ downstream regions. The details of the array design are described in the Supplementary Material, Table S2.

Array-CGH analysis was performed on DNA from 104 individuals diagnosed with STGD1, for each of whom only one mutation in *ABCA4* had been found by sequencing. A DNA sample with a known, previously reported 1030 bp heterozygous deletion of exon 18 in *ABCA4*, was used as positive control (15). Experimental procedures of aCGH were performed as described previously (27) with minor modifications. Agilent Genomic Workbench version 7.0 software (Agilent Technologies, CA) was used for data analysis, and PCR validations were performed for the plausible true-positive CNVs after being filtered against several criteria.

ABCA4 RNA analysis from rhesus macaque retina

Total RNA was isolated from a snap frozen rhesus macaque (*M. mulatta*) retina using AllPrep DNA/RNA Mini Kit (QIAGEN Cat. No 80204) with a fast spin-column procedure. cDNA synthesis was achieved using TaqMan Reverse Transcription Reagents (Life technologies, Carlsbad, CA) in a 30 min incubation at 48°C. The primer pair for further PCR amplification was designed to encompass *ABCA4* exons 36 and 37, with the forward primer in the exon 36 of the *ABCA4* gene, 5′GATTTTCTCCATGTCCTTCG3′ and the reverse primer in the exon 37, 5′CTTTCTTCTGAAACCCGATG3′, resulting in an amplicon of 205 bp, in the case of correct splicing.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

The authors thank Dr John Fingert for the analysis of a subset of Asian-Indian samples and Dr Dwight Stambolian for providing African-American general population samples.

Conflict of Interest statement. None declared.

FUNDING

This work was supported, in part, by grants from the National Eye Institute/NIH EY021163, EY019861, EY021237 and EY019007 (Core Support for Vision Research); Foundation Fighting Blindness (Owings Mills, MD), Harold and Pauline Price Foundation, unrestricted funds from Research to Prevent Blindness (New York, NY) to the Departments of Ophthalmology, Columbia University and UCLA, the Pangere Family

Foundation, Chicago Lighthouse, FIS PI13/00226 & CIBERER from ISCIII, Madrid, Spain, ONCE and Fundaluce (Spain).

REFERENCES

1. Allikmets, R., Singh, N., Sun, H., Shroyer, N.F., Hutchinson, A., Chidambaram, A., Gerrard, B., Baird, L., Stauffer, D., Peiffer, A. *et al.* (1997) A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. *Nat. Genet.*, **15**, 236–246.
2. Cremers, F.P., van de Pol, D.J., van Driel, M., den Hollander, A.I., van Haren, F.J., Knoers, N.V., Tijmes, N., Bergen, A.A., Rohrschneider, K., Blankenagel, A. *et al.* (1998) 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.
3. Maugeri, A., Klevering, B.J., Rohrschneider, K., Blankenagel, A., Brunner, H.G., Deutman, A.F., Hoyng, C.B. and Cremers, F.P. (2000) Mutations in the ABCA4 (ABCR) gene are the major cause of autosomal recessive cone–rod dystrophy. *Am. J. Hum. Genet.*, **67**, 960–966.
4. Martinez-Mir, A., Paloma, E., Allikmets, R., Ayuso, C., del Rio, T., Dean, M., Vilageliu, L., Gonzalez-Duarte, R. and Balcells, S. (1998) Retinitis pigmentosa caused by a homozygous mutation in the Stargardt disease gene ABCR. *Nat. Genet.*, **18**, 11–12.
5. Shroyer, N.F., Lewis, R.A., Yatsenko, A.N. and Lupski, J.R. (2001) Null missense ABCR (ABCA4) mutations in a family with Stargardt disease and retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.*, **42**, 2757–2761.
6. Allikmets, R. (2007) Tombran-Tink, J. and Barnstable, C.J. (eds), In *Retinal Degenerations: Biology, Diagnostics and Therapeutics*, Humana Press, Totowa, NJ, in press, pp. 105–118.
7. Burke, T.R., Fishman, G.A., Zernant, J., Schubert, C., Tsang, S.H., Smith, R.T., Ayyagari, R., Koenekoop, R.K., Umfress, A., Ciccarelli, M.L. *et al.* (2012) Retinal phenotypes in patients homozygous for the G1961E mutation in the ABCA4 gene. *Invest. Ophthalmol. Vis. Sci.*, **53**, 4458–4467.
8. Maugeri, A., van Driel, M.A., van de Pol, D.J., Klevering, B.J., van Haren, F.J., Tijmes, N., Bergen, A.A., Rohrschneider, K., Blankenagel, A., Pinckers, A.J. *et al.* (1999) 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.
9. Rivera, A., White, K., Stohr, H., Steiner, K., Hemmrich, N., Grimm, T., Jurklies, B., Lorenz, B., Scholl, H.P., Apfelstedt-Sylla, E. *et al.* (2000) 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.
10. Valverde, D., Riveiro-Alvarez, R., Bernal, S., Jaakson, K., Baiget, M., Navarro, R. and Ayuso, C. (2006) Microarray-based mutation analysis of the ABCA4 gene in Spanish patients with Stargardt disease: evidence of a prevalent mutated allele. *Mol. Vis.*, **12**, 902–908.
11. Rosenberg, T., Klie, F., Garred, P. and Schwartz, M. (2007) N965S is a common ABCA4 variant in Stargardt-related retinopathies in the Danish population. *Mol. Vis.*, **13**, 1962–1969.
12. Chacon-Camacho, O.F., Granillo-Alvarez, M., Ayala-Ramirez, R. and Zenteno, J.C. (2013) 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.
13. Zernant, J., Schubert, C., Im, K.M., Burke, T., Brown, C.M., Fishman, G.A., Tsang, S.H., Gouras, P., Dean, M. and Allikmets, R. (2011) Analysis of the ABCA4 gene by next-generation sequencing. *Invest Ophthalmol. Vis. Sci.*, **52**, 8479–8487.
14. Braun, T.A., Mullins, R.F., Wagner, A.H., Andorf, J.L., Johnston, R.M., Bakall, B.B., Deluca, A.P., Fishman, G.A., Lam, B.L., Weleber, R.G. *et al.* (2013) Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. *Hum. Mol. Genet.*, **22**, 5136–5145.
15. Yatsenko, A.N., Shroyer, N.F., Lewis, R.A. and Lupski, J.R. (2003) An ABCA4 genomic deletion in patients with Stargardt disease. *Hum. Mutat.*, **21**, 636–644.
16. Stenirri, S., Battistella, S., Fermo, I., Manitto, M.P., Martina, E., Brancato, R., Ferrari, M. and Cremonesi, L. (2006) De novo deletion removes a conserved motif in the C-terminus of ABCA4 and results in cone–rod dystrophy. *Clin. Chem. Lab. Med.*, **44**, 533–537.
17. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M. and Shendure, J. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.*, **46**, 310–315.
18. Kong, J., Kim, S.R., Binley, K., Pata, I., Doi, K., Mannik, J., Zernant-Rajang, J., Kan, O., Iqball, S., Naylor, S. *et al.* (2008) Correction of the disease phenotype in the mouse model of Stargardt disease by lentiviral gene therapy. *Gene Ther.*, **15**, 1311–1320.
19. Sun, H., Smallwood, P.M. and Nathans, J. (2000) Biochemical defects in ABCR protein variants associated with human retinopathies. *Nat. Genet.*, **26**, 242–246.
20. Shroyer, N.F., Lewis, R.A., Yatsenko, A.N., Wensel, T.G. and Lupski, J.R. (2001) 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.
21. Quazi, F. and Molday, R.S. (2013) 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.
22. Quazi, F. and Molday, R.S. (2014) ATP-binding cassette transporter ABCA4 and chemical isomerization protect photoreceptor cells from the toxic accumulation of excess 11-cis-retinal. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 5024–5029.
23. Liu, Y.H., Li, C.G. and Zhou, S.F. (2009) Prediction of deleterious functional effects of non-synonymous single nucleotide polymorphisms in human nuclear receptor genes using a bioinformatics approach. *Drug Metab. Lett.*, **3**, 242–286.
24. Hardison, R.C. and Taylor, J. (2012) Genomic approaches towards finding cis-regulatory modules in animals. *Nat. Rev. Genet.*, **13**, 469–483.
25. Utz, V.M., Chappelow, A.V., Marino, M.J., Beight, C.D., Sturgill-Short, G.M., Pauer, G.J., Crowe, S., Hagstrom, S.A. and Traboulsi, E.I. (2013) Identification of three ABCA4 sequence variations exclusive to African American patients in a cohort of patients with Stargardt disease. *Am J Ophthalmol*, **156**, 1220–1227. e1222.
26. Genomes Project, C., Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T. and McVean, G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.
27. Gonzaga-Jauregui, C., Zhang, F., Towne, C.F., Batish, S.D. and Lupski, J.R. (2010) *GJB1*/connexin 32 whole gene deletions in patients with X-linked Charcot-Marie-Tooth disease. *Neurogenetics*, **11**, 465–470.

Supplemental Table 1. The 141 non coding new *ABCA4* locus variants that have not been reported in 1000 Genomes Database, with C-scores and DNaseI hypersensitivity and Transcription Factor (TF) scores.

Position on chr1	Ref Nucleotide	Variant Nucleotide	Variant	Alleles in 114 samples	C score	DNaseI score	TF binding score	TF	Comments
94593916	GAGT	T	c.-7318_-7316del	1	6.174				no predicted effect
94589709	CAAAAA	A	c.-3113_-3109del	1	0.124				no predicted effect
94589707	CAAAAAAA	A	c.-3113_-3107del	1	0.155				no predicted effect
94589000	TC	C	c.-2400del	1	7.266				no predicted effect
94588938	A	G	c.-2337A>G	1	5.946				no predicted effect
94588085	CA	A	c.-1485del	1	3.092		353	FOS	no predicted effect
94588081	TCAA	A	c.-1483_-1481del	1	4.690		353	FOS	no predicted effect
94587006	A	T	c.-405A>T	2	5.550	818	444	TCF7L2	no predicted effect, on the same allele with c.570+1798A>G
94586760	T	C	c.-159T>C	2	9.106	1000			no predicted effect, on the same allele with c.5196+1056A>G
94584729	AC	A	c.66+1806del	1	4. 175				no predicted effect
94583189	C	T	c.66+3347C>T	1	4.288				no predicted effect
94583170	A	G	c.66+3366A>G	1	2.870				no predicted effect
94582524	AAGAG	G	c.67-3906_-3903del	1	0.714				no predicted effect
94582524	AAGAAAGAG	G	c.67-3910_-3903del	1	0.622				no predicted effect
94582523	AAGAGA	A	c.67-3906_-3902del	1	0.674				no predicted effect
94582509	AGGAG	G	c.67-3891_-3888del	1	2.650				no predicted effect
94582401	T	G	c.67-3779T>G	1	14.53	1000	352	MYBL2	no predicted effect
94582226	G	A	c.67-3604G>A	1	7.314	1000	352	MYBL2	no predicted effect
94581788	C	T	c.67-3166C>T	2	2.809	1000	242	MAFK	no predicted effect, on the same allele with p.T1253M
94580855	C	T	c.67-2233C>T	1	6.460	171			no predicted effect
94580700	G	A	c.67-2078G>A	1	8.637	171			no predicted effect
94580645	T	G	c.67-2023T>G	1	1.888	171			predicted effect on splicing, conservation issues
94578195	TG	G	c.160+333del	1	0.726				no predicted effect
94576926	C	T	c.302+68C>T	2	3.148				no predicted effect
94575178	A	G	c.303-906A>G	1	12.97				no predicted effect
94573386	G	C	c.442+747G>C	1	1.762				no predicted effect
94573340	C	T	c.442+793C>T	2	3.120				no predicted effect, on the same allele with p.[L541P;A1038V]
94573295	G	C	c.442+838G>C	1	5.989				no predicted effect
94571986	T	C	c.442+2147T>C	1	5.762	638	241	POLR2A	no predicted effect
94570537	T	C	c.443-1839T>C	1	1.824		154	BCL3	no predicted effect
94570168	G	A	c.443-1470G>A	1	2.554		236	REST	no predicted effect
94569553	G	A	c.443-855G>A	1	5.116				no predicted effect
94569519	C	T	c.443-821C>T	1	7.466				no predicted effect
94569284	G	T	c.443-586G>T	1	4.905	429			no predicted effect

94568157	C	T	c.570+414C>T	1	2.564				no predicted effect
94568063	G	A	c.570+508G>A	1	0.406				no predicted effect
94567407	C	T	c.570+1164C>T	2	9.362	589	411	GATA3	no predicted effect, on the same allele with p.V931M
94567072	C	A	c.570+1499C>A	1	8.229	589	1000	TCF7L2	no predicted effect
94566773	A	G	c.570+1798A>G	2	8.955	589	400	GATA3	predicted effect on splicing
94566254	C	T	c.571-1707C>T	1	2.075	589	1000	REST	no predicted effect
94565855	C	T	c.571-1308C>T	2	7.352				no predicted effect, on the same allele with p.V931M
94565288	C	A	c.571-741C>A	1	2.954	742			no predicted effect
94563992	C	T	c.768+358C>T	2	9.324				no predicted effect
94563912	G	A	c.768+438G>A	2	2.838				no predicted effect, on the same allele with p.R212C
94563549	T	C	c.768+801T>C	1	1.447				no predicted effect
94562756	C	T	c.768+1594C>T	1	7.069	708			no predicted effect
94561105	T	C	c.768+3245T>C	2	9.753	1000	214	CTCF	no predicted effect, on the same allele with p.V931M
94556070	C	T	c.769-7073C>T	1	9.193	251			no predicted effect
94554455	C	T	c.769-5458C>T	1	2.729				no predicted effect
94552149	CT	T	c.769-3153del	2	0.744				no predicted effect
94550775	T	C	c.769-1778T>C	2	4.085	262			no predicted effect
94549993	T	G	c.769-996T>G	1	6.050				no predicted effect
94549609	G	A	c.769-612G>A	1	5.244	280			no predicted effect
94549354	C	T	c.769-357C>T	1	6.935	280	270	GATA3	no predicted effect
94548697	C	T	c.858+211C>T	1	5.969				no predicted effect
94548645	A	T	c.858+263A>T	1	3.713				no predicted effect
94546315	T	C	c.859-41T>C	1	12.05				no predicted effect
94546283	T	C	c.859-9T>C	2	4.495				predicted effect on splicing
94545481	C	G	c.1100-464C>G	1	0.004				no predicted effect
94543664	T	A	c.1357-221T>A	1	10.52				no predicted effect
94541551	A	G	c.1554+1695A>G	1	4.221				no predicted effect
94540683	G	C	c.1554+2563G>C	1	0.848				no predicted effect
94540656	T	C	c.1554+2590T>C	2	0.574				no predicted effect
94540648	C	T	c.1554+2598C>T	2	0.118				no predicted effect
94540446	A	T	c.1554+2800A>T	1	1.665				no predicted effect
94535498	TA	A	c.1555-6626del	1	5.755	1000	1000	POLR2A	no predicted effect
94533530	T	C	c.1555-4657T>C	1	0.133		342	FOXA1	no predicted effect
94533427	G	A	c.1555-4554G>A	1	1.990	135			no predicted effect
94531734	A	C	c.1555-2861A>C	1	2.438				no predicted effect
94531618	A	G	c.1555-2745A>G	1	4.558				no predicted effect
94531573	A	C	c.1555-2700A>C	1	2.224				no predicted effect
94531233	TC	C	c.1555-2361del	1	9.491				no predicted effect
94528032	G	A	c.1937+101G>A	1	8.994	631	200	JUN	no predicted effect
94527472	C	A	c.1937+661C>A	1	8.564	347	252	POLR2A	no predicted effect
94526934	A	G	c.1938-619A>G	1	2.211		252	POLR2A	predicted effect on splicing

94526932	A	G	c.1938-617A>G	1	5.952		252	POLR2A	no predicted effect
94525509	A	G	c.2160+584A>G	1	12.53	122			predicted effect on splicing
94525483	G	A	c.2160+610G>A	1	5.854	122			no predicted effect
94525085	A	G	c.2160+1008A>G	1	7.396				no predicted effect
94522386	G	A	c.2161-8G>A	2	13.25				predicted effect on splicing
94521740	AC	C	c.2382+416del	1	7.093				no predicted effect
94520496	G	A	c.2587+171G>A	1	5.440				no predicted effect
94517827	G	A	c.2588-573G>A	1	1.824				no predicted effect
94517010	C	T	c.2653+179C>T	1	5.574		209	MAFK	no predicted effect
94514521	T	G	c.2654-8T>G	1	11.30	706	373	POLR2A	predicted effect on splicing
94513208	G	A	c.2744-559G>A	2	3.835				no predicted effect, on the same allele with p.V931M
94512967	C	T	c.2744-318C>T	1	0.173				no predicted effect
94511126	T	A	c.2919-826T>A	1	1.513	1000	541	POLR2A	predicted effect on splicing, conservation issues
94510881	A	G	c.2919-581A>G	1	0.344				no predicted effect
94510549	C	T	c.2919-249C>T	1	1.372				no predicted effect
94509901	C	G	c.3050+268C>G	1	0.054	861	295	MYBL2	no predicted effect
94509799	C	T	c.3050+370C>T	1	1.072	861	295	MYBL2	predicted effect on splicing
94508746	C	T	c.3190+146C>T	1	3.969				no predicted effect
94507654	TTTTG	G	c.3328+659_+662del	1	0.590				no predicted effect
94507654	TG	G	c.3328+662del	1	0.526				no predicted effect
94507634	C	T	c.3329-676C>T	1	1.688				no predicted effect
94507338	G	A	c.3329-380G>A	1	7.181				no predicted effect
94504333	A	T	c.3607+1266A>T	1	3.619	1000	760	POLR2A	no predicted effect
94504010	AG	G	c.3608-1105del	1	6.471		760	POLR2A	no predicted effect
94503655	T	G	c.3608-749T>G	1	8.353		770	NFIC	no predicted effect
94501750	A	G	c.3862+546A>G	2	3.088				no predicted effect, on the same allele with p.V931M
94500855	C	T	c.3862+1441C>T	1	6.518				no predicted effect
94499842	T	C	c.3863-2243T>C	1	1.321	969			no predicted effect
94498982	T	G	c.3863-1383T>G	1	0.569				no predicted effect
94495923	G	A	c.4352+61G>A	1	6.914				predicted effect on splicing, conservation issues
94493272	G	T	c.4539+1729G>T	1	1.586				predicted effect on splicing
94493030	T	C	c.4539+1971T>C	1	15.14	192			no predicted effect
94493000	G	A	c.4539+2001G>A[a]	2	4.653	192			no predicted effect
94492973	C	T	c.4539+2028C>T[a]	2	3.531	192			no predicted effect
94492937	C	T	c.4539+2064C>T	4	3.598	192			the most frequent new variant, no predicted effect
94491148	C	T	c.4540-544C>T	1	2.482	342	151	CTCF	no predicted effect
94486507	T	C	c.5018+289T>C	1	6.814	1000	177	GATA2	no predicted effect
94486092	AATTC	C	c.5018+700_+703del	1	7.484				no predicted effect
94484903	G	A	c.5196+235G>A	1	4.845				no predicted effect
94484367	G	A	c.5196+771G>A	1	2.188				no predicted effect

94484082	A	G	c.5196+1056A>G[a]	3	2.025				predicted effect on splicing
94484001	G	A	c.5196+1137G>A[a]	1	2.681				predicted effect on splicing
94483524	G	A	c.5196+1614G>A	1	2.794	172			no predicted effect
94479748	CTTCCTTCTTTCT	T	c.5460+339_+350del	1	3.644				no predicted effect
94478330	C	A	c.5461-1389C>A	4	4.805	841	362	SIN3A	the most frequent new variant, no predicted effect, on the same allele with c.4539+2064C>T
94477890	G	A	c.5461-949G>A	1	0.622	513	259	ZNF263	no predicted effect
94477216	G	A	c.5461-275G>A	1	3.496	197			no predicted effect
94475341	T	C	c.5715-914T>C	1	4.189	356			no predicted effect
94473856	C	A	c.5836-3C>A	1	11.84				predicted effect on splicing
94471747	T	A	c.6006-609T>A	3	7.220				interesting variant based on frequency, no predicted effect
94471443	C	A	c.6006-305C>A	1	1.588				no predicted effect
94470470	T	C	c.6147+527T>C	1	6.293		1000	CEBPB	no predicted effect
94468019	C	T	c.6148-471C>T	1	8.470				predicted effect on splicing
94466914	A	G	c.6283-253A>G	1	7.705				no predicted effect
94464822	C	T	c.6480-1156C>T	1	0.424				no predicted effect
94464821	G	A	c.6480-1155G>A	1	0.207				no predicted effect
94462550	C	G	c.6730-799C>G	2	1.574		317	CEBPB	no predicted effect, on the same allele with c.5196+1056A>G
94462292	T	C	c.6730-541T>C	1	5.682				predicted effect on splicing
94462243	TA	A	c.6730-493del	1	5.138				no predicted effect
94461930	G	A	c.6730-179G>A	1	5.704				no predicted effect
94460488	G	C	c.6816+1177G>C	1	0.352				no predicted effect
94459907	C	AC	c.6817-1110_-1109insA	1	3.637				no predicted effect
94457747	G	A	c.*1046G>A	1	3.807				no predicted effect
94456653	G	T	c.*2140G>T	1	0.805				no predicted effect
94456489	C	A	c.*2304C>A	1	4.194				no predicted effect
94454413	T	A	c.*4380T>A	1	1.856	706	300	REST	no predicted effect

[a] Variant is previously reported by Braun et al 2013.

Supplemental Table 2. Detailed information of the 8X60K custom aCGH microarray. The probe densities are differentially distributed according to the genetic loci. The “primary loci” are more intensively covered than the “secondary loci”. Control probes serve as the “backbone” of the array for the normalization purpose.

	Region	Size (kb)	Approximate average probe spacing (bp) or number of probes
Primary loci (n=9) [*ABCA4*, *ELOVL4*, *PRPH2* (*RDS*), *BEST1 (VMD2*), *RS1*, *CNGB3*, *CFH* locus[a], *C2*/*CFB* locus[b], *ARMS2* locus (*ARMS2*/*HTRA1*)]	Gene	Variable	50
	"Promoter"	10	50
	High density 3'	10	50
	5' conserved or to next gene or to 100kb from gene[c]	100+	100
	3' conserved or to next gene or to 100kb from gene[c]	100+	100
	5' flanking for breakpoint mapping	900	5000
	3' flanking for breakpoint mapping	900	5000
Secondary loci (n=18) [*C3*, *APOE*, *CFI*, *LIPC*, *SYN3*/ *TIMP3*, *CETP*, *COL8A1*, *BBX*, *PLD1*, *SPEF2*, *ADAM19*, *VEGFA*, *FRK*, *MEPCE*, *CHMP7*/*LOXL2*, *TGFBR1*, *NPS*, *PICK1*]	Gene	Variable	400
	"Promoter"	10	400
	High density 3'	10	400
	5' flanking for breakpoint mapping	500	5000
	3' flanking for breakpoint mapping	500	5000
	Exons	Variable	8 probes/exon
Control probes	Spread throughout genome	-	5000 probes

[a] From *CFH* to *F13B*.
[b] From *C2* to *C4A*/*C4B*.
[c] Whichever is longer.

A c.5196+1056A>G

B c.570+1798A>G

C c.2161-8G>A

D c.859-9T>C

E c.5836-3C>A

F c.2654-8T>G

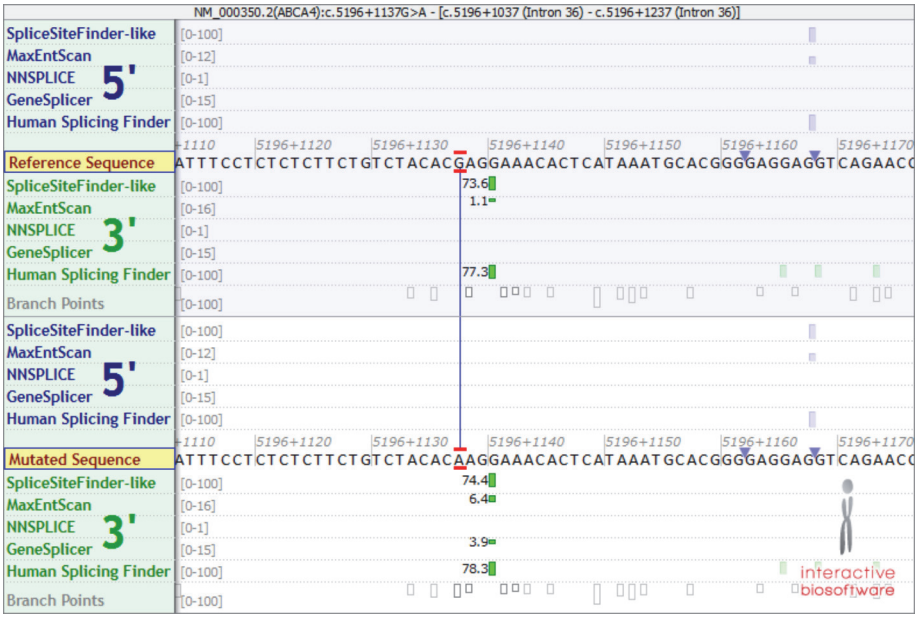
G c.1938-619A>G

H c.3050+370C>T

I c.2160+584A>G

J c.4539+1729G>T

K c.5196+1137G>A



L c.6148-471C>T

M c.6730-541T>C

Supplemental Figure 1. Examples of the possible predicted effect of new non-coding *ABCA4* variants on splicing, assessed using 5 different algorithms (SpliceSiteFinder, MaxEntScan, NNSPLICE, GeneSplicer, Human Splicing Finder) via Alamut software (http://www.interactive-biosoftware.com).

Publication III

Zernant J, Lee W, Collison FT, Fishman GA, Sergeev YV, Schuerch K, Sparrow JR, Tsang SH, Allikmets R (2017). Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration. J Med Genet. 54(6):404-412.

ORIGINAL ARTICLE

Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration

Jana Zernant,[1] Winston Lee,[1] Frederick T Collison,[2] Gerald A Fishman,[2] Yuri V Sergeev,[3] Kaspar Schuerch,[1] Janet R Sparrow,[1,4] Stephen H Tsang,[1,4] Rando Allikmets[1,4]

► Additional material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/jmedgenet-2017-104540).

[1]Department of Ophthalmology, Columbia University, New York, New York, USA
[2]The Pangere Center for Hereditary Retinal Diseases, The Chicago Lighthouse, Chicago, Illinois, USA
[3]Ophthalmic Genetics and Visual Function Branch, National Eye Institute, National Institutes of Health, Bethesda, MD, USA
[4]Department of Pathology & Cell Biology, Columbia University, New York, New York, USA

Correspondence to
Dr Rando Allikmets, Department of Ophthalmology, Columbia University, Edward S. Harkness Eye Institute Research Annex, 635 West 165th Street, Box # 28, New York, NY 10032, USA; rla22@columbia.edu

JZ and WL contributed equally.

Received 16 January 2017
Revised 28 February 2017
Accepted 7 March 2017
Published Online First 26 April 2017

ABSTRACT

Background Variation in the *ABCA4* gene is causal for, or associated with, a wide range of phenotypes from early onset Mendelian retinal dystrophies to late-onset complex disorders such as age-related macular degeneration (AMD). Despite substantial progress in determining the causal genetic variation, even complete sequencing of the entire open reading frame and splice sites of *ABCA4* identifies biallelic mutations in only 60%–70% of cases; 20%–25% remain with one mutation and no mutations are found in 10%–15% of cases with clinically confirmed ABCA4 disease. This study was designed to identify missing causal variants specifically in monoallelic cases of ABCA4 disease.

Methods Direct sequencing and analysis were performed in a large familial ABCA4 disease cohort of predominately European descent (n=643). Patient phenotypes were assessed from clinical and retinal imaging data.

Results We determined that a hypomorphic *ABCA4* variant c.5603A>T (p.Asn1868Ile), previously considered benign due to high minor allele frequency (MAF) (~7%) in the general population, accounts for 10% of the disease, >50% of the missing causal alleles in monoallelic cases, ~80% of late-onset cases and distinguishes ABCA4 disease from AMD. It results in a distinct clinical phenotype characterised by late-onset of symptoms (4th decade) and foveal sparing (85%). Intragenic modifying effects involving this variant and another, c.2588G>C (p.Gly863Ala) allele, were also identified.

Conclusions These findings substantiate the causality of frequent missense variants and their phenotypic outcomes as a significant contribution to ABCA4 disease, particularly the late-onset phenotype, and its clinical variation. They also suggest a significant revision of diagnostic screening and assessment of *ABCA4* variation in aetiology of retinal diseases.

INTRODUCTION

Mutations in the *ABCA4* gene (STGD1, MIM #248 200)[1] are the most frequent cause of Mendelian inherited recessive retinal dystrophies with phenotypes ranging from late-onset mild cases of central atrophy to very early onset, retinitis pigmentosa (RP)-like, panretinal degeneration.[2–5] Similar to phenotypic variability, the disease-associated genetic variation in *ABCA4* is also extensive.[6] Currently, >1000 definitely or possibly disease-causing variants have been determined in the coding sequences and splice sites of the *ABCA4* gene (JZ and RA, unpublished data). While the understanding of genetic variation in *ABCA4* has substantially improved since the discovery of the gene in 1997, diagnostic screening in patients diagnosed with ABCA4 disease (often collectively referred to as Stargardt's disease) is successful in only 60%–70% of cases where both disease-causing alleles are identified.[7 8] The remaining cases include patients with one mutation (20%–25%) and no mutation (10%–15%).[7] The population frequency of *ABCA4* alleles is high—1:20 people carry a potentially pathogenic *ABCA4* allele.[7 9] Therefore, recent efforts have been directed towards two goals; first, to find the missing *ABCA4* disease-causing alleles in non-coding sequences[8] and, second, to perform extensive whole exome sequencing (WES) in patients with 0 *ABCA4* alleles (and ABCA4-like phenotypes) to differentiate non-*ABCA4* phenocopies.[10–13] Substantial progress has been made with respect to both objectives: several deep intronic *ABCA4* variants have been proven to affect splicing[8 14] and preliminary data (not shown) indicate that such is also true of rare variants affecting *ABCA4* expression. We have also shown that large CNVs, which elude detection by sequencing, are exceptionally rare in the *ABCA4* locus and, as a result, account for only a small fraction of 'missing' alleles.[15] WES of patients with 0 *ABCA4* mutations has uncovered causal genes in ~70% of cases with *ABCA4*-like phenotypes, including known retinal disease genes (*CRB1*, *CRX*, etc)[11 13] and new genes (*RAB28*, *RDH11*).[10 12] However, despite all these advances, the second causative mutation in most patients with 1 and 0 *ABCA4* alleles remains unidentified following complete sequencing.[8] This study demonstrates that some known, frequent *ABCA4* variants, which have been considered benign due to high minor allele frequency (MAF) in the general population, are in fact hypomorphic alleles, which result in disease expression under certain conditions.

MATERIALS AND METHODS

Patient cohort

All study subjects were consented before participating in the study under the protocols #AAAI9906 approved by the Institutional Review Board at Columbia University and #20130770 by the

To cite: Zernant J, Lee W, Collison FT, *et al*. *J Med Genet* 2017;**54**:404–412.

Western Institutional Review Board at the Chicago Lighthouse. The study adhered to tenets established in the Declaration of Helsinki. Complete ophthalmic examinations were provided by a retinal specialist (SHT and GAF), including slit-lamp and dilated fundus examinations. Clinical assessments (WL and FTC) 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 excitation) and viewing the resultant fluorescence through a band pass filter with a short wavelength cut-off at 495 nm. Colour fundus photos were obtained with a FF 450plus Fundus Camera (Carl Zeiss Meditec AG, Jena, Germany) and CR-1 Mark II Fundus Camera (Canon, Tokyo, Japan).

Sequence analysis

The *ABCA4* gene and locus sequencing were performed as previously described, using the Illumina TruSeq Custom Amplicon protocol (Illumina, San Diego, California, USA), followed by sequencing on Illumina MiSeq platform, or the RainDance microdroplet-PCR target enrichment (RainDance Technologies, Billerica, Massachusetts, USA), with subsequent sequencing on Roche 454 platform.[8 15] The next-generation sequencing reads were analysed and compared with the *ABCA4* sequence in the reference genome GRCh37/hg19, using the variant discovery software NextGENe (SoftGenetics, State College, Pennsylvania, USA). All detected possibly disease-associated variants and their segregation with the disease in available family members were analysed and confirmed by Sanger sequencing. Nucleotide positions and protein translation correspond to CCDS747.1 and NP_000341.2, respectively. Nucleotide numbering uses the A of the ATG translation initiation start site as nucleotide 1. The possible effect of all *ABCA4* variants was assessed as described in detail previously.[8 15] The variant prediction algorithms were accessed via Alamut Visual version 2.7 (Interactive Biosoftware, Rouen, France). The allele frequencies of all variants were compared with the gnomAD database,[16] and The 1000 Genomes Project database.[17] All *ABCA4* variants reported in this manuscript were submitted to Leiden Open Variation Database V.3.0 (www.lovd.nl/3.0/home).

Molecular modelling

Amino acid sequence P78363 (ABCA4_HUMAN) was obtained from the Uniprot database (http://www.uniprot.org). Two transmembrane and nucleotide-binding motifs of the human ABCA4 structure were built by homology modelling similar to that of previously described for the RS1 protein.[18 19] Briefly, the homology model of structure of transmembrane and nucleotide-binding domains of ABCA4 was built using the automatic segment matching and self-consistent ensemble optimisation by the Gene Mine Look program, V.3.5.2.[18] Bacterial lipid 'flippase' MsbA was chosen from the Crystallographic Database as a structural template (PDB: 3b60). Hydrogen atoms were added to the structure and the predicted ABCA4 structure was 2 ns equilibrated in water using molecular dynamics (MD). MD simulations were performed using the Impact module of the Maestro program package (V.8.0.308, Schrodinger, New York, New York, USA). The quality of the structure was accessed with the PROCHECK program. Protein unfolding propensities for mutant variants p.Asn1868Ile and p.Gly863Ala were evaluated using the unfolding mutation screen as previously described in detail.[20 21]

RESULTS

Analysis of genetic data

The study cohort consisted of 643 individuals of (mostly Eastern) European descent. Of these, 2 *ABCA4* mutations were identified in 437 cases (68%), 1 mutation in 117 cases (18%) and 0 mutations in 89 cases (14%) (see online supplementary table 1), leaving ~23% of disease-associated alleles in 32% of patients yet to be identified. Almost all patients with no *ABCA4* mutations and ~50% of patients with 1 mutation have been screened by whole exome sequencing to determine if variants in other genes were causal in these cases. All cases, where disease-associated variants in other genes were detected, were excluded from this cohort.

The *ABCA4* gene harbours many common missense variants with MAFs >5% in the general population. The c.5603A>T (p.Asn1868Ile) allele was present in patients with ABCA4 disease three to four times more frequently than expected (~20% vs 6.6%), an observation which had also been previously reported.[22 23] The allele frequency of the p.Asn1868Ile variant is ~6.6% in Europe, 2% in Latinos and South Asians, and 1% in Africans (gnomAD). This allele was also observed to be in *cis* with several more frequent *ABCA4* disease-causing alleles, notably p.5461–10T>C (p.?) and c.2588G>C (p.Gly863Ala) (table 1 and see online supplementary table 1). Therefore, we assumed that the more frequent occurrence of p.Asn1868Ile in patients with ABCA4 disease is due to linkage disequilibrium with other truly pathogenic alleles. However, when complex alleles with known pathogenic mutations were excluded from these calculations, the allele frequency in patients remained significantly higher.

In 643 screened patients, p.Asn1868Ile was detected in 176 cases (197 alleles); 21 cases were homozygous and 155 cases heterozygous. When dividing the cohort into three groups based on the number of detected *ABCA4* pathogenic alleles as described above, the fractions of the p.Asn1868Ile alleles were 11.6% (101/874) in 2 mutation cases, 33% (77/234) in 1 mutation cases and 10.6% (19/178) in 0 mutation cases. Therefore, the allele frequency of p.Asn1868Ile was two times higher as compared with the matched general population in cases with 2 and 0 mutations and 5X higher in cases with 1 mutation (33% vs 6%; $p<0.0001$). More frequent disease-causing alleles which always carried p.Asn1868Ile in *cis* were c.5461–10T>C (p.?), c.4469G>A (p.Cys1490Tyr), c.4594G>A (p.Asp1532Asn) and c.2588G>C (p.Gly863Ala) (see online supplementary table 1). Other, rarer disease-causing variants which were found allelic with p.Asn1868Ile after determining the phase included c.319C>T (p.Arg107*), c.1253T>C (p.Phe418Ser), c.2552G>A (p.Gly851Asp), c.5114G>A (p.Arg1705Gln), c.5318C>T (p.Ala1773Val) and c.5572T>A (p.Tyr1858Asn).

After accounting for known disease-associated alleles that harbour the p.Asn1868Ile variant in *cis*, we determined that over half of the cases with 1 mutation (67/117, 57.3%) carried the variant in *trans* (table 1). We then assessed all 67 cases of whom 57 were heterozygous for the p.Asn1868Ile variant and 10 were homozygous. The homozygous fraction included cases with p.Asp1868Ile as a complex allele, mostly due to the c.5461–10T>C variant (n=6, 60%) in *trans*. Six other heterozygous

Table 1 Genotypes of cases with single *ABCA4* pathogenic variants and the c.5603A>T (p.Asn1868Ile) status

c.5603A>T genotype	Pathogenic *ABCA4* variant	No. of cases	c.5603A>T genotype	Pathogenic *ABCA4* variant	No. of cases
++	c.1027_1028del (p.Asn343*)	1	+/–	c.6229C>T (p.Arg2077Trp)	2
++	c.2069del (p.Gly690fs)	1	+/–	c.6386+1G>A (p.?)	1
++	c.3754G>T (p.Glu1252*)	1	+/–	c.6386+2C>G (p.?)	1
++	c.4469G>A (p.Cys1490Tyr)	1	+/–	c.6543_6578del (p.Leu2182_Phe2193del)	1
++	c.5461–10T>C (p.?)	6	–/–	c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	1
+/–	c.45G>A (p.Trp15*)	1	–/–	c.[3758C>T;5882G>A] (p.[Thr1253Met;Gly1961Glu])	1
+/–	c.52C>T (p.Arg18Trp)	1	–/–	c.[5882G>A;6445C>T] (p.[Gly1961Glu;Arg2149*])	1
+/–	c.161–1G>A (p.?)	1	–/–	c.[735T>G;2300T>A] (p.[Tyr245*;Val767Asp])	1
+/–	c.161G>A (p.Cys54Tyr)	2	–/–	c.161G>A (p.Cys54Tyr)	1
+/–	c.247_250dup (p.Ser84fs)	2	–/–	c.184C>T (p.Pro62Ser)	1
+/–	c.262G>A (p.Gly88Arg)	1	–/–	c.247_250dup (p.Ser84fs)	1
+/–	c.571–1G>T (p.?)	1	–/–	c.634C>T (p.Arg212Cys)	2
+/–	c.655A>T (p.Arg219*)	1	–/–	c.733T>C (p.Tyr245His)	1
+/–	c.868C>T (p.Arg290Trp)	1	–/–	c.1034A>G (p.Tyr345Cys)	1
+/–	c.1222C>T (p.Arg408*)	1	–/–	c.1699G>A (p.Val567Met)	1
+/–	c.[1622T>C;3113C>T] (p.[L541P;A1038V])	8	–/–	c.2041C>T (p.Arg681*)	1
+/–	c.1804C>T (p.Arg602Trp)	1	–/–	c.2616del (p.Leu274fs)	1
+/–	c.1903C>T (p.Gln635*)	1	–/–	c.2966T>C (p.Val989Ala)	2
+/–	c.1906C>T (p.Gln636*)	1	–/–	c.2971G>C (p.Gly991Arg)	1
+/–	c.1921T>C (p.Cys641Arg)	1	–/–	c.3043T>C (p.Phe1015Leu)	1
+/–	c.1957C>T (p.Arg653Cys)	1	–/–	c.3113C>T(p.Ala1038Val)	2
+/–	c.2588G>C (p.Gly863Ala)	2	–/–	c.3148G>A (p.Gly1050Ser)	2
+/–	c.2894A>G (p.Asn965Ser)	2	–/–	c.3292C>T (p.Arg1098Cys)	1
+/–	c.3259G>A (p.Glu1087Lys)	1	–/–	c.3322C>T (p.Arg1108Cys)	1
+/–	c.3261A>C (p.Glu1087Asp)	1	–/–	c.4139C>T (p.Pro1380Leu)	1
+/–	c.3328+1G>A (p.?)	1	–/–	c.4248C>A (p.Phe1416Leu)	1
+/–	c.[3481C>A;5714+5G>A] (p.[Arg1161Ser;?])	1	–/–	c.4353–4T>C (p.?)	1
+/–	c.3812A>G (p.Glu1271Gly)	1	–/–	c.4577C>T (p.Thr1526Met)	2
+/–	c.3846del (p.Gly1283fs)	1	–/–	c.4793C>A (p.Ala1598Asp)	1
+/–	c.4139C>T (p.Pro1380Leu)	2	–/–	c.4918C>T (p.Arg1640Trp)	1
+/–	c.[4222T>C;4918C>T] (p.[Trp1408Arg;Arg1640Trp])	1	–/–	c.5312+1G>A (p.?)	1
+/–	c.4234C>T (p.Gln1412*)	2	–/–	c.5714+5G>A (p.?)	3
+/–	c.4249_4251del (p.Phe1417del)	1	–/–	c.5761G>A (p.Val1921Met)	1
+/–	c.4670A>G (p.Tyr1557Cys)	1	–/–	c.5881G>A (p.Gly1961Arg)	1
+/–	c.4918C>T (p.Arg1640Trp)	2	–/–	c.5882G>A (p.Gly1961Glu)	5
+/–	c.5196+1G>A (p.?)	1	–/–	c.6088C>T (p.Arg2030*)	1
+/–	c.5395A>G (p.Asn1799Asp)	1	–/–	c.6089G>A (p.Arg2030Gln)	1
+/–	c.5461–10T>C (p.?)	4	–/–	c.6229C>T (p.Arg2077Trp)	1
+/–	c.5923G>C (p.Gly1975Arg)	1	–/–	c.6320G>A (p.Arg2107His)	4
+/–	c.6166A>T (p.Lys2056*)	1	–/–	c.6694G>A (p.Glu2232Lys)	1

Nucleotide positions and protein translation correspond to CCDS747.1 and NP_000341.2, respectively. Nucleotide numbering uses the A of the ATG translation initiation start site as nucleotide 1. Number of cases indicates the number of individuals with identical genotypes.

cases harboured the p.Asn1868Ile allele in *cis* with a known mutation; the remainder were in *trans* as determined by phase analysis in available families (table 1 and supplementary table 1). The haplotype containing the c.5603A>T (p.Asn1868Ile) allele has been established.[22] In cases where phase determination was impossible due to the absence of family members, the haplotype was identified by subtracting the genotype of the known allele. Further examination established that the mutations in *trans* with the p.Asn1868Ile variant are deleterious, predominantly null mutations based on *in silico* predictions and experimental data (tables 1 and 2). The breakdown of mutations in *trans* with p.Asn1868Ile include the following: 6 c.5461–10T>C, 9 stop mutations, 7 small indels resulting in a frameshift, 6 severe splicing-altering variants (eg, at +/−1 and +/−2 positions) and 33 missense mutations, including well-known severe variants such as two cases of c.2894A>G (p.Asn965Ser)[24] and eight cases of the complex allele c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val]).[25–27] Among others, seven cases with mutations of arginine to tryptophan—c.52C>T (p.Arg18Trp), c.868C>T (p.Arg290Trp), c.1804C>T (p.Arg602Trp), c.4918C>T (p.Arg1640Trp) and c.6229C>T (p.Arg2077Trp) especially stood out (table 1).

Sequencing of the entire *ABCA4* genomic locus had been performed for 93/117 monoallelic *ABCA4* cases as previously described.[8] In many cases with a single pathogenic *ABCA4* variant, another confirmed or highly likely pathogenic *ABCA4* deep intronic variant(s) were identified.[8] However, none of the 61 cases with the p.Asn1868Ile allele harboured any definitely

Table 2 Type and frequency of the allele in *trans* for selected *ABCA4* disease alleles

	Type and fraction (%) of the allele in *trans*			
ABCA4 disease allele (no. of cases)	Deleterious (X, FS, splice)	Deleterious missense	Non-deleterious missense	Hypomorphic
p.Asn1868Ile (60)	46.7	53.3	0	0
p.[Gly863Ala;Asn1868Ile] (35)	51.4	22.9	22.9	2.9
p.Gly1961Glu (118)	40.7	32.2	20.3	6.8
p.Arg2107His (21)	23.8	33.3	38.1	4.8
c.5461−10T>C (p.?) (36)	11.1	30.6	25	33.3

FS, frameshift mutation; splice, mutation affecting splice site; X, stop mutation.

or very likely pathogenic deep intronic variants, suggesting that the disease phenotype was indeed conferred by the p.Asn1868Ile variant in *trans* with a deleterious *ABCA4* mutation.

The frequency of another presumed 'mild' p.Gly863Ala variant is high in populations of European descent: ~2% in populations of Dutch and Swedish ancestry[22 28] and ~0.8% in all Europeans (gnomAD), therefore, it was considered hypomorphic, that is, requiring to be paired, in *trans*, with a 'severe' *ABCA4* allele in order to present phenotypically.[28] This is in part true, the analysis of our dataset of 643 patients (table 2) suggests that the *ABCA4* allele in *trans* from p.Gly863Ala is deleterious or very likely deleterious in >70% of cases. However, we have encountered at least two exceptions of patients homozygous for p.Gly863Ala and in some cases the opposing allele is known to be mild, for example, the missense alleles c.5882G>A (p.Gly1961Glu), c.2947A>G (p.Thr983Ala) and c.4919G>A (Arg1640Gln) (see online supplementary table 1).

The p.Gly863Ala allele is always in *cis* with the p.Asn1868Ile in patients with ABCA4 disease, which lead us to initially consider the latter to be a benign allele on the same haplotype, as also observed by Maugeri *et al* in 2002,[22] where disease-associated p.Gly863Ala haplotypes were defined. When analysing data from our large cohorts of patients with age-related macular degeneration (AMD) and matched controls,[29–31] we observed that ~90% of cases with the p.Gly863Ala variant did not carry the p.Asn1868Ile allele. Analysis of these and previously published data[22] revealed that the p.Asn1868Ile variant is allelic with the p.Gly863Ala in only ~13% of cases in populations of European ancestry.

Further evidence of the complex allele c.[2588G>C;5603A>T] (p.[Gly863Ala;Asn1868Ile]) being disease-associated and the p.Gly863Ala alone not, comes from the calculation of expected and observed allele frequencies in our patient cohort and matched general population. If we consider the frequency of the complex allele, p.[Gly863Ala;Asn1868Ile] to be ~13% of all cases carrying the p.Gly863Ala allele, which frequency is ~1% on average in the general population of European descent, then our cohort of 643 cases should harbour 13 cases with the p.Gly863Ala allele and 2 cases with the complex allele p.[Gly863Ala;Asn1868Ile], if neither of these would be disease-causing. In fact, we observe 35 cases (17X more of the expected at random) of the complex allele ($p<0.0001$) and only 4 cases (3X less of the expected at random) of the simplex allele ($p<0.0001$). These data additionally support the pathogenicity of the complex p.[Gly863Ala;Asn1868Ile] allele and not of the p.Gly863Ala alone.

Finally, we determined that 2 out of 643 patients were homozygous for the p.[Gly863Ala;Asn1868Ile] complex allele and none for the p.Gly863Ala allele alone. Since the population frequency of the p.Gly863Ala allele is 1% and the complex allele ~0.1%, the random occurrence of homozygotes is 1:10 000 and 1:1 000 000, respectively. Since we did not detect, among our cohort of patients with ABCA4 disease, any homozygotes for the simplex allele and found two for the complex allele, the odds for p.Gly863Ala being disease-causing without the p.Asn1868Ile allele are extremely low.

Genotype-phenotype correlations

Although previously described as clinically 'mild' alleles, patients harbouring the c.5882G>A (p.Gly1961Glu) or c.2588G>C (p.Gly863Ala) alleles have substantially different phenotypes. Patients with p.Gly1961Glu indeed exhibit milder disease expression,[32 33] although not in the overall rate of disease progression but in a distinct phenotypic pattern that, interestingly, overlaps with patients harbouring p.Asn1868Ile. ABCA4 disease invariably begins as a maculopathy with an enlarging lesion of outer retinal atrophy and accumulation of yellow foci, or flecks, at the level of the retinal pigment epithelium (RPE). Patients with most other *ABCA4* disease alleles exhibit progressively severe fleck patterns from very few in the macula to a stage of 'absolute confluence' across the posterior pole between the ages 30 and 40 years (figure 1).

Patients harbouring p.Gly1961Glu or p.Asn1868Ile consistently exhibit milder spatiotemporal fleck patterns even at advanced age, which never progresses to the absolute confluence stage as illustrated by fundus autofluorescence imaging which detects lipofuscin accumulation (figure 1). The resistance to fleck accumulation as seen in p.Gly1961Glu or p.Asn1868Ile patients is further exemplified by the occurrence of a particular phenotype that is seen only in this patient group where a well-defined, unifocal lesion of dark atrophy ranging from 8.6 to 56.4 mm^2 in size with proximally bordering, lesion-centric flecks appears (see online supplementary figure 1). Similar lesions of dark atrophy are typically numerous (multifocal) and observed in the background of more advanced fleck stages with most other ABCA4 disease cases (figure 1). Additionally, the individual morphology of flecks in p.Gly1961Glu or p.Asn1868Ile appear to be distinct in that they are predominantly larger in size and are well-defined in shape and generally more sparsely distributed (see online supplementary figure 1).

Despite sharing a similar phenotypic trajectory with patients carrying p.Gly1961Glu with respect to fleck formation and lesion progression, patients harbouring the p.Asn1868Ile variant also exhibit distinct clinical characteristics. ABCA4 disease is predominantly juvenile-onset as patients (including those with p.Gly1961Glu) report visual symptoms within the first and second decades of life; however, p.Asn1868Ile patients report a significantly delayed onset within the fourth decade of life (mean age=36.3 years; $p<0.0001$) (figure 2A). The prevalence of foveal sparing, defined as the structural and function preservation of outer retinal layers in the fovea despite the progressing atrophy of the macula, was observed in approximately one-third of patients with ABCA4 disease, but in 84.7% of p.Asn1868Ile

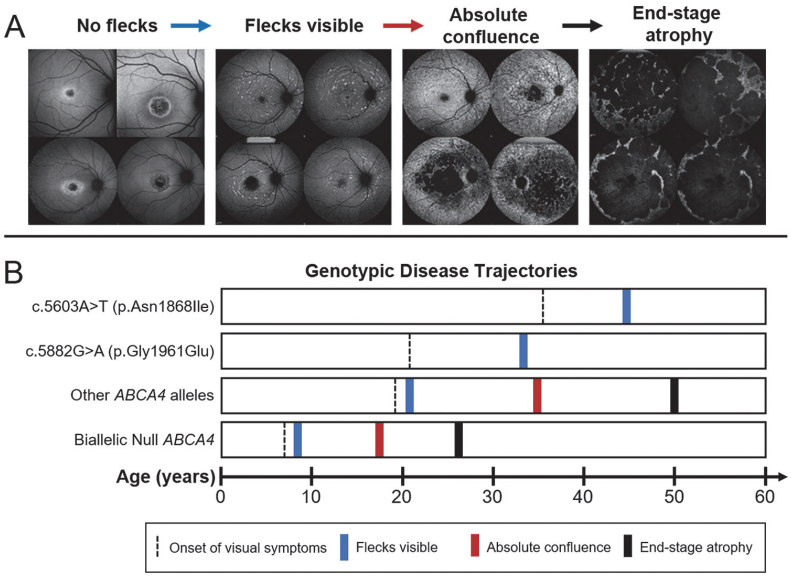


Figure 1 Summary of disease trajectories associated with autofluorescent fleck accumulation in ABCA4 disease. (A) Autofluorescence images illustrating milestones of fleck accumulation beginning with confined or bull's eye lesions with no visible flecks, to early accumulations within and around the macula, state of absolute confluence with fleck atrophy and end-stage of multiple coalescing lesions across the posterior pole. (B) Timeline of milestones (mean age) disease trajectories with respect to genotype groups. Symptomatic onset (black dashed line) varies across genotypic groups. Notably, patients harbouring c.5882G>A (p.Gly1961Glu) and c.5603A>T (p.Asn1868Ile) present at a significantly later age as compared with patients harbouring other *ABCA4* variant combinations. Both c.5882G>A (p.Gly1961Glu) and c.5603A>T (p.Asn1868Ile) patient groups exhibit a delayed onset of visible flecks in the retina (blue line) and, notably, do not progress to the absolute confluence (red line) or end-stage stage (black line) phenotypes as seen in all other cases.

patients (figure 2B). Visual function and acuity is often preserved in cases where the outer retinal layers of the fovea (RPE, ellipsoid zone and external limiting membrane) (figure 2C and D) are structurally intact within a region of macular atrophy, which may be connected to the delay in perception of visual symptoms within this group.

As expected, the phenotypes of biallelic patients with p.Asn1868Ile in *cis* with other mutations (eg, c.5461–10T>C, p.Cys1490Tyr, etc) were typically consistent with the effect of overall genotype and presented with mostly early onset, severe phenotypes. As a complex allele, p.Asn1868Ile/p.Gly863Ala acted as a fully penetrant allele with a moderate effect and resulted in variable phenotypes, which appeared to be, as expected, reflective also of the variant on the opposite allele. Intermediate phenotypes were associated with missense variants, for example, c.2947A>G (p.Thr983Ala) and c.1964T>G (p.Phe655Cys), while advanced cases were associated with deletions and splice site variants, for example, c.5778_5779del (p.Arg1927fs) and c.6088C>T (p.Arg2030*). The only two patients homozygous for the complex p.[Gly863Ala;Asn1868Ile] allele presented with later onset and milder disease and foveal sparing in one of the two cases.

Late-onset ABCA4 disease and AMD

There is an ongoing discussion about the prevalence and role of heterozygous *ABCA4* alleles in a late-onset, common disease, AMD.[25 31 34–37] While the answer is still not unequivocal—some studies find statistically significantly elevated *ABCA4* alleles in AMD and some do not—it has been suggested that many of AMD cases are actually misdiagnosed late-onset patients with ABCA4 disease.[37 38] It has also been suggested that patients with late-onset ABCA4 disease remain monoallelic after complete *ABCA4* sequencing much more often than those with early onset disease.[38] Finally, data on phenotypes of individuals who represent true carriers, such as parents of patients with ABCA4 disease, are also very conflicting, where some, including us[39] suggest that there is no detectable disease phenotype or increased lipofuscin accumulation, at least up to 60 years of age, some other studies have frequently found disease phenotypes in parents of STGD1 patients.[40 41] While unequivocal answers to these questions are beyond of the scope of this study, we queried the frequencies and composition of *ABCA4* alleles in late-onset ABCA4 disease, in patients with AMD, and in matched controls.

We defined late onset as ≥45 years of age, in accordance with previous studies,[38] who had found in a small cohort of 21 unrelated patients that a greater proportion of late-onset cases were monoallelic (52%) than biallelic (38%) from which the authors proposed that slower disease progression from defects on just one disease allele may eventually cause ABCA4 disease. When examining our larger group of patients for whom similar disease onset data were available, an even larger disparity among monoallelic and biallelic late-onset cases; 27.6% vs 7.7%, respectively, was observed. However, following the discovery of p.Asn1868Ile to be the second causal variant in ~80% of these monoallelic late-onset cases, a similar fraction of the late-onset phenotype (~10%) was observed in both monoallelic and biallelic cases of the late-onset disease. The remaining missing alleles in late onset, monoallelic disease cases could be another hypomorphic variant or, more likely, another yet unknown modifier allele. Our data therefore do not corroborate an association between late-onset

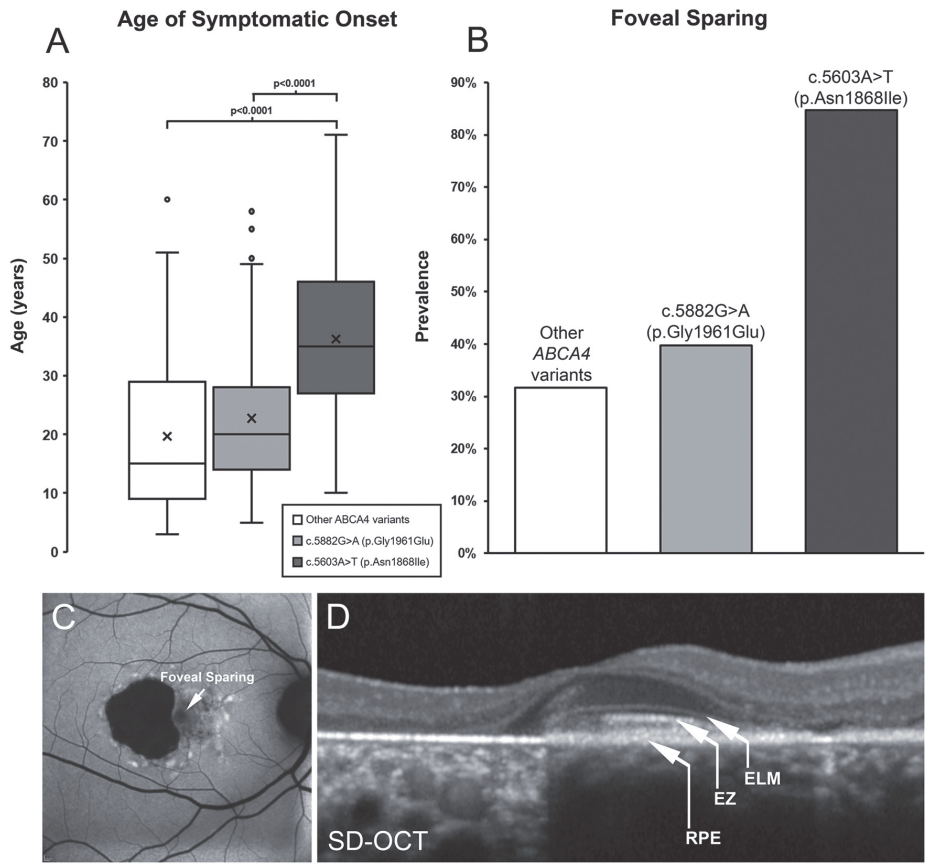


Figure 2 Summary of phenotypic distinction in patients harbouring the c.5603A>T (p.Asn1868Ile) allele of *ABCA4*. (A) Patients harbouring biallelic *ABCA4* variants, including null alleles, and c.5882G>A (p.Gly1961Glu) report visual symptoms at a mean age of 19.7 and 22.7 years, respectively, while the disease is significantly delayed (to 36.3 years, p<0.0001) in patients with the c.5603A>T (p.Asn1868Ile) allele. (B) The prevalence of foveal sparing is highest among patients with p.Asn1868Ile at 84.7% while observed only in ~33% of cases with other ABCA4 variants including p.Gly1961Glu. (C) Autofluorescence imaging across the macula in a p.Asn1868Ile patient exhibiting foveal sparing within an area of retinal pigment epithelium (RPE) and photoreceptor cell atrophy. (D) A spectral domain-optical coherence tomographic scan across the fovea of the same patient reveals the presence of outer retinal layers: RPE, ellipsoid zone (EZ) and external limiting membrane (ELM) in the fovea. SD-OCT, spectral domain-optical coherence tomography.

ABCA4 disease phenotype and monoallelic cases. Genetic analysis taking into account the p.Asn1868Ile variant would allow distinguishing true monoallelic carriers affected with AMD from biallelic late-onset ABCA4 disease in most cases under investigation.

DISCUSSION

Since the *ABCA4* gene was cloned 20 years ago,[1] the *ABCA4* locus continues to be the most heterogeneous of Mendelian eye diseases in terms of both genetic and phenotypic expression.[42] While loci with similar genetic heterogeneity are not rare—the related ABC transporter, CFTR, locus presents with more pathogenic variants,[43] the *ABCA4* gene continues to emerge as an example of complexity in Mendelian disorders. Due to a high prevalence of disease-associated alleles, the variation in the *ABCA4* locus underlies a wide spectrum of disease phenotypes, from early onset, rapidly progressing disease resulting in panretinal degeneration to late-onset, milder disease with limited visual deterioration that can phenotypically overlap with the more prevalent AMD. The c.5603A>T (p.Asn1868Ile) allele is, pending unequivocal functional evidence, one of the most frequent alleles (MAF >6.5% in the matched general population) in Mendelian diseases which causes a phenotype. The *ABCA4* locus has several frequent disease-causing alleles in some populations. The best known is the p.Gly1961Glu variant, with the frequency as high as 10% in Somalian population.[44] While several compound heterozygous and even homozygous cases of ABCA4 disease due to the p.Gly1961Glu allele of Somalian ancestry have been described,[32] the penetrance of the variant is unknown in that population since very few cases of Somalian ancestry have been screened due to limited access to patients. However, it is a fully penetrant hypomorphic allele in all other populations where significant numbers of patients have been sequenced. For example, the frequency of the p.Gly1961Glu in South Asian (Indian, etc) populations is 1.5% and we detected it in ~50% of patients (~24% of alleles)

with ABCA4 disease from India.[45] Another recently described frequent variant is the c.6320G>A (p.Arg2107His) allele, with MAF ~2% in African-Americans and MAF >19% in patients of AA descent.[46]

Interestingly, while this study unequivocally demonstrates the non-pathogenicity of the p.Gly863Ala variant alone, that is, without p.Asn1868Ile, it was shown to have an effect in *in vitro* assays where less protein production, reduced ATP binding and hydrolysis[47] and reduced retinal transfer[48] were demonstrated. It has also been shown to have a dual effect: it either results in a deletion of Gly863, due to aberrant splicing, or a substitution of the same glycine to alanine.[28] The p.Asn1868Ile variant, on the contrary, has been showing minor, if any effects in the ATP binding and hydrolysis assays, which were comparable to the wild-type protein.[47] Many reasons could explain these discrepancies, the foremost being the *in vitro* assays likely not adequately recapitulating the *in vivo* conditions. One of known issues with the *in vitro* protein assays are the possible detrimental effects of detergent solubilisation on the stability of certain *ABCA4* variants. Functional studies of knock-in mouse models, which will bypass the issues with *in vitro* assays, are under way to assess the effects of the p.Asn1868Ile, p.Gly863Ala, and the complex allele carrying both variants. Here, the influence of the variants on protein structure and function were assessed by prediction from the atomic level of protein structure as previously described for X linked retinoschisin.[18 19] The hypothetical structure of transmembrane and nucleotide-binding domains of the ABCA4 protein responsible for the flipping of N-retinylidene-PE across photoreceptor disc membranes is shown in figure 3. In this model, both variants, p.Asn1868Ile and p.Gly863Ala, are located on the same side the of the photoreceptor outer segment disc membrane and both were predicted to affect the protein folding by the unfolding mutation screen.[20 21] While not directly experimentally demonstrated, it is predicted and very plausible that both variants forming the complex allele will have a synergistic effect (figure 3).

In fact, the severity of many missense mutations, in addition to those already proven by functional assays or statistical analyses in large patient cohorts (p.[Leu541Pro;Ala1038Val] p.Cys1490Tyr, p.Asn965Ser, p.Cys54Tyr, etc)[24 26 27 47] can be assessed by their presence in *trans* with the p.Asn1868Ile variant, which acts as a 'litmus test' for severity of an *ABCA4* variant. By applying this test and *in silico* analysis, missense mutations determined to be deleterious included c.3261A>C (p.Glu1087Asp), c.3259G>A (p.Glu1087Lys), c.5395A>G (p.Asn1799Asp), c.1921T>C (p.Cys641Arg), c.262G>A (p.Gly88Arg) and c.5923G>C (p.Gly1975Arg).

CONCLUSIONS

First, our study identified the conditions under which the very frequent, previously considered benign c.5603A>T (p.Asn1868Ile) variant is pathogenic. It is an 'extremely' hypomorphic allele and is phenotypically expressed only when in *trans* with a deleterious mutation.

Second, the third most frequent *ABCA4* mutation, the c.2588G>C (p.Gly863Ala) variant, is not disease causing alone, but rather requires the c.5603A>T (p.Asn1868Ile) variant in *cis* for pathogenicity. Therefore, the p.Gly863Ala variant acts not as a disease-associated hypomorphic variant, but rather as a modifier, since it causes disease only when in *cis* with p.Asn1868Ile.

Third, the phenotypes caused by the p.Asn1868Ile variant are late onset, frequently associated with foveal sparing and distinguishable from other ABCA4 disease phenotypes.

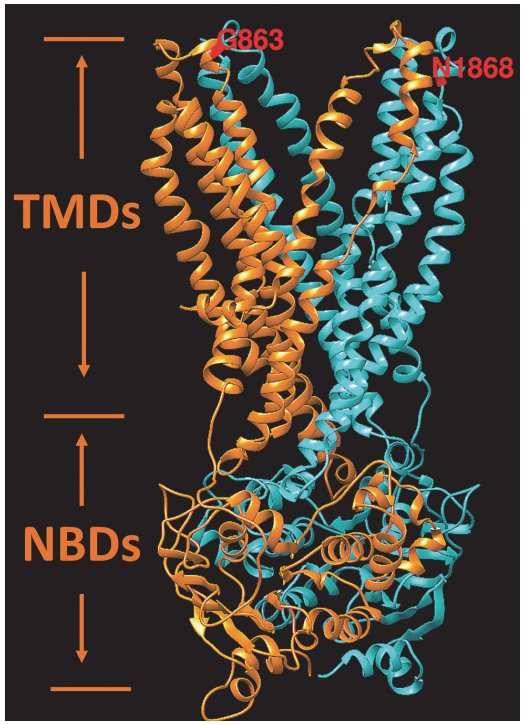


Figure 3 Hypothetical structural model of ABCA4 transmembrane and nucleotide-binding domains. Two symmetrical structural motifs are shown by different colours. Ribbon structure of motif 1, which include transmembrane and nucleotide-binding domains localised in the first half of the protein, is shown in orange. Motif 2 is localised in the second half of the ABCA4 protein and shown in cyan. Transmembrane and nucleotide-binding domains are labelled according to Dawson and Locher[49] as the TMDs and NBDs, respectively. Residues G863 and N1868 are exposed to disk lumen. Energetic (ΔΔG, kcal/mol) and protein folding effects of structural perturbations p.Asn1868Ile and p.Gly863Ala were predicted from the model using the unfolding mutation screen.[20 21]

Fourth, the cases compound heterozygous for p.Asn1868Ile account for at least 10% of all known ABCA4 disease and refine another 10% where they form a complex allele with the p.Gly863Ala variant.

Fifth, the p.Asn1868Ile alleles account for >50% of previously considered 'one mutation' cases of ABCA4 disease and for a very large fraction (~80%) of late-onset cases. While it had been suggested before that many late-onset cases are monoallelic, these are in fact biallelic and harbour the extremely hypomorphic p.Asn1868Ile variant. Therefore, the p.Asn1868Ile variant explains most of the late-onset biallelic ABCA4 disease sometimes masquerading as AMD.

In summary, these data suggest a substantial revision of genetic screening and assessment of pathogenic variants in ABCA4 disease including the absolute necessity of determining of the phase of all alleles. It also serves as an example for all other Mendelian diseases, where careful consideration should be given to seemingly benign and (very) frequent variants, and to specific combinations of these.

Contributors All the authors contributed significantly to this research. Study conceptualisation and design: RA. Sequencing: JZ. Genetic analysis and

interpretation: RA and JZ. Acquisition of clinical data and specimens: WL and FTC. Clinical analysis and interpretation: WL, FTC, SHT, GF and KS. Drafting of manuscript: RA, JZ and WL. Critical revisions: All authors. All authors agree to be accountable for all aspects of the work.

Funding This work was supported, in part, by grants from the National Eye Institute/NIH EY021163, EY019861 and EY019007 (Core Support for Vision Research); Pangere Family Foundation, Pangere Center, Chicago Lighthouse, OPOS Stiftung zugunsten Wahrnehmungsbehinderten, St. Gallen, Switzerland and Alfred-Vogt-Stifung, St. Gallen, Switzerland and unrestricted funds from Research to Prevent Blindness (New York, NY) to the Department of Ophthalmology, Columbia University.

Competing interests None declared.

Patient consent All data used and presented in the study cannot be traced or identified to an individual subject. All identifiable information were removed in accordance with the Health Insurance Portability and Accountability Act of 1996 regulations and Columbia University Medical Center Institutional Review Board protocols.

Ethics approval Columbia University Medical Center Institutional Review Board.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Allikmets R, Singh N, Sun H, Shroyer NF, Hutchinson A, Chidambaram A, Gerrard B, Baird L, Stauffer D, Peiffer A, Rattner A, Smallwood P, Li Y, Anderson KL, Lewis RA, Nathans J, Leppert M, Dean M, Lupski JR. A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. *Nat Genet* 1997;15:236–46.
2 Cremers FP, van de Pol DJ, van Driel M, den Hollander AI, van Haren FJ, Knoers NV, Tijmes N, Bergen AA, Rohrschneider K, Blankenagel A, Pinckers AJ, Deutman AF, Hoyng CB. Autosomal recessive retinitis pigmentosa and cone-rod dystrophy caused by splice site mutations in the Stargardt's disease gene ABCR. *Hum Mol Genet* 1998;7:355–62.
3 Martínez-Mir A, Paloma E, Allikmets R, Ayuso C, del Rio T, Dean M, Vilageliu L, Gonzàlez-Duarte R, Balcells S. Retinitis Pigmentosa caused by a homozygous mutation in the Stargardt disease gene ABCR. *Nat Genet* 1998;18:11–12.
4 Maugeri A, Klevering BJ, Rohrschneider K, Blankenagel A, Brunner HG, Deutman AF, Hoyng CB, Cremers FP. Mutations in the ABCA4 (ABCR) gene are the major cause of autosomal recessive cone-rod dystrophy. *Am J Hum Genet* 2000;67:960–6.
5 Shroyer NF, Lewis RA, Yatsenko AN, Lupski JR. Null missense ABCR (ABCA4) mutations in a family with stargardt disease and retinitis pigmentosa. *Invest Ophthalmol Vis Sci* 2001;42:2757–61.
6 Allikmets R. Stargardt disease: from gene discovery to therapy. In: Tombran-Tink J, Barnstable CJ, eds. *Retinal degenerations: biology, diagnostics and therapeutics*. Totowa, NJ: Humana Press, 2007:105–18.
7 Zernant J, Schubert C, Im KM, Burke T, Brown CM, Fishman GA, Tsang SH, Gouras P, Dean M, Allikmets R. Analysis of the ABCA4 gene by next-generation sequencing. *Invest Ophthalmol Vis Sci* 2011;52:8479–87.
8 Zernant J, Xie YA, Ayuso C, Riveiro-Alvarez R, Lopez-Martinez MA, Simonelli F, Testa F, Gorin MB, Strom SP, Bertelsen M, Rosenberg T, Boone PM, Yuan B, Ayyagari R, Nagy PL, Tsang SH, Gouras P, Collison FT, Lupski JR, Fishman GA, Allikmets R. Analysis of the ABCA4 genomic locus in Stargardt disease. *Hum Mol Genet* 2014;23:6797–806.
9 Yatsenko AN, Shroyer NF, Lewis RA, Lupski JR. Late-onset Stargardt disease is associated with missense mutations that map outside known functional regions of ABCR (ABCA4). *Hum Genet* 2001;108:346–55.
10 Riveiro-Álvarez R, Xie YA, López-Martínez MÁ, Gambin T, Pérez-Carro R, Ávila-Fernández A, López-Molina MI, Zernant J, Jhangiani S, Muzny D, Yuan B, Boerwinkle E, Gibbs R, Lupski JR, Ayuso C, Allikmets R. New mutations in the RAB28 gene in 2 Spanish families with cone-rod dystrophy. *JAMA Ophthalmol* 2015;133:133–9.
11 Tsang SH, Burke T, Oll M, Yzer S, Lee W, Xie YA, Allikmets R. Whole exome sequencing identifies CRB1 defect in an unusual maculopathy phenotype. *Ophthalmology* 2014;121:1773–82.
12 Xie YA, Lee W, Cai C, Gambin T, Nõupuu K, Sujirakul T, Ayuso C, Jhangiani S, Muzny D, Boerwinkle E, Gibbs R, Greenstein VC, Lupski JR, Tsang SH, Allikmets R. New syndrome with retinitis pigmentosa is caused by nonsense mutations in retinol dehydrogenase RDH11. *Hum Mol Genet* 2014;23:5774–80.
13 Yamamoto S, Jaiswal M, Charng WL, Gambin T, Karaca E, Mirzaa G, Wiszniewski W, Sandoval H, Haelterman NA, Xiong B, Zhang K, Bayat V, David G, Li T, Chen K, Gala U, Harel T, Pehlivan D, Penney S, Vissers LE, de Ligt J, Jhangiani SN, Xie Y, Tsang SH, Parman Y, Sivaci M, Battaloglu E, Muzny D, Wan YW, Liu Z, Lin-Moore AT, Clark RD, Curry CJ, Link N, Schulze KL, Boerwinkle E, Dobyns WB, Allikmets R, Gibbs RA, Chen R, Lupski JR, Wangler MF, Bellen HJ. A Drosophila genetic resource of mutants to study mechanisms underlying human genetic diseases. *Cell* 2014;159:200–14.
14 Sangermano R, Bax NM, Bauwens M, van den Born LI, De Baere E, Garanto A, Collin RW, Goercharn-Ramlal AS, den Engelsman-van Dijk AH, Rohrschneider K, Hoyng CB, Cremers FP, Albert S. Photoreceptor progenitor mRNA analysis reveals Exon skipping resulting from the ABCA4 c.5461-10T→C mutation in stargardt disease. *Ophthalmology* 2016;123:1375–85.
15 Lee W, Xie Y, Zernant J, Yuan B, Bearelly S, Tsang SH, Lupski JR, Allikmets R. Complex inheritance of ABCA4 disease: four mutations in a family with multiple macular phenotypes. *Hum Genet* 2016;135:9–19.
16 Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation C; Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285–91.
17 Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR; 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015;526:68–74.
18 Sergeev YV, Caruso RC, Meltzer MR, Smaoui N, MacDonald IM, Sieving PA. Molecular modeling of retinoschisin with functional analysis of pathogenic mutations from human X-linked retinoschisis. *Hum Mol Genet* 2010;19:1302–13.
19 Sergeev YV, Vitale S, Sieving PA, Vincent A, Robson AG, Moore AT, Webster AR, Holder GE. Molecular modeling indicates distinct classes of missense variants with mild and severe XLRS phenotypes. *Hum Mol Genet* 2013;22:4756–67.
20 McCafferty CL, Sergeev YV. In silico mapping of protein unfolding mutations for inherited disease. *Sci Rep* 2016;6:37298.
21 McCafferty CL, Sergeev YV. Dataset of eye disease-related proteins analyzed using the unfolding mutation screen. *Sci Data* 2016;3:160112.
22 Maugeri A, Flothmann K, Hemmrich N, Ingvast S, Jorge P, Paloma E, Patel R, Rozet JM, Tammur J, Testa F, Balcells S, Bird AC, Brunner HG, Hoyng CB, Metspalu A, Simonelli F, Allikmets R, Bhattacharya SS, D'Urso M, Gonzàlez-Duarte R, Kaplan J, te Meerman GJ, Santos R, Schwartz M, Van Camp G, Wadelius C, Weber BH, Cremers FP. The ABCA4 2588G>C Stargardt mutation: single origin and increasing frequency from South-West to North-East europe. *Eur J Hum Genet* 2002;10:197–203.
23 Aguirre-Lamban J, González-Aguilera JJ, Riveiro-Alvarez R, Cantalapiedra D, Avila-Fernandez A, Villaverde-Montero C, Corton M, Blanco-Kelly F, Garcia-Sandoval B, Ayuso C. 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* 2011;52:6206–12.
24 Bertelsen M, Zernant J, Larsen M, Duno M, Allikmets R, Rosenberg T. Generalized choriocapillaris dystrophy, a distinct phenotype in the spectrum of ABCA4-associated retinopathies. *Invest Ophthalmol Vis Sci* 2014;55:2766–76.
25 Rivera A, White K, Stöhr H, Steiner K, Hemmrich N, Grimm T, Jurklies B, Lorenz B, Scholl HP, Apfelstedt-Sylla E, Weber BH. A comprehensive survey of sequence variation in the ABCA4 (ABCR) gene in Stargardt disease and age-related macular degeneration. *Am J Hum Genet* 2000;67:800–13.
26 Zhang N, Tsybovsky Y, Kolesnikov AV, Rozanowska M, Swider M, Schwartz SB, Stone EM, Palczewska G, Maeda A, Kefalov VJ, Jacobson SG, Cideciyan AV, Palczewski K. Protein misfolding and the pathogenesis of ABCA4-associated retinal degenerations. *Hum Mol Genet* 2015;24:3220–37.
27 Cideciyan AV, Swider M, Aleman TS, Tsybovsky Y, Schwartz SB, Windsor EA, Roman AJ, Sumaroka A, Steinberg JD, Jacobson SG, Stone EM, Palczewski K. ABCA4 disease progression and a proposed strategy for gene therapy. *Hum Mol Genet* 2009;18:931–41.
28 Maugeri A, van Driel MA, van de Pol DJ, Klevering BJ, van Haren FJ, Tijmes N, Bergen AA, Rohrschneider K, Blankenagel A, Pinckers AJ, Dahl N, Brunner HG, Deutman AF, Hoyng CB, Cremers FP. 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;64:1024–35.
29 Gold B, Merriam JE, Zernant J, Hancox LS, Taiber AJ, Gehrs K, Cramer K, Neel J, Bergeron J, Barile GR, Smith RT, Hageman GS, Dean M, Allikmets R; AMD Genetics Clinical Study Group. Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration. *Nat Genet* 2006;38:458–62.
30 Hageman GS, Anderson DH, Johnson LV, Hancox LS, Taiber AJ, Hardisty LI, Hageman JL, Stockman HA, Borchardt JD, Gehrs KM, Smith RJ, Silvestri G, Russell SR, Klaver CC, Barbazetto I, Chang S, Yannuzzi LA, Barile GR, Merriam JC, Smith RT, Olsh AK, Bergeron J, Zernant J, Merriam JE, Gold B, Dean M, Allikmets R. A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc Natl Acad Sci U S A* 2005;102:7227–32.

31 Fritsche LG, Igl W, Bailey JN, Grassmann F, Sengupta S, Bragg-Gresham JL, Burdon KP, Hebbring SJ, Wen C, Gorski M, Kim IK, Cho D, Zack D, Souied E, Scholl HP, Bala E, Lee KE, Hunter DJ, Sardell RJ, Mitchell P, Merriam JE, Cipriani V, Hoffman JD, Schick T, Lechanteur YT, Guymer RH, Johnson MP, Jiang Y, Stanton CM, Buitendijk GH, Zhan X, Kwong AM, Boleda A, Brooks M, Gieser L, Ratnapriya R, Branham KE, Foerster JR, Heckenlively JR, Othman MI, Vote BJ, Liang HH, Souzeau E, McAllister IL, Isaacs T, Hall J, Lake S, Mackey DA, Constable IJ, Craig JE, Kitchner TE, Yang Z, Su Z, Luo H, Chen D, Ouyang H, Flagg K, Lin D, Mao G, Ferreyra H, Stark K, von Strachwitz CN, Wolf A, Brandl C, Rudolph G, Olden M, Morrison MA, Morgan DJ, Schu M, Ahn J, Silvestri G, Tsironi EE, Park KH, Farrer LA, Orlin A, Brucker A, Li M, Curcio CA, Mohand-Saïd S, Sahel JA, Audo I, Benchaboune M, Cree AJ, Rennie CA, Goverdhan SV, Grunin M, Hagbi-Levi S, Campochiaro P, Katsanis N, Holz FG, Blond F, Blanché H, Deleuze JF, Igo RP, Truitt B, Peachey NS, Meuer SM, Myers CE, Moore EL, Klein R, Hauser MA, Postel EA, Courtenay MD, Schwartz SG, Kovach JL, Scott WK, Liew G, Tan AG, Gopinath B, Merriam JC, Smith RT, Khan JC, Shahid H, Moore AT, McGrath JA, Laux R, Brantley MA, Agarwal A, Ersoy L, Caramoy A, Langmann T, Saksens NT, de Jong EK, Hoyng CB, Cain MS, Richardson AJ, Martin TM, Blangero J, Weeks DE, Dhillon B, van Duijn CM, Doheny KF, Romm J, Klaver CC, Hayward C, Gorin MB, Klein ML, Baird PN, den Hollander AI, Fauser S, Yates JR, Allikmets R, Wang JJ, Schaumberg DA, Klein BE, Hagstrom SA, Chowers I, Lotery AJ, Léveillard T, Zhang K, Brilliant MH, Hewitt AW, Swaroop A, Chew EY, Pericak-Vance MA, DeAngelis M, Stambolian D, Haines JL, Iyengar SK, Weber BH, Abecasis GR, Heid IM. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat Genet* 2016;48:134–43.

32 Burke TR, Fishman GA, Zernant J, Schubert C, Tsang SH, Smith RT, Ayyagari R, Koenekoop RK, Umfress A, Ciccarelli ML, Baldi A, Iannaccone A, Cremers FP, Klaver CC, Allikmets R. Retinal phenotypes in patients homozygous for the G1961E mutation in the ABCA4 gene. *Invest Ophthalmol Vis Sci* 2012;53:4458–67.

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

34 Allikmets R, Shroyer NF, Singh N, Seddon JM, Lewis RA, Bernstein PS, Peiffer A, Zabriskie NA, Li Y, Hutchinson A, Dean M, Lupski JR, Leppert M. Mutation of the Stargardt disease gene (ABCR) in age-related macular degeneration. *Science* 1997;277:1805–7.

35 Allikmets R. Further evidence for an association of ABCR alleles with age-related macular degeneration. the international ABCR screening consortium. *Am J Hum Genet* 2000;67:487–91.

36 Stone EM, Webster AR, Vandenburgh K, Streb LM, Hockey RR, Lotery AJ, Sheffield VC. Allelic variation in ABCR associated with Stargardt disease but not age-related macular degeneration. *Nat Genet* 1998;20:328–9.

37 Fritsche LG, Fleckenstein M, Fiebig BS, Schmitz-Valckenberg S, Bindewald-Wittich A, Keilhauer CN, Renner AB, Mackensen F, Mößner A, Pauleikhoff D, Adrion C, Mansmann U, Scholl HP, Holz FG, Weber BH. A subgroup of age-related macular degeneration is associated with mono-allelic sequence variants in the ABCA4 gene. *Invest Ophthalmol Vis Sci* 2012;53:2112–8.

38 Westeneng-van Haaften SC, Boon CJ, Cremers FP, Hoefsloot LH, den Hollander AI, Hoyng CB. Clinical and genetic characteristics of late-onset Stargardt's disease. *Ophthalmology* 2012;119:1199–210.

39 Duncker T, Stein GE, Lee W, Tsang SH, Zernant J, Bearelly S, Hood DC, Greenstein VC, Delori FC, Allikmets R, Sparrow JR. Quantitative fundus autofluorescence and optical coherence tomography in ABCA4 carriers. *Invest Ophthalmol Vis Sci* 2015;56:7274–85.

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

41 Kjellström U. Reduced macular function in ABCA4 carriers. *Mol Vis* 2015;21:767–82.

42 Allikmets R. Simple and complex ABCR: genetic predisposition to retinal disease. *Am J Hum Genet* 2000;67:793–9.

43 Sosnay PR, Raraigh KS, Gibson RL. Molecular genetics of cystic fibrosis transmembrane conductance regulator: genotype and phenotype. *Pediatr Clin North Am* 2016;63:585–98.

44 Guymer RH, Héon E, Lotery AJ, Munier FL, Schorderet DF, Baird PN, McNeil RJ, Haines H, Sheffield VC, Stone EM. Variation of codons 1961 and 2177 of the stargardt disease gene is not associated with age-related macular degeneration. *Arch Ophthalmol* 2001;119:745–51.

45 Lee W, Schuerch K, Zernant J, Collison FT, Bearelly S, Fishman GA, Tsang SH, Sparrow JR, Allikmets R. Clinical and genetic spectrum of ABCA4-associated disease in patients of south asian descent. *Eur J Hum Genet* 2017.

46 Zernant J, Collison FT, Lee W, Fishman GA, Noupuu K, Yuan B, Cai C, Lupski JR, Yannuzzi LA, Tsang SH, Allikmets R. Genetic and clinical analysis of ABCA4-associated disease in african american patients. *Hum Mutat* 2014;35:1187–94.

47 Sun H, Smallwood PM, Nathans J. Biochemical defects in ABCR protein variants associated with human retinopathies. *Nat Genet* 2000;26:242–6.

48 MoldayLLAR, Molday RS. Localization and functional analysis of ABCA4 variants associated with stargardt disease. *Association for research in vision and ophthalmology* 2016. Seattle, WA.

49 Dawson RJ, Locher KP. Structure of a bacterial multidrug ABC transporter. *Nature* 2006;443:180–5.

Supplemental Table 1. ABCA4 patient cohort with pathogenic ABCA4 changes and the c.5603A>T (p.Asn1868Ile) status. Nucleotide positions and protein translation correspond to CCDS747.1 and NP_000341.2, respectively. Nucleotide numbering uses the A of the ATG translation initiation start site as nucleotide 1.

PATIENT #	segregation of variants performed	Age of Onset	Foveal Sparing	ABCA4 disease alleles	c.5603A>T (p.Asn1868Ile) genotype	ABCA4 locus seq	ABCA4 pathogenic variant	ABCA4 pathogenic variant
3007	x			2	++		c.2588G>C (p.Gly863Ala)	c.5461-10T>C (p.?)
3807	x	51	N	2	++		c.2588G>C (p.Gly863Ala)	c.5461-10T>C (p.?)
3306				2	++		c.2588G>C (p.Gly863Ala)	c.5461-10T>C (p.?)
3465		10	N	2	++		c.2588G>C (p.Gly863Ala)	c.5461-10T>C (p.?)
3334				2	++		c.2588G>C (p.Gly863Ala)	c.5461-10T>C (p.?)
3467		33		2	++		c.6191C>T (p.Ala2064Val)	c.6191C>T (p.Ala2064Val)
4247				2	++		c.4594G>A (p.Asp1532Asn)	c.4594G>A (p.Asp1532Asn)
3356		41	N	2	++		c.2588G>C (p.Gly863Ala)	c.2588G>C (p.Gly863Ala)
3122	x	32	Y	2	++		c.2588G>C (p.Gly863Ala)	c.2588G>C (p.Gly863Ala)
3341		35	Y	2	++		c.2588G>C (p.Gly863Ala)	c.5177C>A (p.Thr1726Asn)
3611	x			2	+/-		c.[1715G>A;2588G>C] (p.[Arg572Gln;Gly863Ala])	c.5898+1G>A (p.?)
3597		21	N	2	+/-		c.[2549A>G;5882G>A] (p.[Tyr850Cys;Gly1961Glu])	c.5461-10T>C (p.?)
3408	x	18	N	2	+/-	x	c.5882G>A (p.Gly1961Glu)	c.2160+584A>G (p.?)
3269		60	Y	2	+/-	x	c.2588G>C (p.Gly863Ala)	c.4539+2028C>T (p.?)
3179	x			2	+/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5018+5G>A (p.?)
3358				2	+/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5461-10T>C (p.?)
3084	x			2	+/-		c.[4222T>C;4918C>T] (p.[Trp1408Arg;Arg1640Trp])	c.5461-10T>C (p.?)
3427		16	N	2	+/-		c.1592A>G (p.Glu531Gly)	c.5461-10T>C (p.?)
3336	x			2	+/-		c.161G>A (p.Cys54Tyr)	c.5461-10T>C (p.?)
3837	x	11	N	2	+/-		c.1648G>A (p.Gly550Arg)	c.5461-10T>C (p.?)
3250				2	+/-		c.3292C>T (p.Arg1098Cys)	c.5461-10T>C (p.?)
3347				2	+/-		c.3322C>T (p.Arg1108Cys)	c.5461-10T>C (p.?)
3546				2	+/-		c.4139C>T (p.Pro1380Leu)	c.5461-10T>C (p.?)
3839	x	9	Y	2	+/-		c.4139C>T (p.Pro1380Leu)	c.5461-10T>C (p.?)
7565				2	+/-		c.4139C>T (p.Pro1380Leu)	c.5461-10T>C (p.?)
3186				2	+/-		c.4139C>T (p.Pro1380Leu)	c.5461-10T>C (p.?)
3303	x			2	+/-		c.5882G>A (p.Gly1961Glu)	c.5461-10T>C (p.?)
3284	x	10	N	2	+/-		c.5461-10T>C (p.?)	c.5714+5G>A (p.?)
3342	x			2	+/-		c.5461-10T>C (p.?)	c.5714+5G>A (p.?)
3004	x			2	+/-		c.4469G>A (p.Cys1490Tyr)	c.5714+5G>A (p.?)
5020	x			2	+/-		c.770T>G (p.Leu257Arg)	c.5714+5G>A (p.?)
3857				2	+/-		c.2588G>C (p.Gly863Ala)	c.5898+1G>A (p.?)
4581		45	N	2	+/-		c.5461-10T>C (p.?)	c.6006-609T>A (p.?)
3063	x			2	+/-		c.[1715G>A;2588G>C] (p.[Arg572Gln;Gly863Ala])	c.[3322C>T;6320G>A] (p.[Arg1108Cys;Arg2107His])
3595	x		Y	2	+/-		c.1805G>A (p.Arg602Gln)	c.4469G>A (p.Cys1490Tyr)

3129			Y	2	+/-		c.3113C>T (p.Ala1038Val)	c.4469G>A (p.Cys1490Tyr)
3003				2	+/-		c.2588G>C (p.Gly863Ala)	c.4506C>A (p.Cys1502*)
3738		25	N	2	+/-		c.3050+5G>A (p.?)	c.4594G>A (p.Asp1532Asn)
3204		18	N	2	+/-		c.4139C>T (p.Pro1380Leu)	c.4594G>A (p.Asp1532Asn)
3077				2	+/-		c.4139C>T (p.Pro1380Leu)	c.4594G>A (p.Asp1532Asn)
Y703				2	+/-		c.454C>T (p.Arg152*)	c.4594G>A (p.Asp1532Asn)
3553				2	+/-		c.2588G>C (p.Gly863Ala)	c.6286G>A (p.Glu2096Lys)
3104				2	+/-		c.5461-10T>C (p.?)	c.5882G>A (p.Gly1961Glu)
Y647				2	+/-		c.5461-10T>C (p.?)	c.5882G>A (p.Gly1961Glu)
3332		23	N	2	+/-		c.5461-10T>C (p.?)	c.5882G>A (p.Gly1961Glu)
3747		17	N	2	+/-		c.5461-10T>C (p.?)	c.5882G>A (p.Gly1961Glu)
3021				2	+/-		c.3322C>T (p.Arg1108Cys)	c.5882G>A (p.Gly1961Glu)
3458		13	N	2	+/-		c.4216C>T (p.His1406Tyr)	c.5882G>A (p.Gly1961Glu)
3946		42	Y	2	+/-	x	c.4216C>T (p.His1406Tyr)	c.5882G>A (p.Gly1961Glu)
3635		17	Y	2	+/-		c.4469G>A (p.Cys1490Tyr)	c.5882G>A (p.Gly1961Glu)
3473	x		Y	2	+/-		c.4469G>A (p.Cys1490Tyr)	c.5882G>A (p.Gly1961Glu)
2020	x	55	Y	2	+/-		c.4594G>A (p.Asp1532Asn)	c.5882G>A (p.Gly1961Glu)
3724		58	Y	2	+/-		c.4594G>A (p.Asp1532Asn)	c.5882G>A (p.Gly1961Glu)
4351	x	12	Y	2	+/-	x	c.5318C>T (p.Ala1773Val)	c.5882G>A (p.Gly1961Glu)
3508		38	Y	2	+/-		c.67-2A>G (p.?)	c.2588G>C (p.Gly863Ala)
3009				2	+/-		c.768G>T (p.?)	c.2588G>C (p.Gly863Ala)
3125				2	+/-		c.1015T>G (p.Trp339Gly)	c.2588G>C (p.Gly863Ala)
3345				2	+/-		c.1574T>C (p.Phe525Ser)	c.2588G>C (p.Gly863Ala)
9026	x	22	N	2	+/-	x	c.1819G>A (p.Gly607Arg)	c.2588G>C (p.Gly863Ala)
3691			N	2	+/-		c.1849C>T (p.Gln617*)	c.2588G>C (p.Gly863Ala)
3787		21	Y	2	+/-		c.1921T>G (p.Cys641Gly)	c.2588G>C (p.Gly863Ala)
3454				2	+/-		c.1964T>G (p.Phe655Cys)	c.2588G>C (p.Gly863Ala)
3264				2	+/-		c.318T>G (p.Tyr106*)	c.2588G>C (p.Gly863Ala)
3366	x	5		2	+/-		c.5572T>A (p.Tyr1858Asn)	c.5822A>C (p.His1941Pro)
3255		18	N	2	+/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5537T>C (p.Ile1846Thr)
3101				2	+/-		c.5461-10T>C (p.?)	c.6079C>T (p.Leu2027Phe)
3298	x	11	N	2	+/-		c.2552G>A (p.Gly851Asp)	c.6079C>T (p.Leu2027Phe)
3155			Y	2	+/-		c.1253T>C (p.Phe418Ser)	c.4326C>A (p.Asn1442Lys)
3800		18	N	2	+/-		c.2588G>C (p.Gly863Ala)	c.3085C>T (p.Gln1029*)
3216	x			2	+/-		c.2588G>C (p.Gly863Ala)	c.4234C>T (p.Gln1412*)
3292			N	2	+/-		c.2894A>G (p.Asn965Ser)	c.5138A>G (p.Gln1713Arg)
5007	x	5	N	2	+/-		c.4537del (p.Gln1513fs)	c.319C>T (p.Arg107*)
4563	x	48	N	2	+/-		c.2588G>C (p.Gly863Ala)	c.3322C>T (p.Arg1108Cys)
3237	x	9	N	2	+/-		c.1253T>C (p.Phe418Ser)	c.3322C>T (p.Arg1108Cys)
3070				2	+/-		c.2588G>C (p.Gly863Ala)	c.4919G>A (p.Arg1640Gln)
3965	x	12	N	2	+/-	x	c.2564G>A (p.Trp855*)	c.5114G>A (p.Arg1705Gln)

3648	x			2	+/-		c.2588G>C (p.Gly863Ala)	c.5778_5779del (p.Arg1927fs)
3865	x		Y	2	+/-		c.4469G>A (p.Cys1490Tyr)	c.6089G>A (p.Arg2030Gln)
4514		7	N	2	+/-		c.[2041C>T;2588G>C] (p.[Arg681*;Gly863Ala])	c.6088C>T (p.Arg2030*)
3981				2	+/-		c.2588G>C (p.Gly863Ala)	c.6088C>T (p.Arg2030*)
3400				2	+/-		c.5461-10T>C (p.?)	c.6119G>A (p.Arg2040Gln)
3902		50	Y	2	+/-		c.1957C>T (p.Arg653Cys)	c.1957C>T (p.Arg653Cys)
9018		45	N	2	+/-		c.2966T>C (p.Val989Ala)	c.3210_3211dup (p.Ser1071fs)
3257				2	+/-		c.2588G>C (p.Gly863Ala)	c.3335C>A (p.Thr1112Asn)
3600		8	N	2	+/-		c.3322C>T (p.Arg1108Cys)	c.4577C>T (p.Thr1526Met)
3188				2	+/-		c.2919-2A>G (p.?)	c.6718A>G (p.Thr2240Ala)
3049	x			2	+/-		c.2588G>C (p.Gly863Ala)	c.2947A>G (p.Thr983Ala)
3074	x	36	Y	2	+/-		c.5461-10T>C (p.?)	c.1843G>T (p.Val615Phe)
3349				2	+/-		c.5461-10T>C (p.?)	c.122G>A (p.Trp41*)
3509	x			2	+/-		c.2588G>C (p.Gly863Ala)	c.[3758C>T;5882G>A] (p.[Thr1253Met;Gly1961Glu])
3477	x			2	+/-		c.2588G>C (p.Gly863Ala)	c.[4222T>C;4918C>T] (p.[Trp1408Arg;Arg1640Trp])
3348	x			2	-/-	x	c.[3758C>T;5882G>A] (p.[Thr1253Met;Gly1961Glu])	c.1938-619A>G (p.?)
3449	x			2	-/-		c.571-1G>T (p.?)	c.2160+1G>C (p.?)
5225	x			2	-/-		c.161G>A (p.Cys54Tyr)	c.2160+1G>C (p.?)
3193				2	-/-		c.3322C>T (p.Arg1108Cys)	c.3522+5del (p.?)
3061				2	-/-		c.3113C>T (p.Ala1038Val)	c.3813+1G>T (p.?)
3326			N	2	-/-		c.2971G>C (p.Gly991Arg)	c.4253+4C>T (p.?)
3241				2	-/-		c.2588G>C (p.Gly863Ala)	c.4253+5G>T (p.?)
4542	x			2	-/-	x	c.2966T>C (p.Val989Ala)	c.4253+5G>T (p.?)
3031				2	-/-		c.3386G>A (p.Arg1129His)	c.4352+1G>A (p.?)
3525	x			2	-/-	x	c.768G>T (p.?)	c.4539+2001G>A (p.?)
3134		5	N	2	-/-	x	c.2564G>A (p.Trp855*)	c.4539+2001G>A (p.?)
3786		14	N	2	-/-	x	c.1022A>G (p.Glu341Gly)	c.4539+2028C>T (p.?)
3513	x			2	-/-	x	c.161G>A (p.Cys54Tyr)	c.4539+2028C>T (p.?)
3291				2	-/-	x	c.4577C>T (p.Thr1526Met)	c.4539+2028C>T (p.?)
3798		6	N	2	-/-	x	c.5113C>T (p.Arg1705Trp)	c.4539+2028C>T (p.?)
3142	x			2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.4540-2A>G (p.?)
3221	x			2	-/-		c.4328G>A (p.Arg1443His)	c.4667+1G>A (p.?)
9004				2	-/-	x	c.161G>A (p.Cys54Tyr)	c.4773+3A>G (p.?)
3399	x			2	-/-		c.5018+2T>C (p.?)	c.5018+2T>C (p.?)
3642		19	N	2	-/-		c.1726G>C (p.Asp576His)	c.5196+1056A>G (p.?)
3556				2	-/-		c.3413T>A (p.Leu1138His)	c.5196+1056A>G (p.?)
4271	x			2	-/-	x	c.5882G>A (p.Gly1961Glu)	c.5196+1056A>G (p.?)
3330				2	-/-	x	c.5882G>A (p.Gly1961Glu)	c.5196+1056A>G (p.?)
3430	x			2	-/-	x	c.161G>A (p.Cys54Tyr)	c.5196+1137G>A (p.?)
3827	x			2	-/-	x	c.161G>A (p.Cys54Tyr)	c.5196+1137G>A (p.?)
3151				2	-/-	x	c.666_678del (p.Lys223_Arg226delfs)	c.5196+1137G>A (p.?)

3788	x	32	Y	2	-/-	x	c.666_678del (p.Lys223_Arg226delfs)	c.5196+1137G>A (p.?)
3097				2	-/-		c.571-2_c.575del (p.?)	c.5313-2_c.5316del (p.?)
3143				2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5460+1G>A (p.?)
3295				2	-/-	x	c.2971G>C (p.Gly991Arg)	c.570+1798A>G (p.?)
Y417				2	-/-	x	c.52C>T (p.Arg18Trp)	c.570+1798A>G (p.?)
3954	x			2	-/-	x	c.571-1G>T (p.?)	c.571-1G>T (p.?)
3192				2	-/-		c.3522+5del (p.?)	c.5714+5G>A (p.?)
9025		42	Y	2	-/-	x	c.5714+5G>A (p.?)	c.5714+5G>A (p.?)
3353	x			2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5714+5G>A (p.?)
3730		13	N	2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5714+5G>A (p.?)
3225				2	-/-		c.1622T>C (p.Leu541Pro)	c.5714+5G>A (p.?)
3570		24	N	2	-/-		c.1648G>A (p.Gly550Arg)	c.5714+5G>A (p.?)
3226				2	-/-	x	c.2092T>C (p.Cys698Arg)	c.5714+5G>A (p.?)
3725				2	-/-		c.2915C>A (p.Thr972Asn)	c.5714+5G>A (p.?)
9001	x	17	Y	2	-/-	x	c.3352C>T (p.His1118Tyr)	c.5714+5G>A (p.?)
3709	x	19	N	2	-/-		c.4139C>T (p.Pro1380Leu)	c.5714+5G>A (p.?)
3859	x	8	N	2	-/-		c.4139C>T (p.Pro1380Leu)	c.5714+5G>A (p.?)
3888	x			2	-/-		c.4139C>T (p.Pro1380Leu)	c.5714+5G>A (p.?)
3071				2	-/-		c.4139C>T (p.Pro1380Leu)	c.5714+5G>A (p.?)
3387		43	Y	2	-/-		c.4139C>T (p.Pro1380Leu)	c.5714+5G>A (p.?)
3059	x			2	-/-		c.106del (p.Leu37fs)	c.5714+5G>A (p.?)
3378				2	-/-	x	c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5836-3C>A (p.?)
3391	x			2	-/-		c.5882G>A (p.Gly1961Glu)	c.6005+1G>T (p.?)
3124		42	Y	2	-/-	x	c.3814-2A>G (p.?)	c.6006-609T>A (p.?)
3023	x			2	-/-	x	c.5312+2T>G (p.?)	c.6006-609T>A (p.?)
9029		34	Y	2	-/-	x	c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.6006-609T>A (p.?)
3782	x			2	-/-	x	c.161G>A (p.Cys54Tyr)	c.6006-609T>A (p.?)
3001				2	-/-	x	c.1988G>A (p.Trp663*)	c.6006-609T>A (p.?)
3638	x			2	-/-		c.5882G>A (p.Gly1961Glu)	c.6148-698_ c.6670del (p.?)
3971	x			2	-/-	x	c.4577C>T (p.Thr1526Met)	c.6386+2C>G (p.?)
3398				2	-/-		c.5882G>A (p.Gly1961Glu)	c.6387-1G>T (p.?)
3809	x	17	N	2	-/-		c.5882G>A (p.Gly1961Glu)	c.6479+1G>A (p.?)
Y416				2	-/-		c.71G>A (p.Arg24His)	c.6479+1G>C (p.?)
3925		31		2	-/-	x	c.5882G>A (p.Gly1961Glu)	c.67-1845_c.160+345del (p.?)
3539	x			2	-/-		c.5882G>A (p.Gly1961Glu)	c.6729+5_6729+19del (p.?)
3928				2	-/-		c.5882G>A (p.Gly1961Glu)	c.6729+5_6729+19del (p.?)
3572				2	-/-		c.859-9T>C (p.?)	c.859-9T>C (p.?)
3106	x			2	-/-		c.859-9T>C (p.?)	c.859-9T>C (p.?)
3392				2	-/-		c.5318C>T (p.Ala1773Val)	c.4773G>C (p.?)
3035				2	-/-	x	c.5882G>A (p.Gly1961Glu)	c.6342G>A (p.?)
3516				2	-/-		c.2894A>G (p.Asn965Ser)	c.[3416A>G;6119G>A] (p.[Tyr1139Cys;Arg2040Gln])

3263				2	-/-		c.194G>A (p.Gly65Glu)	c.3113C>T (p.Ala1038Val)
3384	x			2	-/-		c.1988G>A (p.Trp663*)	c.3113C>T (p.Ala1038Val)
9084	x			2	-/-		c.247_250dup (p.Ser84fs)	c.3113C>T (p.Ala1038Val)
3511	x			2	-/-		c.4457C>T (p.Pro1486Leu)	c.4793C>A (p.Ala1598Asp)
3198	x			2	-/-		c.4793C>A (p.Ala1598Asp)	c.4793C>A (p.Ala1598Asp)
3385	x			2	-/-		c.4793C>A (p.Ala1598Asp)	c.4793C>A (p.Ala1598Asp)
4515	x	12	N	2	-/-		c.3386G>T (p.Arg1129Leu)	c.5318C>T (p.Ala1773Val)
3785		11	N	2	-/-		c.5380G>C (p.Ala1794Pro)	c.5380G>C (p.Ala1794Pro)
3052				2	-/-	x	c.3287C>T (p.Ser1096Leu)	c.6108_6130dup (p.Ala2044ValinsProGlyPheGluValTyrGln)
3483		10	N	2	-/-		c.2453G>A (p.Gly818Glu)	c.4462T>C (p.Cys1488Arg)
3818	x			2	-/-		c.5882G>A (p.Gly1961Glu)	c.6448T>C (p.Cys2150Arg)
3416	x			2	-/-		c.[2560G>A;3113C>T] (p.[Ala854Thr;Ala1038Val])	c.6449G>A (p.Cys2150Tyr)
3498	x			2	-/-		c.2461T>A (p.Trp821Arg)	c.6449G>A (p.Cys2150Tyr)
3561				2	-/-		c.5882G>A (p.Gly1961Glu)	c.6449G>A (p.Cys2150Tyr)
4185		12	N	2	-/-		c.5714+5G>A (p.?)	c.161G>A (p.Cys54Tyr)
3231				2	-/-		c.247_250dup (p.Ser84fs)	c.323A>T (p.Asp108Val)
3075				2	-/-		c.52C>T (p.Arg18Trp)	c.3606T>A (p.Asp1202Glu)
3748	x	3	N	2	-/-		c.834del (p.Asp279fs)	c.834delT (p.Asp279fs)
3732		32	Y	2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.1726G>C (p.Asp576His)
3423		12	N	2	-/-		c.[1622T>C;1933G>A;3113C>T] (p.[Leu541Pro;Asp645Asn;Ala1038Val])	c.1933G>A (p.Asp645Asn)
3821	x			2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.3259G>A (p.Glu1087Lys)
3960	x	17	N	2	-/-	x	c.2401G>A (p.Ala801Thr)	c.3259G>A (p.Glu1087Lys)
9064	x			2	-/-		c.3988G>T (p.Glu1330*)	c.3988G>T (p.Glu1330*)
3113				2	-/-		c.4854G>T (p.Trp1618Cys)	c.6694G>A (p.Glu2232Lys)
4333	x			2	-/-		c.1222C>T (p.Arg408*)	c.4249_4251del (p.Phe1417del)
3211				2	-/-		c.3323G>A (p.Arg1108His)	c.4318T>G (p.Phe1440Val)
3478				2	-/-		c.71G>A (p.Arg24His)	c.751_753del (p.Phe251del)
3220	x			2	-/-		c.101_106del (p.Ser34_Leu35delfs)	c.3547G>T (p.Gly1183Cys)
3578	x			2	-/-		c.3113C>T (p.Ala1038Val)	c.4322G>A (p.Gly1441Asp)
3494	x			2	-/-		c.[1622T>C;4328G>A] (p.[Leu541Pro;Arg1443His])	c.5882G>A (p.Gly1961Glu)
3262	x	32	Y	2	-/-		c.[3758C>T;5882G>A] (p.[Thr1253Met;Gly1961Glu])	c.5882G>A (p.Gly1961Glu)
3586	x			2	-/-		c.[1A>G;6089G>A] (p.[?;Arg2030Gln])	c.5882G>A (p.Gly1961Glu)
3401	x			2	-/-		c.[3608G>A;4537del] (p.[Gly1203Glu;Gln1513fs])	c.5882G>A (p.Gly1961Glu)
3501	x			2	-/-		c.1100-1G>A (p.?)	c.5882G>A (p.Gly1961Glu)
3418		19	N	2	-/-		c.160+2T>C (p.?)	c.5882G>A (p.Gly1961Glu)
3267				2	-/-		c.160+5G>C (p.?)	c.5882G>A (p.Gly1961Glu)
3039				2	-/-		c.2382+1G>A (p.?)	c.5882G>A (p.Gly1961Glu)
3343	x			2	-/-		c.3050+5G>A (p.?)	c.5882G>A (p.Gly1961Glu)
3594	x			2	-/-		c.3050+5G>A (p.?)	c.5882G>A (p.Gly1961Glu)
3708		13		2	-/-		c.3050+5G>A (p.?)	c.5882G>A (p.Gly1961Glu)

4385				2	-/-		c.3050+5G>A (p.?)	c.5882G>A (p.Gly1961Glu)
3309				2	-/-		c.4539+1G>T (p.?)	c.5882G>A (p.Gly1961Glu)
3533				2	-/-		c.4540-2A>G (p.?)	c.5882G>A (p.Gly1961Glu)
3453				2	-/-		c.5018+2T>C (p.?)	c.5882G>A (p.Gly1961Glu)
3260	x			2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5882G>A (p.Gly1961Glu)
3214	x			2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5882G>A (p.Gly1961Glu)
4547				2	-/-	x	c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5882G>A (p.Gly1961Glu)
3248				2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5882G>A (p.Gly1961Glu)
3324				2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5882G>A (p.Gly1961Glu)
3964		11		2	-/-	x	c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5882G>A (p.Gly1961Glu)
3167		10	Y	2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5882G>A (p.Gly1961Glu)
3249				2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5882G>A (p.Gly1961Glu)
3322				2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5882G>A (p.Gly1961Glu)
3373				2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5882G>A (p.Gly1961Glu)
3091		33	Y	2	-/-		c.1208A>T (p.Asp403Val)	c.5882G>A (p.Gly1961Glu)
3066				2	-/-		c.161G>A (p.Cys54Tyr)	c.5882G>A (p.Gly1961Glu)
3289	x			2	-/-		c.1622T>C (p.Leu541Pro)	c.5882G>A (p.Gly1961Glu)
3371				2	-/-		c.1804C>T (p.Arg602Trp)	c.5882G>A (p.Gly1961Glu)
3842				2	-/-	x	c.1819G>T (p.Gly607Trp)	c.5882G>A (p.Gly1961Glu)
3000				2	-/-		c.1908G>T (p.Gln636His)	c.5882G>A (p.Gly1961Glu)
3852				2	-/-		c.1957C>T (p.Arg653Cys)	c.5882G>A (p.Gly1961Glu)
3109				2	-/-		c.1964T>G (p.Phe655Cys)	c.5882G>A (p.Gly1961Glu)
3409	x			2	-/-		c.1988G>A (p.Trp663*)	c.5882G>A (p.Gly1961Glu)
3016	x			2	-/-		c.1A>G (p.?)	c.5882G>A (p.Gly1961Glu)
3175				2	-/-		c.2453G>T (p.Gly818Val)	c.5882G>A (p.Gly1961Glu)
3824	x			2	-/-		c.286A>G (p.Asn96Asp)	c.5882G>A (p.Gly1961Glu)
3792				2	-/-		c.3007C>T (p.Gln1003*)	c.5882G>A (p.Gly1961Glu)
3130		36	Y	2	-/-		c.3322C>T (p.Arg1108Cys)	c.5882G>A (p.Gly1961Glu)
3208				2	-/-		c.3364G>A (p.Glu1122Lys)	c.5882G>A (p.Gly1961Glu)
3493				2	-/-		c.3385C>T (p.Arg1129Cys)	c.5882G>A (p.Gly1961Glu)
3475				2	-/-		c.4070C>T (p.Ala1357Val)	c.5882G>A (p.Gly1961Glu)
3435	x			2	-/-		c.4139C>T (p.Pro1380Leu)	c.5882G>A (p.Gly1961Glu)
3451	x			2	-/-		c.4139C>T (p.Pro1380Leu)	c.5882G>A (p.Gly1961Glu)
3543	x			2	-/-		c.4139C>T (p.Pro1380Leu)	c.5882G>A (p.Gly1961Glu)
3565	x			2	-/-		c.4139C>T (p.Pro1380Leu)	c.5882G>A (p.Gly1961Glu)
7076				2	-/-		c.4139C>T (p.Pro1380Leu)	c.5882G>A (p.Gly1961Glu)
Y391				2	-/-		c.4139C>T (p.Pro1380Leu)	c.5882G>A (p.Gly1961Glu)
4361				2	-/-		c.4139C>T (p.Pro1380Leu)	c.5882G>A (p.Gly1961Glu)
Y395				2	-/-		c.4217A>G (p.His1406Arg)	c.5882G>A (p.Gly1961Glu)
3233				2	-/-		c.4234C>T (p.Gln1412*)	c.5882G>A (p.Gly1961Glu)
3382				2	-/-		c.4234C>T (p.Gln1412*)	c.5882G>A (p.Gly1961Glu)

3784		27	N	2	-/-		c.4462T>C (p.Cys1488Arg)	c.5882G>A (p.Gly1961Glu)
3323				2	-/-		c.454C>T (p.Arg152*)	c.5882G>A (p.Gly1961Glu)
3254	x			2	-/-		c.45G>A (p.Trp15*)	c.5882G>A (p.Gly1961Glu)
3067				2	-/-		c.5087G>A (p.Ser1696Asn)	c.5882G>A (p.Gly1961Glu)
3196				2	-/-		c.5114G>T (p.Arg1705Leu)	c.5882G>A (p.Gly1961Glu)
3062				2	-/-		c.5316G>A (p.Trp1772*)	c.5882G>A (p.Gly1961Glu)
3469				2	-/-	x	c.5318C>T (p.Ala1773Val)	c.5882G>A (p.Gly1961Glu)
6046	x			2	-/-		c.5882G>A (p.Gly1961Glu)	c.5882G>A (p.Gly1961Glu)
3329	x			2	-/-		c.5882G>A (p.Gly1961Glu)	c.5882G>A (p.Gly1961Glu)
3190				2	-/-		c.5882G>A (p.Gly1961Glu)	c.5882G>A (p.Gly1961Glu)
3235				2	-/-		c.5882G>A (p.Gly1961Glu)	c.5882G>A (p.Gly1961Glu)
4085	x			2	-/-		c.634C>T (p.Arg212Cys)	c.5882G>A (p.Gly1961Glu)
3432	x			2	-/-		c.634C>T (p.Arg212Cys)	c.5882G>A (p.Gly1961Glu)
3222				2	-/-		c.564del (p.Glu189fs)	c.5882G>A (p.Gly1961Glu)
3804				2	-/-		c.2191_2199del (p.Phe731_Leu733del)	c.5882G>A (p.Gly1961Glu)
3266				2	-/-		c.4734del (p.Leu1580*)	c.5882G>A (p.Gly1961Glu)
3275		13	Y	2	-/-		c.4537dup (p.Gln1513fs)	c.5882G>A (p.Gly1961Glu)
9028		16	Y	2	-/-		c.5196+1G>A (p.?)	c.5882G>A (p.Gly1961Glu)
4602	x			2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5882G>A (p.Gly1961Glu)
9090	x			2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5882G>A (p.Gly1961Glu)
9056				2	-/-		c.3407G>T (p.Gly1136Val)	c.5882G>A (p.Gly1961Glu)
3993				2	-/-		c.5882G>A (p.Gly1961Glu)	c.5882G>A (p.Gly1961Glu)
3998		49	Y	2	-/-		c.5882G>A (p.Gly1961Glu)	c.5882G>A (p.Gly1961Glu)
9068	x			2	-/-		c.4947del (p.Glu1650fs)	c.5882G>A (p.Gly1961Glu)
3480				2	-/-		c.2947A>G (p.Thr983Ala)	c.5882G>A (p.Gly1961Glu)
3174				2	-/-		c.286A>G (p.Asn96Asp)	c.[5959G>T;c.5961_5964del] (p.Gly1987*)
9019				2	-/-		c.2588G>C (p.Gly863Ala)	c.2897G>A (p.Gly966Glu)
3044				2	-/-		c.2971G>C (p.Gly991Arg)	c.4875T>A (p.His1625Gln)
3904		20	Y	2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5051T>A (p.Ile1684Asn)
4060	x			2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.3543del (p.Lys1182fs)
3688		42	Y	2	-/-		c.2925C>G (p.Ile975Met)	c.5932A>G (p.Lys1978Glu)
3137		36	Y	2	-/-	x	c.2971G>C (p.Gly991Arg)	c.3377T>C (p.Leu1126Pro)
3470	x			2	-/-	x	c.2971G>C (p.Gly991Arg)	c.3413T>C (p.Leu1138Pro)
3015				2	-/-		c.3279C>A (p.Asp1093Glu)	c.3819dup (p.Leu1274fs)
3956		14	Y	2	-/-	x	c.2813T>C (p.Phe938Ser)	c.5186T>C (p.Leu1729Pro)
3502				2	-/-		c.2947A>G (p.Thr983Ala)	c.5186T>C (p.Leu1729Pro)
3968				2	-/-		c.5059del (p.Ile1687fs)	c.5351T>G (p.Leu1784Arg)
3173				2	-/-		c.4139C>T (p.Pro1380Leu)	c.5819T>A (p.Leu1940Gln)
9012		22	Y	2	-/-	x	c.[71G>A;5882G>A] (p.[Arg24His;Gly1961Glu])	c.6079C>T (p.Leu2027Phe)
3313				2	-/-		c.4253+5G>T (p.?)	c.6079C>T (p.Leu2027Phe)
3010	x			2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.6079C>T (p.Leu2027Phe)

3778	x			2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.6079C>T (p.Leu2027Phe)
3466				2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.6079C>T (p.Leu2027Phe)
3604	x			2	-/-		c.2915C>A (p.Thr972Asn)	c.6079C>T (p.Leu2027Phe)
3252	x			2	-/-		c.3323G>A (p.Arg1108His)	c.6079C>T (p.Leu2027Phe)
3209				2	-/-		c.4139C>T (p.Pro1380Leu)	c.6079C>T (p.Leu2027Phe)
3765				2	-/-		c.4457C>T (p.Pro1486Leu)	c.6079C>T (p.Leu2027Phe)
3011				2	-/-		c.4577C>T (p.Thr1526Met)	c.6079C>T (p.Leu2027Phe)
3962		44	Y	2	-/-	x	c.5882G>A (p.Gly1961Glu)	c.6079C>T (p.Leu2027Phe)
3242				2	-/-		c.91T>C (p.Trp31Arg)	c.6079C>T (p.Leu2027Phe)
3817		12	Y	2	-/-		c.5882G>A (p.Gly1961Glu)	c.6098T>A (p.Leu2033His)
3160				2	-/-		c.3292C>T (p.Arg1098Cys)	c.6179T>G (p.Leu2060Arg)
3403	x			2	-/-		c.4437G>A (p.Trp1479*)	c.6419T>A (p.Leu2140Gln)
3164		47	Y	2	-/-		c.6543_6578del (p.Leu2182_Phe2193del)	c.6543_6578del (p.Leu2182_Phe2193del)
3395				2	-/-	x	c.859-9T>C (p.?)	c.885del (p.Leu296fs)
3178				2	-/-		c.1622T>C (p.Leu541Pro)	c.1622T>C (p.Leu541Pro)
3412				2	-/-		c.1804C>T (p.Arg602Trp)	c.5646G>A (p.Met1882Ile)
3411				2	-/-		c.2023G>A (p.Val675Ile)	c.5646G>A (p.Met1882Ile)
3390				2	-/-		c.634C>T (p.Arg212Cys)	c.4326C>A (p.Asn1442Lys)
3686		5	N	2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.2894A>G (p.Asn965Ser)
3811		6	N	2	-/-		c.2461T>A (p.Trp821Arg)	c.2894A>G (p.Asn965Ser)
3364	x			2	-/-		c.2894A>G (p.Asn965Ser)	c.2894A>G (p.Asn965Ser)
3136				2	-/-	x	c.2041C>T (p.Arg681*)	c.294C>G (p.Asn98Lys)
3942		13	N	2	-/-	x	c.[3758C>T;5882G>A] (p.[Thr1253Met;Gly1961Glu])	c.4139C>T (p.Pro1380Leu)
3887				2	-/-		c.[3758C>T;5882G>A] (p.[Thr1253Met;Gly1961Glu])	c.4139C>T (p.Pro1380Leu)
3413				2	-/-		c.[3758C>T;5882G>A] (p.[Thr1253Met;Gly1961Glu])	c.4139C>T (p.Pro1380Leu)
3236				2	-/-		c.[1A>G;6089G>A] (p.[?;Arg2030Gln])	c.4139C>T (p.Pro1380Leu)
5180	x	6	N	2	-/-		c.3050+5G>A (p.?)	c.4139C>T (p.Pro1380Leu)
3419				2	-/-	x	c.859-9T>C (p.?)	c.4139C>T (p.Pro1380Leu)
3548				2	-/-		c.3322C>T (p.Arg1108Cys)	c.4139C>T (p.Pro1380Leu)
3243				2	-/-		c.3322C>T (p.Arg1108Cys)	c.4139C>T (p.Pro1380Leu)
3801				2	-/-		c.3322C>T (p.Arg1108Cys)	c.4139C>T (p.Pro1380Leu)
3858				2	-/-		c.3322C>T (p.Arg1108Cys)	c.4139C>T (p.Pro1380Leu)
3126	x			2	-/-		c.3323G>A (p.Arg1108His)	c.4139C>T (p.Pro1380Leu)
Y423				2	-/-		c.3323G>A (p.Arg1108His)	c.4139C>T (p.Pro1380Leu)
4428		10	N	2	-/-		c.3911T>A (p.Leu1304*)	c.4139C>T (p.Pro1380Leu)
3072	x			2	-/-		c.4139C>T (p.Pro1380Leu)	c.4139C>T (p.Pro1380Leu)
3402				2	-/-		c.4139C>T (p.Pro1380Leu)	c.4139C>T (p.Pro1380Leu)
Y485				2	-/-		c.4139C>T (p.Pro1380Leu)	c.4139C>T (p.Pro1380Leu)
3486				2	-/-		c.4139C>T (p.Pro1380Leu)	c.4139C>T (p.Pro1380Leu)
3953				2	-/-	x	c.4139C>T (p.Pro1380Leu)	c.4139C>T (p.Pro1380Leu)
3006				2	-/-		c.5882G>A (p.Gly1961Glu)	c.4139C>T (p.Pro1380Leu)

3624		5	N	2	-/-		c.768G>T (p.?)	c.4139C>T (p.Pro1380Leu)
3926	x	11	N	2	-/-		c.1025_1038del (p.Asp342_Lys346delfs)	c.4139C>T (p.Pro1380Leu)
3279	x	31	Y	2	-/-		c.2893A>T (p.Asn965Tyr)	c.4457C>T (p.Pro1486Leu)
3805				2	-/-		c.3470T>G (p.Leu1157*)	c.3876G>C (p.Gln1292His)
3154				2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.4234C>T (p.Gln1412*)
4286	x			2	-/-		c.3322C>T (p.Arg1108Cys)	c.4234C>T (p.Gln1412*)
3656	x			2	-/-		c.66G>A (p.?)	c.6658C>T (p.Gln2220*)
3128		32	Y	2	-/-		c.[2627A>C;3322C>T] (p.[Gln876Pro;Arg1108Cys])	c.2627A>C (p.Gln876Pro)
9015				2	-/-		c.1819G>A (p.Gly607Arg)	c.3292C>T (p.Arg1098Cys)
3771		8	N	2	-/-		c.2617T>C (p.Phe873Leu)	c.3292C>T (p.Arg1098Cys)
9060		22	N	2	-/-		c.3322C>T (p.Arg1108Cys)	c.3322C>T (p.Arg1108Cys)
3121				2	-/-		c.2947A>G (p.Thr983Ala)	c.3322C>T (p.Arg1108Cys)
3969	x			2	-/-	x	c.3287C>T (p.Ser1096Leu)	c.3322C>T (p.Arg1108Cys)
3873	x			2	-/-		c.478G>T (p.Glu160*)	c.3322C>T (p.Arg1108Cys)
3234				2	-/-		c.634C>T (p.Arg212Cys)	c.3323G>T (p.Arg1108Leu)
3102				2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.3385C>T (p.Arg1129Cys)
3288				2	-/-		c.2894A>G (p.Asn965Ser)	c.3386G>T (p.Arg1129Leu)
3899	x	9	N	2	-/-		c.1025_1038del (p.Asp342_Lys346delfs)	c.3386G>T (p.Arg1129Leu)
3912	x			2	-/-		c.2813T>C (p.Phe938Ser)	c.3898C>T (p.Arg1300*)
3060	x			2	-/-		c.1804C>T (p.Arg602Trp)	c.3898C>T (p.Arg1300*)
3447				2	-/-		c.4667G>A (p.Arg1556Lys)	c.4667G>A (p.Arg1556Lys)
3081				2	-/-		c.3056C>T (p.Thr1019Met)	c.4919G>A (p.Arg1640Gln)
3036	x			2	-/-		c.3322C>T (p.Arg1108Cys)	c.4919G>A (p.Arg1640Gln)
3365				2	-/-		c.4139C>T (p.Pro1380Leu)	c.4919G>A (p.Arg1640Gln)
3187				2	-/-		c.5196+1G>A (p.?)	c.6089G>A (p.Arg2030Gln)
3183		33	Y	2	-/-		c.1648G>A (p.Gly550Arg)	c.6089G>A (p.Arg2030Gln)
3407	x			2	-/-		c.4139C>T (p.Pro1380Leu)	c.6089G>A (p.Arg2030Gln)
4273	x			2	-/-		c.4639A>T (p.Lys1547*)	c.6089G>A (p.Arg2030Gln)
4405				2	-/-		c.6079C>T (p.L2027Phe)	c.6089G>A (p.Arg2030Gln)
3529		33	Y	2	-/-		c.716G>A (p.Trp239*)	c.6089G>A (p.Arg2030Gln)
3870	x			2	-/-		c.5312+1G>A (p.?)	c.6088C>T (p.Arg2030*)
3978		10	N	2	-/-	x	c.5312+2T>G (p.?)	c.6088C>T (p.Arg2030*)
3650				2	-/-		c.6088C>T (p.Arg2030*)	c.6088C>T (p.Arg2030*)
5010				2	-/-		c.1374del (p.Thr459fs)	c.6088C>T (p.Arg2030*)
9057	x	11	Y	2	-/-		c.4577C>T (p.Thr1526Met)	c.6119G>A (p.Arg2040Gln)
3328		7	N	2	-/-		c.1819G>A (p.Gly607Arg)	c.6119G>A (p.Arg2040Gln)
3670		9	N	2	-/-		c.5882G>A (p.Gly1961Glu)	c.6118C>T (p.Arg2040*)
3338	x			2	-/-		c.688T>A (p.Cys230Ser)	c.6229C>G (p.Arg2077Gly)
3723		10	N	2	-/-		c.3385C>T (p.Arg1129Cys)	c.6229C>T (p.Arg2077Trp)
3259				2	-/-		c.5882G>A (p.Gly1961Glu)	c.6229C>T (p.Arg2077Trp)
3583	x			2	-/-		c.6079C>T (p.L2027Phe)	c.6229C>T (p.Arg2077Trp)

3990	x	16	Y	2	-/-		c.4253+5G>A (p.?)	c.6316C>T (p.Arg2106Cys)
3258				2	-/-		c.1100-6T>A (p.?)	c.6316C>T (p.Arg2106Cys)
3598				2	-/-	x	c.3898C>T (p.Arg1300*)	c.6316C>T (p.Arg2106Cys)
3934				2	-/-		c.4538A>G (p.Gln1513Arg)	c.6316C>T (p.Arg2106Cys)
3812				2	-/-		c.3554_3555CC>GA (p.Ser1185*)	c.6317G>A (p.Arg2106His)
Y265				2	-/-		c.1805G>A (p.Arg602Gln)	c.6317_6323del (p.Arg2107_Met2108delfs)
3986				2	-/-		c.1804C>T (p.Arg602Trp)	c.6320G>A (p.Arg2107His)
3775		21	N	2	-/-		c.302+1G>A (p.?)	c.6320G>A (p.Arg2107His)
3523	x			2	-/-		c.3050+5G>A (p.?)	c.6320G>A (p.Arg2107His)
3261				2	-/-		c.3523-1G>A (p.?)	c.6320G>A (p.Arg2107His)
3517	x			2	-/-		c.4540-2A>G (p.?)	c.6320G>A (p.Arg2107His)
3299				2	-/-		c.768G>T (p.?)	c.6320G>A (p.Arg2107His)
3380				2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.6320G>A (p.Arg2107His)
3559				2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.6320G>A (p.Arg2107His)
3664				2	-/-		c.1015T>G (p.Trp339Gly)	c.6320G>A (p.Arg2107His)
3894	x			2	-/-		c.1577A>C (p.Glu526Ala)	c.6320G>A (p.Arg2107His)
3920				2	-/-		c.1577A>C (p.Glu526Ala)	c.6320G>A (p.Arg2107His)
3337		19	N	2	-/-		c.1609C>T (p.Arg537Cys)	c.6320G>A (p.Arg2107His)
3321		44		2	-/-		c.2966T>C (p.Val989Ala)	c.6320G>A (p.Arg2107His)
4556				2	-/-	x	c.2971G>C (p.Gly991Arg)	c.6320G>A (p.Arg2107His)
3092		25	Y	2	-/-		c.4223G>T (p.Trp1408Leu)	c.6320G>A (p.Arg2107His)
3940	x	7		2	-/-	x	c.428C>T (p.Pro143Leu)	c.6320G>A (p.Arg2107His)
3952		43	Y	2	-/-	x	c.6079C>T (p.L2027Phe)	c.6320G>A (p.Arg2107His)
3191				2	-/-		c.6229C>G (p.Arg2077Gly)	c.6320G>A (p.Arg2107His)
Y415				2	-/-		c.6316C>T (p.Arg2106Cys)	c.6320G>A (p.Arg2107His)
3794		60	Y	2	-/-		c.6320G>A (p.Arg2107His)	c.6320G>A (p.Arg2107His)
3082		50	Y	2	-/-		c.658C>T (p.Arg220Cys)	c.6320G>A (p.Arg2107His)
4360				2	-/-		c.247_250dup (p.Ser84fs)	c.6320G>A (p.Arg2107His)
3760		8	N	2	-/-		c.634C>T (p.Arg212Cys)	c.634C>T (p.Arg212Cys)
3376				2	-/-		c.454C>T (p.Arg152*)	c.6445C>T (p.Arg2149*)
3562	x			2	-/-		c.885del (p.Leu296fs)	c.1609C>T (p.Arg537Cys)
3170				2	-/-		c.885del (p.Leu296fs)	c.1609C>T (p.Arg537Cys)
3397				2	-/-		c.1805G>A (p.Arg602Gln)	c.1805G>A (p.Arg602Gln)
3099				2	-/-		c.4139C>T (p.Pro1380Leu)	c.5087G>A (p.Ser1696Asn)
3580	x			2	-/-		c.5196+1G>A (p.?)	c.6703T>C (p.Ser2235Pro)
3182	x			2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.2385C>G (p.Ser795Arg)
4472	x	7	N	2	-/-		c.[4222T>C;4918C>T] (p.[Trp1408Arg;Arg1640Trp])	c.3056C>T (p.Thr1019Met)
3589				2	-/-		c.3056C>T (p.Thr1019Met)	c.3056C>T (p.Thr1019Met)
3331				2	-/-		c.1037A>C (p.Lys346Thr)	c.3350C>T (p.Thr1117Ile)
3489				2	-/-		c.4036_4037del (p.Thr1346fs)	c.4036_4037del (p.Thr1346fs)
3424	x			2	-/-		c.2564G>A (p.Trp855*)	c.4577C>T (p.Thr1526Met)

3863	x	10	N	2	-/-		c.4577C>T (p.Thr1526Met)	c.4577C>T (p.Thr1526Met)
3240	x			2	-/-		c.1222C>T (p.Arg408*)	c.5935del (p.Thr1979fs)
3351				2	-/-		c.4139C>T (p.Pro1380Leu)	c.5056G>A (p.Val1686Met)
3474	x			2	-/-		c.4537dup (p.Gln1513fs)	c.5056G>A (p.Val1686Met)
3176				2	-/-		c.2966T>C (p.Val989Ala)	c.5289delT (p.Val1764fs)
3945				2	-/-	x	c.2966T>C (p.Val989Ala)	c.5289delT (p.Val1764fs)
3228				2	-/-		c.5714+5G>A (p.?)	c.5917del (p.Val1973*)
3692	x			2	-/-		c.859-9T>C (p.?)	c.5917del (p.Val1973*)
3227				2	-/-		c.4524_4534del (p.Leu1509_Pro1511delfs)	c.5917del (p.Val1973*)
3107	x			2	-/-		c.5714+5G>A (p.?)	c.5917del (p.Val1973*)
3117				2	-/-		c.5882G>A (p.Gly1961Glu)	c.5917del (p.Val1973*)
9081	x			2	-/-		c.2966T>C (p.Val989Ala)	c.5917del (p.Val1973*)
3957	x	11		2	-/-	x	c.5882G>A (p.Gly1961Glu)	c.5917del (p.Val1973*)
9032				2	-/-	x	c.71G>A (p.Arg24His)	c.1561del (p.Val521Serfs)
3002				2	-/-		c.1957C>T (p.Arg653Cys)	c.2966T>C (p.Val989Ala)
3977	x	10	N	2	-/-	x	c.1995C>A (p.Tyr665*)	c.2966T>C (p.Val989Ala)
3510		39	Y	2	-/-		c.2966T>C (p.Val989Ala)	c.2966T>C (p.Val989Ala)
4338				2	-/-		c.52C>T (p.Arg18Trp)	c.2966T>C (p.Val989Ala)
3504				2	-/-		c.1317G>A (p.Trp439*)	c.4347G>T (p.Trp1449Cys)
3446		9	N	2	-/-	x	c.4139C>T (p.Pro1380Leu)	c.4437G>A (p.Trp1479*)
4225		12	N	2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.5316G>A (p.Trp1772*)
3769		8	N	2	-/-		c.91T>C (p.Trp31Arg)	c.91T>C (p.Trp31Arg)
3790	x			2	-/-		c.93G>A (p.Trp31*)	c.93G>A (p.Trp31*)
3687		13	N	2	-/-		c.5018+2T>C (p.?)	c.6107A>G (p.Tyr2036Cys)
9099	x	8	N	2	-/-		c.[179C>T;3830C>T] (p.[Ala60Val;Thr1277Met])	c.[179C>T;3830C>T] (p.[Ala60Val;Thr1277Met])
9040				2	-/-		c.[2549A>G;5882G>A] (p.[Tyr850Cys;Gly1961Glu])	c.[2549A>G;5882G>A] (p.[Tyr850Cys;Gly1961Glu])
3375	x			2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])
4574	x	5	N	2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])
3557				2	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])	c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])
4580		7	N	2	-/-		c.1529T>C (p.Leu510Pro)	c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])
3876	x			2	-/-		c.3352C>T (p.His1118Tyr)	c.[4222T>C;4918C>T] (p.[Trp1408Arg;Arg1640Trp])
3335				2	-/-		c.4139C>T (p.Pro1380Leu)	c.[4222T>C;4918C>T] (p.[Trp1408Arg;Arg1640Trp])
3085	x			2	-/-		c.71G>A (p.Arg24His)	c.[4222T>C;4918C>T] (p.[Trp1408Arg;Arg1640Trp])
3080		30	Y	2	-/-		c.[872C>T;1531C>T] (p.[Pro291Leu;Arg511Cys])	c.[872C>T;1531C>T] (p.[Pro291Leu;Arg511Cys])
3542		43	Y	1	++	x	c.1027_1028del (p.Asn343*)	
3496		35	Y	1	++	x	c.3754G>T (p.Glu1252*)	
9097	x	32	Y	1	++		c.5461-10T>C (p.?)	
9013		35	Y	1	++		c.4469G>A (p.Cys1490Tyr)	
Y522				1	++	x	c.5461-10T>C (p.?)	
3057				1	++		c.5461-10T>C (p.?)	
3087		30	Y	1	++	x	c.5461-10T>C (p.?)	

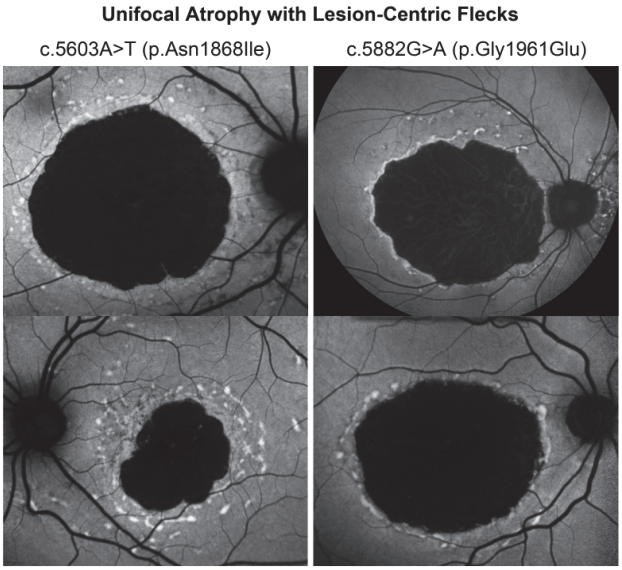
3452		46	Y	1	++	x	c.5461-10T>C (p.?)
3464		24	Y	1	++	x	c.2069del (p.Gly690fs)
3195				1	++	x	c.5461-10T>C (p.?)
3051	x			1	+/-	x	c.4918C>T (p.Arg1640Trp)
3426	x	32	Y	1	+/-	x	c.5395A>G (p.Asn1799Asp)
3094	x			1	+/-	x	c.2588G>C (p.Gly863Ala)
3286	x	42	Y	1	+/-	x	c.4670A>G (p.Tyr1557Cys)
3521		57	Y	1	+/-	x	c.5461-10T>C (p.?)
3064				1	+/-		c.5461-10T>C (p.?)
3584		30	Y	1	+/-	x	c.655A>T (p.Arg219*)
3601	x	22	Y	1	+/-	x	c.[3481C>A;5714+5G>A] (p.[Arg1161Ser;?])
3495	x	10	N	1	+/-	x	c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])
3531		43	N	1	+/-	x	c.4249_4251del (p.Phe1417del)
7203	x	52	Y	1	+/-	x	c.247_250dup (p.Ser84fs)
3100	x		Y	1	+/-	x	c.161G>A (p.Cys54Tyr)
3488		47	Y	1	+/-	x	c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])
3671	x		Y	1	+/-	x	c.571-1G>T (p.?)
3767	x	50	Y	1	+/-	x	c.6229C>T (p.Arg2077Trp)
3892	x	20	Y	1	+/-	x	c.1804C>T (p.Arg602Trp)
3938	x		Y	1	+/-	x	c.5196+1G>A (p.?)
3381	x	45	Y	1	+/-	x	c.2894A>G (p.Asn965Ser)
3405	x	30	Y	1	+/-	x	c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])
9007			Y	1	+/-	x	c.4234C>T (p.Gln1412*)
9011		45	Y	1	+/-	x	c.1957C>T (p.Arg653Cys)
9045				1	+/-		c.5461-10T>C (p.?)
9047				1	+/-		c.6386+1G>A (p.?)
9051			Y	1	+/-		c.52C>T (p.Arg18Trp)
9095		26	N	1	+/-		c.262G>A (p.Gly88Arg)
4546		10		1	+/-	x	c.2588G>C (p.Gly863Ala)
3422				1	+/-	x	c.3328+1G>A (p.?)
4512			Y	1	+/-		c.4234C>T (p.Gln1412*)
3357		48	Y	1	+/-	x	c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])
3293				1	+/-	x	c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])
3028		13	N	1	+/-		c.2894A>G (p.Asn965Ser)
3704		50	Y	1	+/-	x	c.4139C>T (p.Pro1380Leu)
3766		31	Y	1	+/-	x	c.1921T>C (p.Cys641Arg)
3429		27	Y	1	+/-	x	c.4918C>T (p.Arg1640Trp)
3361		28	Y	1	+/-	x	c.6166A>T (p.Lys2056*)
3127		52	Y	1	+/-	x	c.3846del (p.Gly1283fs)
3046				1	+/-		c.6386+2C>G (p.?)
3527		46		1	+/-	x	c.161G>A (p.Cys54Tyr)

4115		44		1	+/-	x	c.1222C>T (p.Arg408*)
3813		40	Y	1	+/-	x	c.868C>T (p.Arg290Trp)
3110		36		1	+/-	x	c.6543_6578del (p.Leu2182_Phe2193del)
4594		35	Y	1	+/-		c.4139C>T (p.Pro1380Leu)
3245		22	N	1	+/-	x	c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])
3862		71	Y	1	+/-	x	c.1903C>T (p.Gln635*)
3803		67	Y	1	+/-	x	c.3261A>C (p.Glu1087Asp)
3098		52	Y	1	+/-	x	c.161-1G>A (p.?)
3485		48	Y	1	+/-	x	c.5923G>C (p.Gly1975Arg)
3456		40	Y	1	+/-	x	c.3259G>A (p.Glu1087Lys)
3089		40	Y	1	+/-	x	c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])
3903		32	Y	1	+/-	x	c.45G>A (p.Trp15*)
3123		31	Y	1	+/-	x	c.3812A>G (p.Glu1271Gly)
3492		29	Y	1	+/-	x	c.6229C>T (p.Arg2077Trp)
3162		21	Y	1	+/-	x	c.1906C>T (p.Gln636*)
3159		18	Y	1	+/-	x	c.5461-10T>C (p.?)
3450				1	+/-	x	c.[4222T>C;4918C>T] (p.[Trp1408Arg;Arg1640Trp])
3045				1	+/-	x	c.247_250dup (p.Ser84fs)
3020				1	+/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])
3008	x			1	-/-		c.3148G>A (p.Gly1050Ser)
3032	x			1	-/-	x	c.[735T>G;2300T>A] (p.[Tyr245*;Val767Asp])
3073	x			1	-/-	x	c.3148G>A (p.Gly1050Ser)
3037	x			1	-/-	x	c.[5882G>A;6445C>T] (p.[Gly1961Glu;Arg2149*])
3468	x			1	-/-	x	c.5714+5G>A (p.?)
3268		42	Y	1	-/-	x	c.4248C>A (p.Phe1416Leu)
3520		59	Y	1	-/-	x	c.5714+5G>A (p.?)
3316				1	-/-	x	c.247_250dup (p.Ser84fs)
3532		57	Y	1	-/-	x	c.3043T>C (p.Phe1015Leu)
3603				1	-/-	x	c.5714+5G>A (p.?)
4367				1	-/-	x	c.2971G>C (p.Gly991Arg)
3609	x			1	-/-	x	c.6320G>A (p.Arg2107His)
3487				1	-/-	x	c.4918C>T (p.Arg1640Trp)
3706		16	N	1	-/-	x	c.2966T>C (p.Val989Ala)
3733	x			1	-/-	x	c.4577C>T (p.Thr1526Met)
3832	x		N	1	-/-	x	c.5882G>A (p.Gly1961Glu)
4565	x		N	1	-/-	x	c.733T>C (p.Tyr245His)
3982	x			1	-/-		c.4793C>A (p.Ala1598Asp)
3999	x			1	-/-		c.5882G>A (p.Gly1961Glu)
3150	x	22	Y	1	-/-	x	c.3322C>T (p.Arg1108Cys)
4638	x			1	-/-		c.5761G>A (p.Val1921Met)
9066	x			1	-/-		c.3292C>T (p.Arg1098Cys)

9096	x			1	-/-		c.5882G>A (p.Gly1961Glu)
9017				1	-/-	x	c.4353-4T>C (p.?)
Y256				1	-/-	x	c.634C>T (p.Arg212Cys)
5360				1	-/-	x	c.5882G>A (p.Gly1961Glu)
4720			N	1	-/-		c.6089G>A (p.Arg2030Gln)
3297				1	-/-	x	c.5882G>A (p.Gly1961Glu)
3943		28	Y	1	-/-	x	c.2616del (p.Leu274fs)
3363				1	-/-	x	c.161G>A (p.Cys54Tyr)
3056				1	-/-		c.3113C>T(p.Ala1038Val)
3118		44	Y	1	-/-	x	c.6320G>A (p.Arg2107His)
3133		41	Y	1	-/-	x	c.5312+1G>A (p.?)
3901		10	Y	1	-/-	x	c.3113C>T(p.Ala1038Val)
5222			N	1	-/-	x	c.2966T>C (p.Val989Ala)
4632			N	1	-/-		c.[1622T>C;3113C>T] (p.[Leu541Pro;Ala1038Val])
4391				1	-/-	x	c.6229C>T (p.Arg2077Trp)
3944				1	-/-	x	c.2041C>T (p.Arg681*)
3579				1	-/-	x	c.6320G>A (p.Arg2107His)
4120				1	-/-	x	c.5881G>A (p.Gly1961Arg)
3278				1	-/-	x	c.6694G>A (p.Glu2232Lys)
4203				1	-/-	x	c.634C>T (p.Arg212Cys)
3898				1	-/-	x	c.1034A>G (p.Tyr345Cys)
3315				1	-/-	x	c.6088C>T (p.Arg2030*)
4541				1	-/-	x	c.1699G>A (p.Val567Met)
3103				1	-/-	x	c.184C>T (p.Pro62Ser)
3327				1	-/-	x	c.[3758C>T;5882G>A] (p.[Thr1253Met;Gly1961Glu])
3740				1	-/-	x	c.4577C>T (p.Thr1526Met)
5056				1	-/-		c.6320G>A (p.Arg2107His)
3018				1	-/-		c.4139C>T (p.Pro1380Leu)
3636		37		0	++		
3212	x			0	+/-		
3119	x	47	Y	0	+/-	x	
3834	x	24	Y	0	+/-		
3939	x	13	N	0	+/-	x	
9093		22	N	0	+/-		
Y382				0	+/-		
Y394				0	+/-		
Y458				0	+/-		
Y533				0	+/-		
3992		10	N	0	+/-		
3314		50	Y	0	+/-		
3280		14	Y	0	+/-		

3120			Y	0	+/-	
3161		44	Y	0	+/-	
3967		3	N	0	+/-	x
3163			N	0	+/-	
3325				0	+/-	
3194				0	-/-	
3224				0	-/-	
3247	x			0	-/-	
3414	x			0	-/-	
3459	x			0	-/-	
3497			Y	0	-/-	
3530	x		Y	0	-/-	
3746	x		Y	0	-/-	
3768	x		Y	0	-/-	
4629	x			0	-/-	
7124				0	-/-	
7728				0	-/-	
9003			N	0	-/-	x
9052			N	0	-/-	
4618			N	0	-/-	
4270				0	-/-	
4184				0	-/-	
4205				0	-/-	
4201				0	-/-	
3294				0	-/-	
3307				0	-/-	
3239				0	-/-	
3795			N	0	-/-	
3796				0	-/-	
3797				0	-/-	
3362				0	-/-	
3189				0	-/-	
3132			N	0	-/-	
3643			N	0	-/-	
3948			N	0	-/-	x
3949			N	0	-/-	x
3088			Y	0	-/-	x
3281			Y	0	-/-	
3659				0	-/-	
3726				0	-/-	
3053				0	-/-	
4172				0	-/-	

3396	0	-/-	x
3029	0	-/-	
3034	0	-/-	
3041	0	-/-	
3069	0	-/-	
3111	0	-/-	
3131	0	-/-	
3135	0	-/-	
3144	0	-/-	
3145	0	-/-	
3157	0	-/-	
3165	0	-/-	
3166	0	-/-	
3205	0	-/-	
3206	0	-/-	
3223	0	-/-	
3251	0	-/-	
3282	0	-/-	
3296	0	-/-	
3311	0	-/-	
3369	0	-/-	
3370	0	-/-	
3372	0	-/-	
3374	0	-/-	
3377	0	-/-	
3383	0	-/-	
3393	0	-/-	
3404	0	-/-	
3448	0	-/-	
3554	0	-/-	
3555	0	-/-	
3558	0	-/-	
3560	0	-/-	
3552	0	-/-	



Supplemental Figure 1. Autofluorescence imaging of the distinct phenotype of a well-delineated, unifocal dark atrophy with bordering lesion-centric flecks observed only in patients harboring c.5882G>A (p.Gly1961Glu) and c.5603A>T (p.Asn1868Ile).

CURRICULUM VITAE

Personal data

Name: Jana Zernant
Date of birth: 29.06.1973
Place of birth: Tartu, Estonia
Citizenship: Estonia

Contact data

E-mail: jz2028@cumc.columbia.edu

Education

2018–2019	Tallinn University of Technology, doctoral studies in gene technology, extern
1996–1998	University of Tartu, Estonia, MSc in microbiology
1992–1993	Ludwig Maximilian University of Munich, Germany, student of biology
1991–1996	University of Tartu, Estonia, BSc in microbiology
1988–1991	Õismäe Classical Gymnasium, Tallinn, Estonia, *cum laude*

Language competence

Estonian	native
English	fluent
German	fluent

Professional employment

2001–current	Columbia University, Department of Ophthalmology, Laboratory of Molecular Genetics, New York, NY, USA, Staff Associate
1999–2001	Asper Biotech Ltd, Tartu, Estonia, Head of Laboratory

Supervised dissertation

Miia Rõõm, BSc, 2006
"Detection of ATM Variants in Patients with Perifoveal Telangiectasia"
Tallinn University of Technology, Faculty of Science, Department of Gene Technology
Co-supervisor

Publications

1. Ma CJ, Lee W, Stong N, **Zernant J**, Chang S, Goldstein D, Nagasaki T, Allikmets R. - 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.* 2019 Jan 10. pii: mcs.a003624. doi: 10.1101/mcs.a003624. [Epub ahead of print]
2. Allikmets R, **Zernant J**, Lee W. - Penetrance of the ABCA4 p.Asn1868Ile Allele in Stargardt Disease. *Invest Ophthalmol Vis Sci.* 2018 Nov 1;59(13):5564-5565.
3. Collison FT, Lee W, Fishman GA, Park JC, **Zernant J**, McAnany JJ, Allikmets R. - CLINICAL CHARACTERIZATION OF STARGARDT DISEASE PATIENTS WITH THE p.N1868I ABCA4 MUTATION. *Retina.* 2018 Sep 7. doi: 10.1097/IAE.0000000000002316.
4. Lee W, **Zernant J**, Nagasaki T, Tsang SH, Allikmets R. - Deep Scleral Exposure: A Degenerative Outcome of End-Stage Stargardt Disease. *Am J Ophthalmol.* 2018 Jul 25. pii: S0002-9394(18)30399-4.
5. **Zernant J**, Lee W, Nagasaki T, Collison FT, Fishman GA, Bertelsen M, Rosenberg T, Gouras P, Tsang SH, Allikmets R. - Extremely hypomorphic and severe deep intronic variants in the ABCA4 locus result in varying Stargardt disease phenotypes. *Cold Spring Harb Mol Case Stud.* 2018 Aug 1;4(4). pii: a002733. doi: 10.1101/mcs.a002733. Print 2018 Aug.
6. Paavo M, Zhao J, Kim HJ, Lee W, **Zernant J**, Cai C, Allikmets R, Tsang SH, Sparrow JR. - Mutations in GPR143/OA1 and ABCA4 Inform Interpretations of Short-Wavelength and Near-Infrared Fundus Autofluorescence. *Invest Ophthalmol Vis Sci.* 2018 May 1;59(6):2459-2469. doi: 10.1167/iovs.18-24213.
7. Albert S, Garanto A, Sangermano R, Khan M, Bax NM, Hoyng CB, **Zernant J**, Lee W, Allikmets R, Collin RWJ, Cremers FPM - Identification and Rescue of Splice Defects Caused by Two Neighboring Deep-Intronic ABCA4 Mutations Underlying Stargardt Disease. *Am J Hum Genet.* 2018 Apr 5;102(4):517-527.
8. Ciccone L, Lee W, **Zernant J**, Tanaka K, Schuerch K, Tsang SH, Allikmets R. - HYPERREFLECTIVE DEPOSITION IN THE BACKGROUND OF ADVANCED STARGARDT DISEASE. *Retina*. 2017 Oct 12. doi: 10.1097/IAE.0000000000001841.
9. Tanaka K, Lee W, **Zernant J**, Schuerch K, Ciccone L, Tsang SH, Sparrow JR, Allikmets R. - The Rapid-Onset Chorioretinopathy Phenotype of ABCA4 Disease. *Ophthalmology.* 2018 Jan;125(1):89-99.
10. **Zernant J**, Lee W, Collison FT, Fishman GA, Sergeev YV, Schuerch K, Sparrow JR, Tsang SH, Allikmets R.- Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration. *J Med Genet.* 2017 Jun;54(6):404-412.
11. Lee W, Schuerch K, **Zernant J**, Collison FT, Bearelly S, Fishman GA, Tsang SH, Sparrow JR, Allikmets R.- Genotypic spectrum and phenotype correlations of ABCA4-associated disease in patients of south Asian descent. *Eur J Hum Genet.* 2017 Jun;25(6):735-743.
12. Scerri TS, Quaglieri A, Cai C, **Zernant J**, Matsunami N, Baird L, Scheppke L, Bonelli R, Yannuzzi LA, Friedlander M; MacTel Project Consortium, Egan CA, Fruttiger M, Leppert M, Allikmets R, Bahlo M. - Genome-wide analyses identify common variants associated with macular telangiectasia type 2. *Nat Genet.* 2017 Apr;49(4):559-567.
13. Cornelis SS, Bax NM, **Zernant J**, Allikmets R, Fritsche LG, den Dunnen JT, Ajmal M, Hoyng CB, Cremers FP.- In Silico Functional Meta-Analysis of 5,962 ABCA4 Variants in 3,928 Retinal Dystrophy Cases. *Hum Mutat.* 2017 Apr;38(4):400-408.

14. Lee W, Schuerch K, Xie Y, **Zernant J**, Tsang SH, Sparrow JR, Allikmets R. - Simultaneous Expression of ABCA4 and GPR143 Mutations: A Complex Phenotypic Manifestation. *Invest Ophthalmol Vis Sci.* 2016 Jun 1;57(7):3409-15.
15. Duncker T, Stein GE, Lee W, Tsang SH, **Zernant J**, Bearelly S, Hood DC, Greenstein VC, Delori FC, Allikmets R, Sparrow JR. - Quantitative Fundus Autofluorescence and Optical Coherence Tomography in ABCA4 Carriers. *Invest Ophthalmol Vis Sci.* 2015 Nov;56(12):7274-85.
16. Lee W, Xie Y, **Zernant J**, Yuan B, Bearelly S, Tsang SH, Lupski JR, Allikmets R. - Complex inheritance of ABCA4 disease: four mutations in a family with multiple macular phenotypes. *Hum Genet.* 2016 Jan;135(1):9-19.
17. Nõupuu K, Lee W, **Zernant J**, Greenstein VC, Tsang S, Allikmets R. - Recessive Stargardt disease phenocopying hydroxychloroquine retinopathy. *Graefes Arch Clin Exp Ophthalmol.* 2016 May;254(5):865-72.
18. Park JC, Collison FT, Fishman GA, Allikmets R, **Zernant J**, Liu M, McAnany JJ. - Objective Analysis of Hyperreflective Outer Retinal Bands Imaged by Optical Coherence Tomography in Patients With Stargardt Disease. *Invest Ophthalmol Vis Sci.* 2015 Jul 1;56(8):4662-7.
19. Sparrow JR, Marsiglia M, Allikmets R, Tsang S, Lee W, Duncker T, **Zernant J**. - Flecks in Recessive Stargardt Disease: Short-Wavelength Autofluorescence, Near-Infrared Autofluorescence, and Optical Coherence Tomography. *Invest Ophthalmol Vis Sci.* 2015 Jul 1;56(8):5029-39.
20. Greenstein VC, Schuman AD, Lee W, Duncker T, **Zernant J**, Allikmets R, Hood DC, Sparrow JR. - Near-infrared autofluorescence: its relationship to short-wavelength autofluorescence and optical coherence tomography in recessive stargardt disease. *Invest Ophthalmol Vis Sci.* 2015 May 1;56(5):3226-34.
21. Duncker T, Tsang SH, Woods RL, Lee W, **Zernant J**, Allikmets R, Delori FC, Sparrow JR. - Quantitative Fundus Autofluorescence and Optical Coherence Tomography in PRPH2/RDS- and ABCA4-Associated Disease Exhibiting Phenotypic Overlap. *Invest Ophthalmol Vis Sci.* 2015 May 1;56(5):3159-70.
22. Marsiglia M, Lee W, Mahajan VB, **Zernant J**, Delori FC, Tsang SH, Sparrow JR. - Quantitative autofluorescence as a clinical tool for expedited differential diagnosis of retinal degeneration. *JAMA Ophthalmol.* 2015 Feb;133(2):219-20.
23. Riveiro-Álvarez R, Xie YA, López-Martínez MÁ, Gambin T, Pérez-Carro R, Ávila-Fernández A, López-Molina MI, **Zernant J**, Jhangiani S, Muzny D, Yuan B, Boerwinkle E, Gibbs R, Lupski JR, Ayuso C, Allikmets R. - New mutations in the RAB28 gene in 2 Spanish families with cone-rod dystrophy. *JAMA Ophthalmol.* 2015 Feb;133(2):133-9.
24. Duncker T, Marsiglia M, Lee W, **Zernant J**, Tsang SH, Allikmets R, Greenstein VC, Sparrow JR. - Correlations among near-infrared and short-wavelength autofluorescence and spectral-domain optical coherence tomography in recessive Stargardt disease. *Invest Ophthalmol Vis Sci.* 2014 Oct 23;55(12):8134-43.
25. Fujinami K, **Zernant J**, Chana RK, Wright GA, Tsunoda K, Ozawa Y, Tsubota K, Robson AG, Holder GE, Allikmets R, Michaelides M, Moore AT. - Clinical and molecular characteristics of childhood-onset Stargardt disease. *Ophthalmology.* 2015 Feb;122(2):326-34.
26. Nõupuu K, Lee W, **Zernant J**, Tsang SH, Allikmets R. - Structural and genetic assessment of the ABCA4-associated optical gap phenotype. *Invest Ophthalmol Vis Sci.* 2014 Oct 9;55(11):7217-26.

27. Duncker T, Tsang SH, Lee W, **Zernant J**, Allikmets R, Delori FC, Sparrow JR. - Quantitative fundus autofluorescence distinguishes ABCA4-associated and non-ABCA4-associated bull's-eye maculopathy. *Ophthalmology.* 2015 Feb;122(2):345-55.
28. Lee W, Noupuu K, Oll M, Duncker T, Burke TR, **Zernant J**, Bearelly S, Tsang SH, Sparrow JR, Allikmets R. - The External Limiting Membrane in Early-Onset Stargardt Disease. *Invest Ophthalmol Vis Sci.* 2014 Aug 19;55(10):6139-49.
29. **Zernant J**, Xie YA, Ayuso C, Riveiro-Alvarez R, Lopez-Martinez MA, Simonelli F, Testa F, Gorin MB, Strom SP, Bertelsen M, Rosenberg T, Boone PM, Yuan B, Ayyagari R, Nagy PL, Tsang SH, Gouras P, Collison FT, Lupski JR, Fishman GA, Allikmets R. - Analysis of the ABCA4 genomic locus in Stargardt disease. *Hum Mol Genet.* 2014 Dec 20;23(25): 6797-806.
30. **Zernant J**, Collison FT, Lee W, Fishman GA, Noupuu K, Yuan B, Cai C, Lupski JR, Yannuzzi LA, Tsang SH, Allikmets R. - Genetic and Clinical Analysis of ABCA4-Associated Disease in African American Patients. *Hum Mutat.* 2014 Oct;35(10):1187-94.
31. Bertelsen M, **Zernant J**, Larsen M, Duno M, Allikmets R, Rosenberg T. - Generalized choriocapillaris dystrophy, a distinct phenotype in the spectrum of ABCA4-associated retinopathies. *Invest Ophthalmol Vis Sci.* 2014 Apr 29;55(4):2766-76.
32. Collison FT, Fishman GA, McAnany JJ, **Zernant J**, Allikmets R. - Psychophysical measurement of rod and cone thresholds in stargardt disease with full-field stimuli. *Retina.* 2014 Sep;34(9):1888-95.
33. Burke TR, Duncker T, Woods RL, Greenberg JP, **Zernant J**, Tsang SH, Smith RT, Allikmets R, Sparrow JR, Delori FC. - Quantitative fundus autofluorescence in recessive Stargardt disease. *Invest Ophthalmol Vis Sci.* 2014 May 1;55(5):2841-52.
34. Chun R, Fishman GA, Collison FT, Stone EM, **Zernant J**, Allikmets R. - The value of retinal imaging with infrared scanning laser ophthalmoscopy in patients with stargardt disease. *Retina.* 2014 Jul;34(7):1391-9.
35. Duncker T, Lee W, Tsang SH, Greenberg JP, **Zernant J**, Allikmets R, Sparrow JR. - Distinct characteristics of inferonasal fundus autofluorescence patterns in stargardt disease and retinitis pigmentosa. *Invest Ophthalmol Vis Sci.* 2013 Oct 17;54(10):6820-6.
36. Fujinami K, Singh R, Carroll J, **Zernant J**, Allikmets R, Michaelides M, Moore AT. - Fine central macular dots associated with childhood-onset Stargardt Disease. *Acta Ophthalmol.* 2014 Mar;92(2):e157-9.
37. Fujinami K, **Zernant J**, Chana RK, Wright GA, Tsunoda K, Ozawa Y, Tsubota K, Webster AR, Moore AT, Allikmets R, Michaelides M. - ABCA4 Gene Screening by Next-Generation Sequencing in a British Cohort. *Invest Ophthalmol Vis Sci.* 2013 Oct 11;54(10):6662-74.
38. Riveiro-Alvarez R, Lopez-Martinez MA, **Zernant J**, Aguirre-Lamban J, Cantalapiedra D, Avila-Fernandez A, Gimenez A, Lopez-Molina MI, Garcia-Sandoval B, Blanco-Kelly F, Corton M, Tatu S, Fernandez-San Jose P, Trujillo-Tiebas MJ, Ramos C, Allikmets R, Ayuso C. - Outcome of ABCA4 Disease-Associated Alleles in Autosomal Recessive Retinal Dystrophies: Retrospective Analysis in 420 Spanish Families. *Ophthalmology*. 2013 Nov;120(11):2332-7.
39. Burke TR, Yzer S, **Zernant J**, Smith RT, Tsang SH, Allikmets R. - Abnormality in the external limiting membrane in early Stargardt Disease. *Ophthalmic Genet.* 2013 Mar-Jun; 34(1-2):75-7.
40. Verdina T, Tsang SH, Greenstein VC, **Zernant J**, Sodi A, Lima LH, Chang S, Allikmets R, Menchini U. - Functional Analysis of Retinal Flecks in Stargardt Disease. *J Clin Exp Ophthalmol.* 2012 Jul 30;3.

41. Burke TR, Fishman GA, **Zernant J**, Schubert C, Tsang SH, Smith RT, Ayyagari R, Koenekoop RK, Umfress A, Ciccarelli ML, Baldi A, Iannaccone A, Cremers FP, Klaver CC, Allikmets R. - Retinal Phenotypes in Patients Homozygous for the G1961E Mutation in the ABCA4 Gene. *Invest Ophthalmol Vis Sci.* 2012 Jul 3;53(8):4458-67.
42. Burke TR, Tsang SH, **Zernant J**, Smith RT, Allikmets R. - Familial discordance in Stargardt disease. *Mol Vis.* 2012;18:227-33.
43. **Zernant J**, Schubert C, Im KM, Burke T, Brown CM, Fishman GA, Tsang SH, Gouras P, Dean M, Allikmets R. - Analysis of the ABCA4 gene by next-generation sequencing. *Invest Ophthalmol Vis Sci*. 2011 Oct 31;52(11):8479-87.
44. Cella W, Greenstein VC, **Zernant-Rajang J**, Smith TR, Barile G, Allikmets R, Tsang SH. - G1961E mutant allele in the Stargardt disease gene ABCA4 causes bull's eye maculopathy. *Exp Eye Res.* 2009 Jun 15;89(1):16-24.
45. Escher P, Gouras P, Roduit R, Tiab L, Bolay S, Delarive T, Chen S, Tsai CC, Hayashi M, **Zernant J**, Merriam JE, Mermod N, Allikmets R, Munier FL, Schorderet DF. - Mutations in NR2E3 can cause dominant or recessive retinal degenerations in the same family. *Hum Mutat.* 2009 Mar;30(3):342-51.
46. Hwang JC, **Zernant J**, Allikmets R, Barile GR, Chang S, Smith RT. –Peripapillary Atrophy in Stargardt Disease. *Retina.* 2009 Feb;29(2):181-6.
47. Tsang SH, Tsui I, Chou CL, **Zernant J**, Haamer E, Iranmanesh R, Tosi J, Allikmets R. - A Novel Mutation and Phenotypes in Phosphodiesterase 6 Deficiency. *Am J Ophthalmol.* 2008 Nov;146(5):780-8.
48. Walia S, Fishman GA, **Zernant-Rajang J**, Raime K, Allikmets R. - Phenotypic expression of a PRPF8 gene mutation in a Large African American family. *Arch Ophthalmol.* 2008 Aug;126(8):1127-32.
49. Kong J, Kim SR, Binley K, Pata I, Doi K, Mannik J, **Zernant-Rajang J**, Kan O, Iqball S, Naylor S, Sparrow JR, Gouras P, Allikmets R. - Correction of the disease phenotype in the mouse model of Stargardt disease by lentiviral gene therapy. *Gene Ther.* 2008 Oct;15(19):1311-20.
50. Despriet DD, Bergen AA, Merriam JE, **Zernant J**, Barile GR, Smith RT, Barbazetto IA, van Soest S, Bakker A, de Jong PT, Allikmets R, Klaver CC.- Comprehensive Analysis of the Candidate Genes CCL2, CCR2, and TLR4 in Age-Related Macular Degeneration. *Invest Ophthalmol Vis Sci.* 2008 Jan;49(1):364-71.
51. Barbazetto IA, Yannuzzi NA, Klais CM, Merriam JE, **Zernant J**, Peiretti E, Yannuzzi LA, Allikmets R. - Pseudo-vitelliform macular detachment and cuticular drusen: exclusion of 6 candidate genes. *Ophthalmic Genet.* 2007 Oct;28(4):192-7.
52. GS Hageman, LS Hancox, AJ Taiber, KM Gehrs, DH Anderson, LV Johnson, MJ Radeke, D Kavanagh, A Richards, J Atkinson, S Meri, J Bergeron, **J Zernant,** J Merriam, B Gold, R Allikmets, M Dean and the AMD clinical study group – Extended haplotypes in the complement factor H (CFH) and CFH-related (CFHR) family of genes protect against age-related macular degeneration: Characterization, ethnic distribution and evolutionary implications. *Annals of Medicine.* 2006; 38: 592-604.
53. Gold B, Merriam JE, **Zernant J**, Hancox LS, Taiber AJ, Gehrs K, Cramer K, Neel J, Bergeron J, Barile GR, Smith RT, Hageman GS, Dean M, Allikmets R, Chang S, Yannuzzi LA, Merriam JC, Barbazetto I, Lerner LE, Russell S, Hoballah J, Hageman J, Stockman H. - Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration. *Nature Genetics* 2006 Apr;38(4):458-62.

54. Hargitai J, **Zernant J**, Somfai GM, Vamos R, Farkas A, Salacz G, Allikmets R.- Correlation of clinical and genetic findings in hungarian patients with stargardt disease. *Invest Ophthalmol Vis Sci.* 2005 Dec;46(12):4402-8.
55. **Zernant J**, Kulm M, Dharmaraj S, den Hollander AI, Perrault I, Preising MN, Lorenz B, Kaplan J, Cremers FP, Maumenee I, Koenekoop RK, Allikmets R. - Genotyping microarray (disease chip) for leber congenital amaurosis: detection of modifier alleles. *Invest Ophthalmol Vis Sci.* 2005 Sep;46(9):3052-9.
56. Simonelli F, Testa F, **Zernant J**, Nesti A, Rossi S, Allikmets R, Rinaldi E. - Genotype-phenotype correlation in Italian families with Stargardt disease. *Ophthalmic Research* 2005 May-Jun;37(3):159-67.
57. Hageman GS, Anderson DH, Johnson LV, Hancox LS, Taiber AJ, Hardisty LI, Hageman JL, Stockman HA, Borchardt JD, Gehrs KM, Smith RJ, Silvestri G, Russell SR, Klaver CC, Barbazetto I, Chang S, Yannuzzi LA, Barile GR, Merriam JC, Smith RT, Olsh AK, Bergeron J, **Zernant J**, Merriam JE, Gold B, Dean M, Allikmets R. - A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proceedings of the National Academy of Sciences of the United States of America* 2005 May 17;102(20):7227-32.
58. Rando Allikmets and **Jana Zernant** – From Gene Chips to Disease Chips – New Approach in Molecular Diagnosis of Eye Diseases. *DNA Microarrays by Taylor and Francis Group.* 2005:83-97.
59. Umeda S, Ayyagari R, Allikmets R, Suzuki MT, Karoukis AJ, Ambasudhan R, **Zernant J**, Okamoto H, Ono F, Terao K, Mizota A, Yoshikawa Y, Tanaka Y, Iwata T. - Early-onset macular degeneration with drusen in a cynomolgus monkey (Macaca fascicularis) pedigree: exclusion of 13 candidate genes and loci. *Invest Ophthalmol Vis Sci.* 2005 Feb;46(2):683-91.
60. Hayashi M, Merriam JE, Klaver CC, **Zernant J**, Bergen AA, Smith RT, Chang S, Merriam JC, Allikmets R. - Evaluation of the ARMD1 locus on 1q25-31 in patients with age-related maculopathy: genetic variation in laminin genes and in exon 104 of HEMICENTIN-1. *Ophthalmic Genetics* 2004 Jun;25(2):111-9.
61. Simonelli F, Testa F, **Zernant J**, Nesti A, Rossi S, Rinaldi E, Allikmets R. - Association of a homozygous nonsense mutation in the ABCA4 (ABCR) gene with cone-rod dystrophy phenotype in an Italian family. *Ophthalmic Research* 2004 Mar-Apr;36(2):82-8.
62. Jaakson K, **Zernant J**, Kulm M, Hutchinson A, Tonisson N, Glavac D, Ravnik-Glavac M, Hawlina M, Meltzer MR, Caruso RC, Testa F, Maugeri A, Hoyng CB, Gouras P, Simonelli F, Lewis RA, Lupski JR, Cremers FP, Allikmets R - Genotyping microarray (gene chip) for the ABCR (ABCA4) gene. *Human Mutation* 2003 Nov;22(5):395-403.
63. Dawson E, Abecasis GR, Bumpstead S, Chen Y, Hunt S, Beare DM, Pabial J, Dibling T, Tinsley E, Kirby S, Carter D, Papaspyridonos M, Livingstone S, Ganske R, Lohmussaar E, **Zernant J** et al - A first-generation linkage disequilibrium map of human chromosome 22. *Nature* 2002 Aug 1; 418(6897):544-8.
64. Tõnisson N, **Zernant J**, Kurg A, Pavel H, Slavin G, Roomere H, Meiel A, Hainaut P and Metspalu A - Evaluating the arrayed primer extension resequencing assay of TP53 tumor suppressor gene. *Proceedings of the National Academy of Sciences of the United States of America* 2002 Apr; Vol. 99(8): 5503-5508.
65. Tover A, **Zernant J**, Chugani SA, Chakrabarty AM, Kivisaar M - Critical nucleotides in the interaction of CatR with the pheBA promoter: conservation of the CatR-mediated regulation mechanisms between the pheBA and catBCA operons. *Microbiology* 2000 Jan;146(Pt 1):173-183.

Presentations in scientific meetings

1. **J. Zernant**, W. Lee, F. Collison, G.A. Fishman, S. Tsang, R. Allikmets - Analysis of *ABCA4* mutation spectrum in Stargardt/ABCA4 disease. *ARVO 2019*
2. Y.V. Sergeev, K. Fujinami, B. Falsini, W.M. Zein, K. Goetz, Y. Yokokawa Fujinami, S. Joon Woo, S. Li, M. Bertelli, W. Lee, **J. Zernant**, R. Allikmets, A. Webster, M. Michaelides, B.P. Brooks, P.A. Sieving - Molecular modeling and global computational mutagenesis of ABCA4 protein show a role of missense changes in Stargardt's disease. *ARVO 2019*
3. R. Fallon, R. Berlow, **J. Zernant**, T. Nagasaki, M. Gantner, S. Harkins-Perry, K. Eade, R. Allikmets, M. Friedlander - MacTel patients carry rare phosphoglycerate dehydrogenase (PHGDH) variants with reduced enzymatic activity. *ARVO 2019*
4. W. Lee, **J. Zernant**, S. Tsang, R. Allikmets - Genotype-Phenotype Correlations in Stargardt/ABCA4 Disease Patients of Advanced Age. *ARVO 2019*
5. R. Fallon, **J. Zernant**, T. Nagasaki, M. Gantner, S. Harkins-Perry, K. Eade, R. Allikmets, M. Friedlander - Rare variants in the phosphoglycerate dehydrogenase (PHGDH) gene in MacTel patients lead to decreased enzymatic activity. *ARVO 2018*
6. W. Lee, M. Paavo, **J. Zernant**, S.H. Tsang, J.R. Sparrow, R. Allikmets - Regions of Disease Sparing Spatially Correlate with Choroidal Thinning in ABCA4 Disease. *ARVO 2018*
7. R. Allikmets, **J. Zernant,** F.T. Collison, G.A. Fishman, Y.V. Sergeev, K Schuerch, J.R. Sparrow, S.H. Tsang, W. Lee - Frequent hypomorphic alleles solve 80% of late-onset ABCA4 disease and distinguish it from age-related macular degeneration. *ARVO 2017*
8. F.P. Cremers, S. Cornelis, N. Bax, **J. Zernant**, R. Allikmets, A. Fritsche, J. den Dunnen, M. Ajmal, C.C. Hoyng - In silico functional meta-analysis of 5,962 ABCA4 variants in 3,928 Stargardt disease and cone-rod dystrophy cases. *ARVO 2017*
9. W. Lee, K. Tanaka, **J. Zernant**, K. Schuerch , L. Ciccone , S. Tsang, J.R. Sparrow, R. Allikmets - The Rapid-Onset Chorioretinopathy (ROC) Phenotype of Stargardt Disease. *ARVO 2017*
10. W. Lee, K. Noupuu, **J. Zernant**, K. Schuerch, P. Gouras, S.H. Tsang, J.R. Sparrow, R. Allikmets - Stargardt Disease Expression on a Background of Low Lipofuscin: The Impact of the p.G1961E Mutation of ABCA4. *ARVO 2016*
11. J.C. Park, F.T. Collison, G.A. Fishman, R. Allikmets, **J. Zernant**, J. McAnany - Objective analysis of outer retinal layers imaged by optical coherence tomography in patients with Stargardt disease. *ARVO 2015*
12. A. Fakin, A.G. Robson, G.E. Holder, K. Fujinami, R. Allikmets, **J. Zernant**, A.T. Moore, M. Michaelides, A. Webster - Phenotype and progression of retinal degeneration in patients nullizygous for ABCA4. *ARVO 2015*
13. **J. Zernant,** C. Ayuso, F. Simonelli, M. Bertelsen, T. Rosenberg, M. Gorin, B. Yuan, P.L. Nagy, R. Allikmets - Analysis of the entire ABCA4 genomic locus in Stargardt disease patients in search for missing mutations. *ARVO 2014*
14. K. Fujinami, **J. Zernant**, Y. Ozawa, K. Tsubota, A.G. Robson, G.E. Holder, A. Webster, R. Allikmets, M. Michaelides, A.T. Moore - Clinical and Genetic Characteristics of Childhood-onset Stargardt Disease. *ARVO 2014*
15. T. Duncker, F.C. Delori, J.P. Greenberg, W. Lee, **J. Zernant**, R. Allikmets, S.H. Tsang, J.R. Sparrow - Quantitative Fundus Autofluorescence in Bull's-Eye Maculopathy. *ARVO 2014*
16. **J. Zernant**, K. Im, T. Burke, P.L. Nagy, M. Pang, Z. Yu, R. Allikmets - Analysis of the *ABCA4* gene by next-generation sequencing. *HUGO Human Genome Meeting 2012*

17. J.E. Merriam, C. Schubert, B. Biziorek, **J. Zernant**, S. Han, K. Kuśmirek, M. Piecyk-Sidor, J. Dolar-Szczasny, Z. Zagórski, R. Allikmets - Analysis of Major Alleles Associated with Age-Related Macular Degeneration in Polish Patients. *ARVO 2009*
18. M.A. Sohrab, G. Barile, L. Xu, L. Perretta, R. Allikmets, J.E. Merriam, **J. Zernant**, B. Mirochnik, R.T. Smith - Association of Homozygous High-Risk Alleles in Complement Factor H and ARMS2 With Drusen Staging of Age-Related Macular Degeneration (AMD). *ARVO 2009*
19. W.P. Cella, V.C. Greenstein, **J. Zernant-Rajang**, R.T. Smith, G.R. Barile, N.-K. Wang, J. Tosi, S. Chang, S.H. Tsang, R. Allikmets - The G1961E Mutant Allele in the Stargardt Disease Gene ABCA4 Causes Bull's Eye Maculopathy. *ARVO 2009*
20. J.C. Hwang, **J. Zernant**, R. Allikmets, G. Barile, S. Chang, R.T. Smith - Peripapillary Atrophy in Stargardt Disease. *ARVO 2008*
21. Tsui, J. Tosi, **J. Zernant**, E. Haamer, C. Chou, I. Corcostegui, S. Tsang, R. Allikmets - Novel Mutations and Phenotypes in Phosphodiesterase 6 Deficiency. *ARVO 2008*
22. J.J. Kang, **J. Zernant**, J. Uksti, J. Merriam, E. Dhrami-Gavazi, L. Kleiman, A. Schrier, S. Trokel, R.E. Braunstein, R. Allikmets - Development of a Genotyping Microarray (Disease Chip) for the Screening of Corneal Dystrophies and Keratoconus. *ARVO 2008*
23. N.L. Parmalee, K. Doi, J. Kong, **J. Zernant**, D.F. Schorderet, R. Allikmets - Controlled Suppression of RPE65 in vitro and in vivo: Implications for Treating Stargardt Disease. *ARVO 2008*
24. R. Allikmets, **J. Zernant**, H. Roomere, I. Lopez, C. Ayuso, S. Banfi, F.P.M. Cremers, R.K. Koenekoop - Analysis of Autosomal Dominant and X-Linked Retinitis Pigmentosa Patients on the ADRP and XlRP Genotyping Microarrays (Disease Chips). *ARVO 2007*
25. J.Hargitai, **J.Zernant**, J.E.Merriam, T.Nagy, R.Allikmets - Analysis of Three Major AMD Loci in the Hungarian Population With Age-Related Macular Degeneration. *ARVO 2007*
26. **J.Zernant**, J.E.Merriam, B.Gold, K.Cramer, J.Neel, G.S.Hageman, M.Dean and R. Allikmets - Common variants in the complement regulatory genes factor H, factor B and the complement component 2 gene predispose to age-related macular degeneration. *HUGO Human Genome Variation 2006*
27. B. Biziorek, J.E. Merriam, **J. Zernant**, J. Mackiewicz, B. Szaja, K. Kuśmirek, Z. Zagórski, R. Allikmets - ABCA4 (ABCR) Gene Mutations in Polish Patients With Stargardt Disease. *ARVO 2006* 47: E-Abstract 1696
28. R. Allikmets, **J. Zernant**, K. Jaakson, I. Lopez, E. Haamer, A. den Hollander, C. Ayuso, S. Banfi, F.P.M. Cremers, R.K. Koenekoop - Analysis of Autosomal Recessive Retinitis Pigmentosa Patients on the ARRP Genotyping Microarray (Disease Chip). *ARVO 2006* 47: E-Abstract 1700
29. **J.Zernant**, K.Jaakson, M.Külm, R.Allikmets. Application of the arrayed primer extension (APEX)–based genotyping microarrays to eye diseases. *HUGO Mutation Detection 2005*
30. J. Hargitai, **J. Zernant**, G.M. Somfai, R. Vámos, Á. Farkas, J. Nemes, G. Salacz, and R. Allikmets - ABCR Mutations and Clinical Phenotype in Hungarian Patients With Stargardt Disease *ARVO 2005* 46: E-Abstract 1638
31. R. Allikmets, **J. Zernant**, I. Perrault, A. den Hollander, S. Dharmaraj, F.P.M. Cremers, J. Kaplan, R.K. Koenekoop, and I. Maumenee - Multiallelic Inheritance and/or modifier alleles in Leber Congenital Amaurosis: Analysis with the LCA disease chip. *ARVO 2004* 45: E-Abstract 2444
32. C.J. MacKay, P. Gouras, M. Hayashi, **J. Zernant**, and R. Allikmets - Enhanced S–cone syndrome in a pedigree including autosomal dominant Retinitis Pigmentosa. *ARVO 2004* 45: E-Abstract 3720

33. J. Hargitai, **J. Zernant**, G.M. Somfai, R. Vamos, A. Farkas, J. Nemes, G. Salacz, and R. Allikmets - Correlation of clinical and genetic findings in Stargardt disease patients from Hungary. *ARVO 2004* 45: E-Abstract 1763
34. **J. Zernant** and R. Allikmets - Advances in genetic diagnosis: current and future approaches for DNA Microarrays in Retinal Dystrophies. *Retina Madrid 2004*
35. R. Allikmets, **J Zernant**, M. Kulm, I. Perrault, A. den Hollander, S. Dharmaraj, R. Koenekoop, J. Kaplan, F. Cremers, I. Maumenee - Genotyping microarray (disease chip) for Leber congenital amaurosis. *ARVO 2003* B690:2851.
36. **J. Zernant**, K. Jaakson, R.A. Lewis, D. Glavac, R. Caruso, P. Gouras, F. Simonelli, J.R. Lupski, F.P.M Cremers, R. Allikmets - Molecular diagnostics of Stargardt disease by genotyping patients on the ABCR (ABCA4) microarray. *ARVO 2003* B766:5107.
37. F. Simonelli, F. Testa, A. Nesti, M. Rinaldi, S. Rossi, **J. Zernant**, R. Allikmets, E. Rinaldi - Clinical expression of cone dystrophy associated with a novel deleterious ABCR (ABCA4) mutation in an Italian family. Ophthalmic Res. (*EVER*) 34: S277, *2002.*
38. R. Allikmets, **J. Zernant**, M. Külm, I. Perrault, A. den Hollander, R. Koenekoop, S. Dharmaraj, F. Cremers, J. Kaplan, I. Maumenee - Genotyping microarray (disease chip) for Leber congenital amaurosis. XV International Congress of Eye Research (*ICER*), Geneva, Switzerland, *2002.*

ELULOOKIRJELDUS

Isikuandmed

Nimi: Jana Zernant
Sünniaeg: 29.06.1973
Sünnikoht: Tartu, Eesti
Kodakondsus: eesti

Kontaktandmed

E-post: jz2028@cumc.columbia.edu

Hariduskäik

2018–2019	Tallinna Tehnikaülikool, geenitehnoloogia doktorantuur, ekstern
1996–1998	Tartu Ülikool, MSc mikrobioloogia erialal
1992–1993	Müncheni Ludwig Maximiliani Ülikool, Saksamaa, bioloogia põhiõpe
1991–1996	Tartu Ülikool, BSc mikrobioloogia erialal
1988–1991	Õismäe Humanitaargümnaasium, Tallinn, *cum laude*

Keelteoskus

eesti	emakeel
inglise	kõrgtase
saksa	kõrgtase

Teenistuskäik

2001–	Columbia Ülikool, oftalmoloogia osakond, molekulaargeneetika labor, New York, NY, USA, teadustöötaja
1999–2001	Asper Biotech Ltd, Tartu, laborijuhataja

Juhendatud lõputöö

Miia Rõõm, BSc, 2006
"*ATM* geeni variatsioonide tuvastamine Perifoveaalse Telangiektaasiaga patsientides" Tallinna Tehnikaülikool, Matemaatika-loodusteaduskond, geenitehnoloogia instituut. Kaasjuhendaja