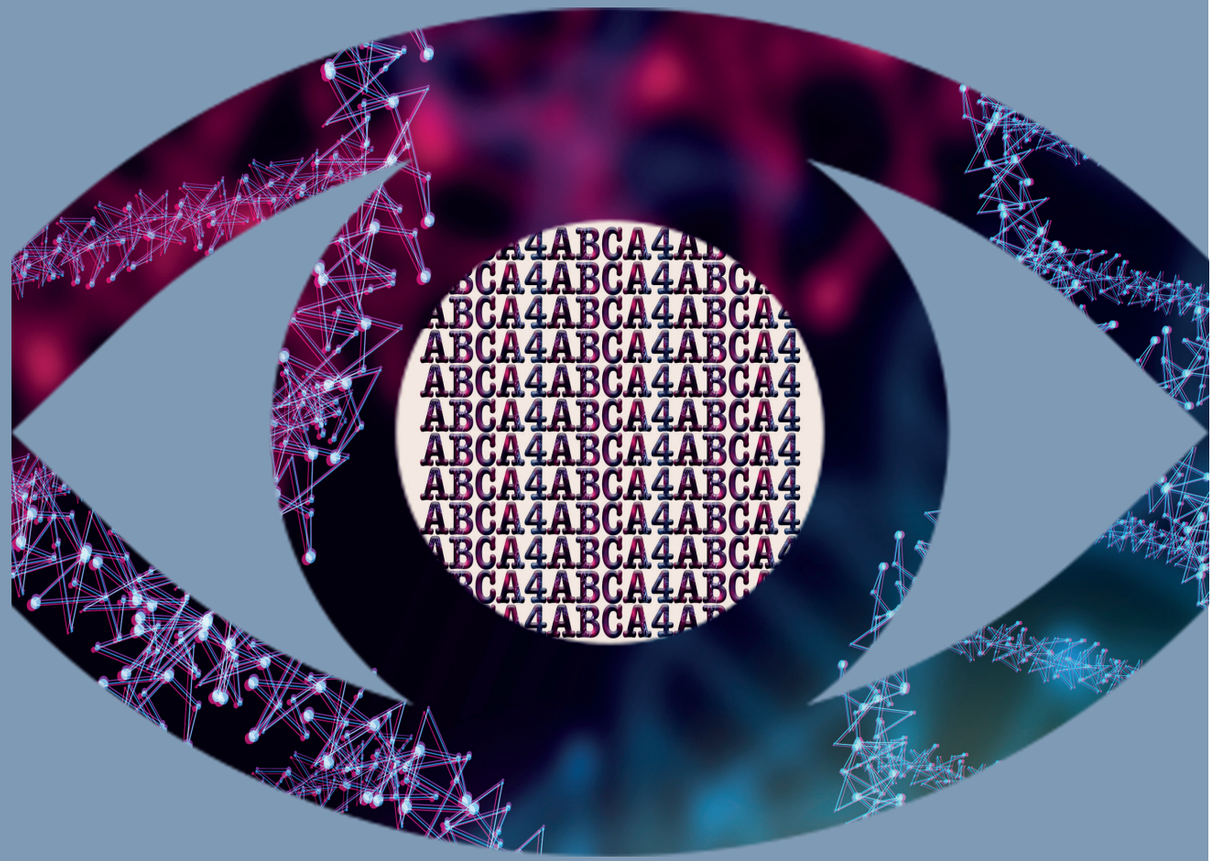
Genomic and transcriptomic landscape of *ABCA4*-associated Stargardt disease

Mubeen Khan

Genomic and transcriptomic landscape of *ABCA4*-associated Stargardt disease



Mubeen Khan

Genomic and transcriptomic landscape of *ABCA4*-associated Stargardt disease

Mubeen Khan

Mubeen Khan, 2020
Genomic and transcriptomic landscape of *ABCA4*-associated Stargardt disease

The work presented in this thesis was carried out within the Department of Human Genetics, Donders institute of Brain, Cognition and Behaviour and Radboud university medical center, Nijmegen, The Netherlands.

The publication of this thesis was financially supported by the Department of Human Genetics, Radboud University Medical Center.

ISBN 978-94-6284-218-2

Design cover by Shahd Elmelik

Design inside and print by Book Builders

Genomic and transcriptomic landscape of *ABCA4*-associated Stargardt disease

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op maandag 29 juni 2020
om 10:30 uur precies

door

Mubeen Khan

geboren op 8 augustus 1988
te Islamabad, Pakistan

Promotor

Prof. dr. Frans P.M. Cremers

Copromotoren

Dr. Rob W.J. Collin

Dr. Alejandro Garanto

Manuscriptcommissie

Prof. dr. Anneke I. den Hollander

Prof. dr. Isabelle Audo (Sorbonne Université, Frankrijk)

Prof. dr. Peter A.C. 't Hoen

Paranimfen

Duaa Elmelik

Zeinab van Gestel

Genomic and transcriptomic landscape of *ABCA4*-associated Stargardt disease

DOCTORAL THESIS

to obtain the degree of doctor from
Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. J.H.J.M. van Krieken,
according to the decision of the Council of the Deans
to be defended in public on Monday, June 29, 2020
at 10:30 hours

by

Mubeen Khan

born on August 8, 1988 in
Islamabad, Pakistan

Supervisor

Prof. dr. Frans P.M. Cremers

Co-supervisors

Dr. Rob W.J. Collin

Dr. Alejandro Garanto

Doctoral thesis committee

Prof. dr. Anneke I. den Hollander

Prof. dr. Isabelle Audo (Sorbonne Université, France)

Prof. dr. Peter A.C. 't Hoen

Paranymphs

Duaa Elmelik

Zeinab van Gestel

To my beloved family

Table of contents

List of abbreviations

2OMe/PS	2-*O*-methyl-modified RNA AONs with phosphorothioate
*	stop codon
A2-E	N-retinylidene-N-retinylethanolamine
AAV	adeno-associated vector
ACHM	achromatopsia
ACMG	American College of Medical Genetics
AD	autosomal dominant
ADP	adenosine diphosphate
adRP	autosomal dominant retinitis pigmentosa
AF	allele frequency
ALK-001	C20-D3-vitamin A
AMD	age-related macular degeneration
AON	antisense oligonucleotide
AR	autosomal recessive
arCRD	autosomal recessive cone-rod dystrophy
arRP	autosomal recessive retinitis pigmentosa
ATP	adenosine triphosphate
BAC	bacterial artificial chromosome
BB	basal body of connecting cilium
BBS	Bardet-Biedl syndrome
BCVA	best-corrected visual acuity
b-FGF	fibroblast growth factor, basic
bp	base pair
BPS	branch point site
c-MYC	MYC proto-oncogene, BHLH transcription factor
CC	connecting cilium
CD	cone dystrophy
cDNA	complementary deoxyribonucleic acid
CHM	choroideremia
CHX	cycloheximide
CK30-NPs	poly lysine-based compacted DNA nanoparticles
CMV	cytomegalovirus
CNV	copy-number variant
Del	deletion
DGV	Database of Genomic Variants
dHPLC	denaturing high performance liquid chromatography
DI	deep intronic
DMD	Duchenne muscular dystrophy
DNA	deoxyribonucleic acid
ERDC	European Retinal Disease Consortium
ERG	electroretinogram
ESCs	embryonic stem cells

ESE	exonic splice enhancer
ESS	exonic splice silencer
EVR	exudative vitreoretinopathy
ExAC	exome aggregation consortium database
EYS	eyes shut homolog
FA	fluorescein angiography
FAF	fundus autofluorescence
ffERG	full-field electroretinography
fs	frameshift
GB	gigabyte
GCL	ganglion cell layer
gnomAD	genome aggregation database
GoNL	Genome of The Netherlands
HEK293T	human embryonic kidney 293 expressing the SV40 large T antigen
hg19	human genome version 19
HGMD	Human Gene Mutation Database
HGVS	Human Genome Variation Society
HSF	human splicing finder
IBD	identical by descent
Indel	insertions or deletions
iPSCs	induced pluripotent stem cells antigen
IRDs	inherited retinal diseases
IS	inner segment
ISE	intronic splice enhancer
ISS	intronic splice silencer
kb	kilobase
LCA	Leber congenital amaurosis
LOD	logarithm of the odds
LOVD	Leiden open (source) variation database
MAF	minor allele frequency
Mb	megabase
MD	macular dystrophy
MI	mitochondrial inheritance
MLPA	multiplex ligation-dependent probe amplification
MT	mutant
mRNA	messenger-RNA
N	nucleus
n.a.	not applicable
NBD	nucleotide binding domain
NCSS	non-canonical splice site
nFE	non-Finnish European
ng	nanogram
NGS	next generation sequencing
NHEJ	non-homologous end joining
nm	nanometer

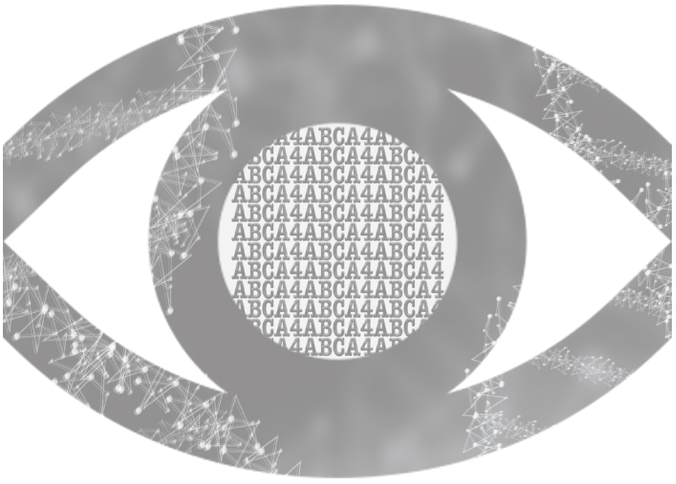
NMD	nonsense-mediate decay
non-LTR	non-long terminal repeats
nt	nucleotide
OCT	optical coherence tomography
OCT3/4	octamer-binding transcription factor
OAT	ornithine aminotransferase
ONL	outer nuclear layer
OS	outer segment
PBS	phosphate buffer saline
PCR	polymerase chain reaction
PD	pseudodominant
PE	pseudoexon
PEG	polyethylene glycol
PPCs	photoreceptor precursor cells
pre-mRNA	precursor-messenger RNA
rAAV	recombinant adeno-associated virus
RmP	rim protein
RNA	ribonucleic acid
RP	retinitis pigmentosa
RP1	retinitis pigmentosa 1
RPE	retinal pigmented epithelium
RT-PCR	reverse transcription PCR
SAS	splice acceptor site
SC35	splice enhancer motif recognized by the human SR protein SC35
SD	standard deviation
SDS	splice donor site
SFs	splicing factors
SF2/ASF	splice enhancer motif recognized by the human SF2/ASF SR protein
SLS	Senior-Løken syndrome
smMIPs	single molecule molecular inversion probes
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
SON	sense oligonucleotide
SRp40	splice enhancer motif recognized by the human SR protein SRp40
SRp55	splice enhancer motif recognized by the human SR protein SRp55
SSCP	single strand conformation polymorphism
SSFL	SpliceSiteFinder-Like
STGD1	Stargardt disease type 1
SVs	structural variants
Tm	melting temperature
UPD	uniparental isodisomy
USH	Usher syndrome

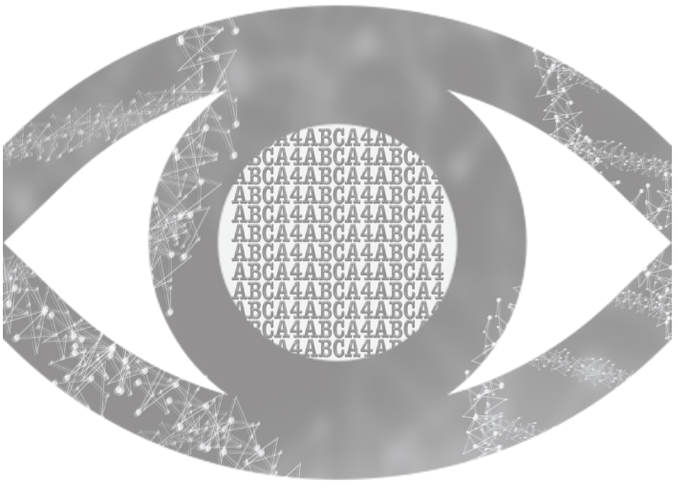
UTR	untranslated region
VEGF	vascular endothelial growth factor
VUS	variant of unknown significance
WES	whole exome sequencing
WGS	whole genome sequencing
WT	wild-type
XL	X-linked recessive

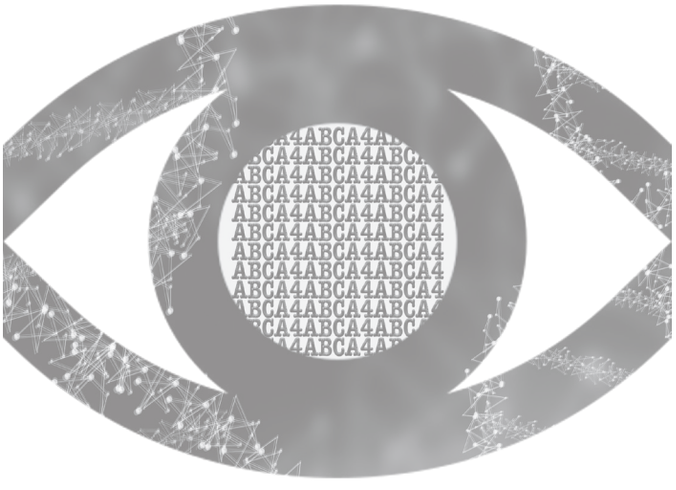
Genes names are not included in this list, Information regarding gene names is available on the Retinal Information Network (RetNet, available at https://sph.uth.edu/retnet/).

Chapter 1

Introduction







Chapter 1.1

General introduction

The eye

The eye is a specialized sensory organ that can detect and partially process visual images. The eye can be divided into an anterior and a posterior compartment. The anterior part comprises the cornea, anterior chamber, conjunctiva, pupil, ciliary body and lens (**Figure 1A**). The posterior part consists of the vitreous body, choroid, retina and optic nerve. Light passes through the different layers of anterior part of the eye, entering via the cornea and travelling through the iris and pupil and continues its path through lens, which refract the light to focus the image on the retina. Here, light is converted into electrochemical signals that are transferred through the optic nerve to the visual centers in the brain (**Figure 1A**). In humans, about one third of the cerebral cortex is dedicated to analyzing and perceiving visual information. Although we are constantly bombarded by a huge volume of these signals, the human visual system has evolved to select the most important stimuli.

Retina

More than a century after the discovery of the retina by Santiago Ramón y Cajal, advanced neuroanatomical studies have explored the retinal structure and function. These studies have shown that like all other structures of the central nervous system, the retina is a highly complex and subtle structure (Masland 2001). It is a light-sensitive tissue that lies at the back of the eye, located between the choroid and the vitreous. It originates from the embryonic neural ectoderm which invaginates further in a two-walled optic cup. The inner wall of the optic cup differentiates into the neural retina, while the outer wall transforms into the retinal pigment epithelium (RPE). The RPE plays an important role in the nourishment of the neural retina and many other processes (see below).

The neural retina is a multilayered tissue comprised of neurons arranged in three cellular layers which are interconnected by two alternating neuronal synapses. Based on the function, retinal cells can be divided into two major types, light-sensitive photoreceptors (rods and cones) and specialized neural cells (i.e., bipolar cells, horizontal cells, amacrine cells and ganglion cells). These cells create an intricate and dynamic neuronal circuit by vertical or horizontal connections. The main purpose of the retina is to convert light into neural signals, which can be then processed by the brain. To serve this purpose, the vertical connections between photoreceptors, bipolar cells and ganglion cells constitute the portion of the circuit that provides excitatory signals, while horizontal and amacrine cells are interneurons able to integrate and modulate the signals flowing through the vertical pathway (**Figure 1B**).

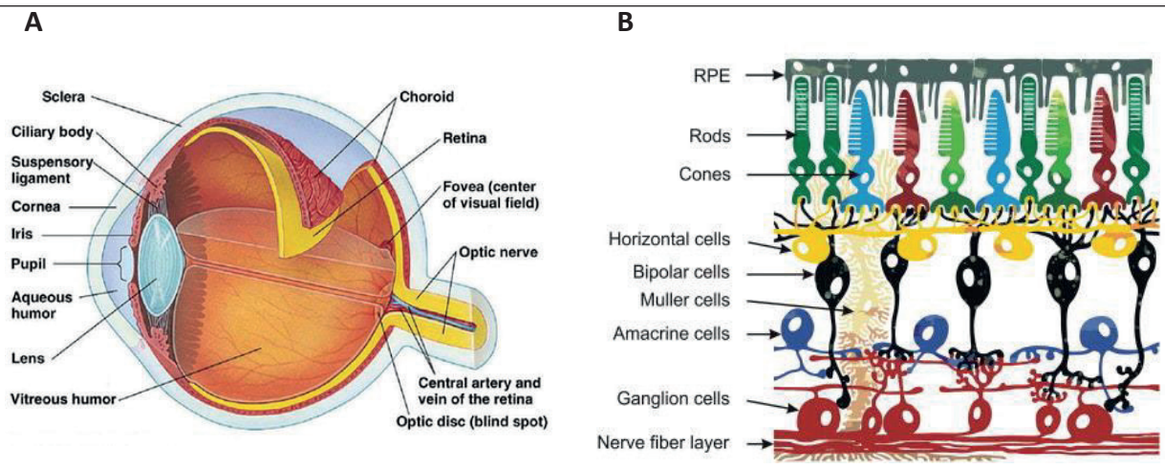


Figure 1. Anatomy of the human eye and cellular organization of the retina. A. Schematic overview of the human eye. Figure adapted from (http://www.eyerisvision.com/anatomy-of-the-eye.html#). **B.** Schematic representation of the various cell types within two retinal layers. Figure adapted with permission from (Slijkerman et al. 2015).

Although the exact role of non-photoreceptor cells is not yet completely known, the function of cones and rods photoreceptors has been well established. Photoreceptors are unequally present and distributed across the human retina. Like all other vertebrates, rod cells are overrepresented in the human retina (Hoon et al. 2014). They account for almost 95% (~120 million) of the photoreceptors. Rods have exquisite sensitivity to light and can detect even a single photon (Rieke, 2000; Sampath and Rieke, 2004). They are responsible for scotopic vision, i.e., vision under dim light. The remaining 5% (~6 million) of human photoreceptors are cones. They are highly concentrated in the central part of the retina called the fovea (~160,000/mm2) and decrease towards the periphery. Cones are 100 times less sensitive than rods, but exhibit much faster response kinetics during phototransduction. Furthermore, each cone photoreceptor type is most sensitive to a specific wavelength of light. Thus, cones are engaged in photopic vision, i.e., bright-light and high acuity color vision. Rods are absent in the fovea but rapidly increase with a maximal density at ~20° eccentricity and subsequently decrease in density towards the periphery of the retina.

The photoreceptors have four distinct compartments, i.e., the cell body, which includes the nucleus; the inner segment (IS); the outer segment (OS); and the synaptic terminal (**Figure 2**). Inner and outer segments perform specific biological functions. For instance, ISs serve the metabolic function by harboring a large number of mitochondria which supply energy for the high metabolic demand of the photoreceptor, while the OSs contains unique proteins including the visual pigments and other components of the phototransduction cascade that allows the conversion of light stimuli to electrical signals.

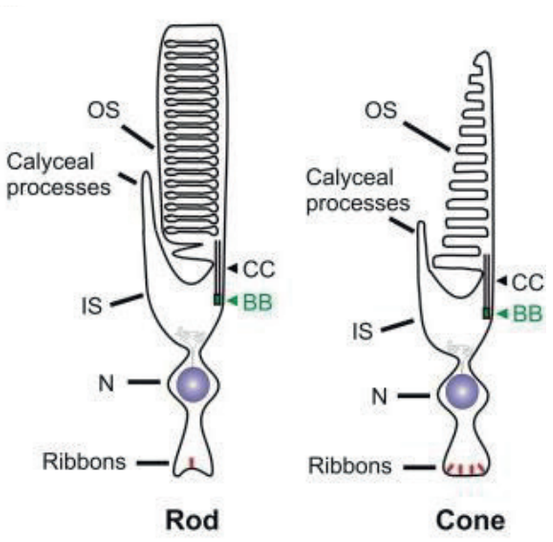


Figure 2. **Schematic representation of photoreceptor cells.** (OS) outer segment; (CC) connecting cilium; (BB) basal body of connecting cilium; (IS) inner segment; (N) nucleus. Figure adapted with permission from (Slijkerman et al. 2015).

The structure of the OS differs significantly between rods and cones. The rod OS is a cylindrical structure comprised of ~1000 membrane discs that are densely packed with the visual pigment rhodopsin for high-sensitivity light detection, and enclosed by a plasma membrane (Gilliam et al. 2012). Cones OSs on the other hand consist of continuous discs derived from the invagination of the plasma membrane and have a conical shape. Membrane disc turnover in rod photoreceptors is notably fast and follows a circadian rhythm: every night, ~10% of rod OS discs are shed and phagocytosed by the RPE at the distal end and new discs are added at the proximal end, which results in the complete renewal of OS within 10-15 days (Lee et al. 2001).

The RPE is a specialized monolayer of pigmented hexagonal cells located between the choroid and neural retina. The apical membrane of the RPE faces the photoreceptor outer segments whereas the basolateral membrane faces Bruch's membrane that separates the RPE from fenestrated vessels of the choroid (i.e., the most outer vascular layer of the eye). The RPE constitutes the outer blood-retinal barrier (Simo et al. 2010). It plays a crucial role in the retinal integrity by preventing the infusion of toxic substances due to tight junctions between neighboring RPE cells (Strauss 2005). Moreover, the RPE performs several important functions including 1) visual pigment regeneration (i.e., enzymatic transition from *all-trans-retinol* to *11-cis-retinal*) is certainly one of the most critical for proper retinal physiology, 2) the reduction of the phototoxic effects to the retina, by absorbing excessive light that is not captured by photoreceptors via melanosomes (Haddad et al. 2012), 3) regulation of the transepithelial transport of ions, fluid, and metabolites from the subretinal space to the blood and vice versa (Halford et al. 2014), and 4) the phagocytosis of shed photoreceptor outer segments to permit their renewal (Kevany and Palczewski 2010). Phagocytized membrane discs are enzymatically digested by

lysosomal activity and undigested portions of outer segment (e.g., lipid material) can be retained in the RPE, accumulating in lipofuscin granules.

References

Gilliam JC, Chang JT, Sandoval IM, Zhang Y, Li T, Pittler SJ, Chiu W, Wensel TG. 2012. Three-dimensional architecture of the rod sensory cilium and its disruption in retinal neurodegeneration. *Cell* **151**: 1029-1041.

Haddad NM, Waked N, Bejjani R, Khoueir Z, Chouery E, Corbani S, Megarbane A. 2012. Clinical and molecular findings in three Lebanese families with Bietti crystalline dystrophy: report on a novel mutation. *Mol Vis* **18**: 1182-1188.

Halford S, Liew G, Mackay DS, Sergouniotis PI, Holt R, Broadgate S, Volpi EV, Ocaka L, Robson AG, Holder GE et al. 2014. Detailed phenotypic and genotypic characterization of bietti crystalline dystrophy. *Ophthalmology* **121**: 1174-1184.

Hoon M, Okawa H, Della Santina L, Wong RO. 2014. Functional architecture of the retina: development and disease. *Prog Retin Eye Res* **42**: 44-84.

Kevany BM, Palczewski K. 2010. Phagocytosis of retinal rod and cone photoreceptors. *Physiology (Bethesda)* **25**: 8-15.

Lee J, Jiao X, Hejtmancik JF, Kaiser-Kupfer M, Gahl WA, Markello TC, Guo J, Chader GJ. 2001. The metabolism of fatty acids in human Bietti crystalline dystrophy. *Invest Ophthalmol Vis Sci* **42**: 1707-1714.

Masland RH. 2001. The fundamental plan of the retina. *Nat Neurosci* **4**: 877-886.

Simo R, Villarroel M, Corraliza L, Hernandez C, Garcia-Ramirez M. 2010. The retinal pigment epithelium: something more than a constituent of the blood-retinal barrier--implications for the pathogenesis of diabetic retinopathy. *J Biomed Biotechnol* **2010**: 190724.

Slijkerman RW, Song F, Astuti GD, Huynen MA, van Wijk E, Stieger K, Collin RW. 2015. The pros and cons of vertebrate animal models for functional and therapeutic research on inherited retinal dystrophies. *Prog Retin Eye Res* **48**: 137-159.

Strauss O. 2005. The retinal pigment epithelium in visual function. *Physiol Rev* **85**: 845-881.

Chapter 1.2

Identification and Analysis of Genes Associated with Inherited Retinal Diseases

Identification and Analysis of Genes Associated with Inherited Retinal Diseases

Mubeen Khan, Zeinab Fadaie, Stéphanie S. Cornelis, Frans P.M. Cremers, Susanne Roosing

Adapted from*:* ***Methods Mol Biol.*** 2019; **1834**: 3-27.

Abstract

Inherited retinal diseases (IRDs) display a very high degree of clinical and genetic heterogeneity, which poses challenges in finding the underlying defects in known IRD-associated genes and in identifying novel IRD-associated genes. Knowledge on the molecular and clinical aspects of IRDs has increased tremendously in the last decade. Here, we outline the state-of-the-art techniques to find the causative genetic variants, with special attention for next-generation sequencing which can combine molecular diagnostics and retinal disease gene identification. An important aspect is the functional assessment of rare variants with RNA and protein effects which can only be predicted *in silico*. We therefore describe the *in vitro* assessment of putative splicing defects in human embryonic kidney cells. In addition, we outline the use of stem cell technology to generate photoreceptor precursor cells from patients' somatic cells which can subsequently be used for RNA and protein studies. Finally, we outline the *in silico* methods to interpret the causality of variants associated with inherited retinal disease and the registry of these variants.

Introduction

Spectrum of Inherited Retinal Diseases

Inherited retinal diseases (IRDs) represent a clinically and genetically heterogeneous group of disorders affecting the retina. These diseases can be classified clinically based on whether they predominantly affect the rods (e.g., retinitis pigmentosa, RP), the cones (e.g., cone and cone-rod dystrophies, CD and CRD), or cause a more generalized photoreceptor disease (e.g., Leber congenital amaurosis, LCA) (Berger et al. 2010). Most IRDs are associated with a gradual deterioration throughout life, while some appear non-progressive (e.g., congenital stationary night blindness, achromatopsia, some forms of LCA). The differential diagnosis between various IRDs can sometimes be complicated, as the clinical features of some IRDs can be very similar, both at early and late stages (**Figure 1**).

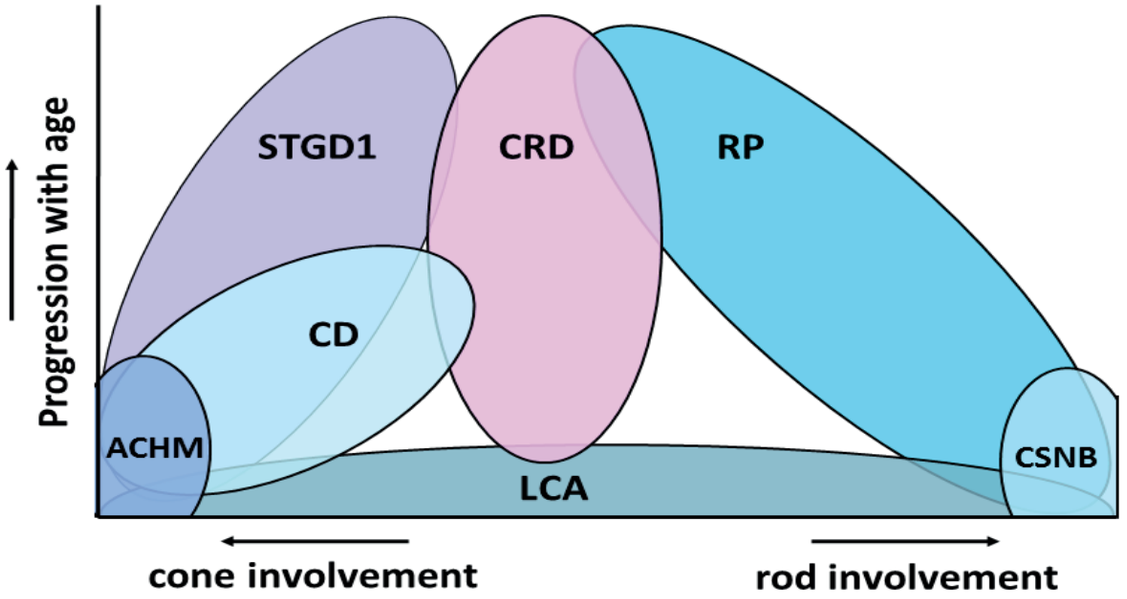


Figure 1. Phenotypic overlap between autosomal recessive retinal diseases. Patients with achromatopsia (ACHM) display a virtually stationary disease course in which cones are principally defective. At end stages, cone dystrophy (CD) can hardly be distinguished from cone-rod dystrophy (CRD). Patients with Stargardt disease (STGD1) later in life show mid-peripheral defects similar to CRD patients. Patients with retinitis pigmentosa (RP) initially display night blindness, followed by tunnel vision due to rod defects which very often progresses to complete blindness when the cones are also afflicted. In patients with Leber congenital amaurosis (LCA), the defects can occur in both types of photoreceptors, or in Müller or RPE cells, and therefore both clinical and molecular genetic overlap with CD, CRD or RP can be expected. Patients with congenital stationary night blindness (CSNB) show a rod-specific defect.

IRDs can be inherited in an autosomal recessive, autosomal dominant, and X-linked mode of inheritance (**Figure 2**). For isolated males and females with RP (**Figures 2B, C, H, I**), all inheritance patterns have been observed which are partially explained by the observation that *de novo* mutations are a significant cause of dominant non-familial disease (Neveling et al. 2013; Glockle et al. 2014).

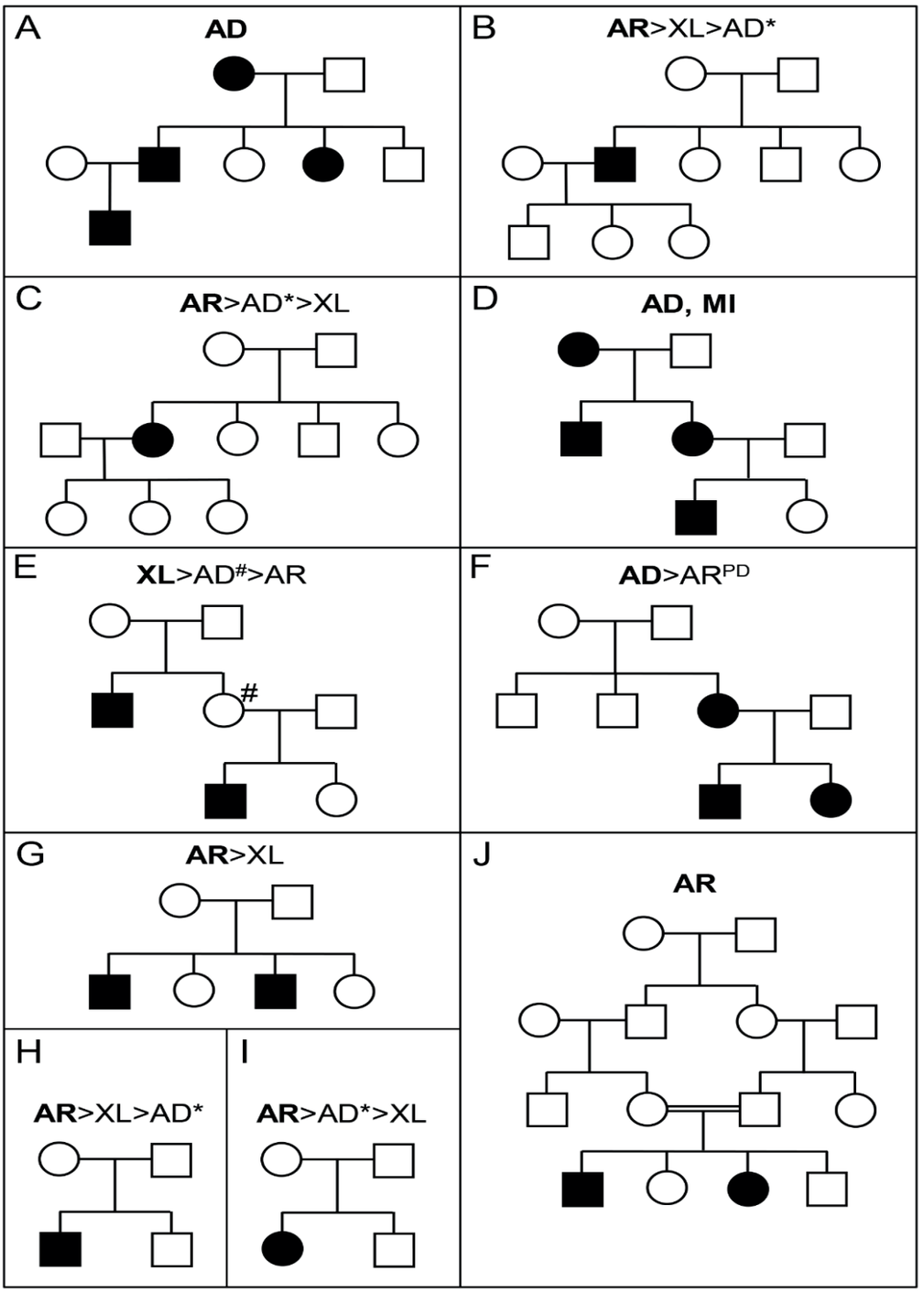


Figure 2. Examples of the different modes of inheritance observed in retinal diseases. Illustration of inheritance models based on the occurrence and gender of affected individuals

and their position in the pedigree. In bold above the pedigrees, the most likely modes of inheritance are given, followed by the less likely modes of inheritance. AD, autosomal dominant; AR, autosomal recessive; MI, mitochondrial inheritance; PD, pseudodominant (= autosomal recessive) inheritance; XL, X-linked recessive; **de novo* mutation; #non-penetrant individual.

Although some diseases are caused by mutations in a relatively small number of genes, the most prevalent IRDs are genetically highly heterogeneous with many causative genes (**Figure 3**). The extreme example is RP, which has currently been associated with mutations in 84 different genes. To date, mutations in 261 genes have been identified in patients with non-syndromic or syndromic IRD, and it is estimated that these genes account for ~80% of the genetic disease load (den Hollander et al. 2010; Roosing et al. 2014). As the most recently identified novel genetic defects in IRDs are found in single cases or families, it is difficult to predict how many IRD-associated genes are yet to be identified. All known IRD-associated genes and the corresponding modes of inheritance can be found at http://www.sph.uth.tmc.edu/RetNet/.

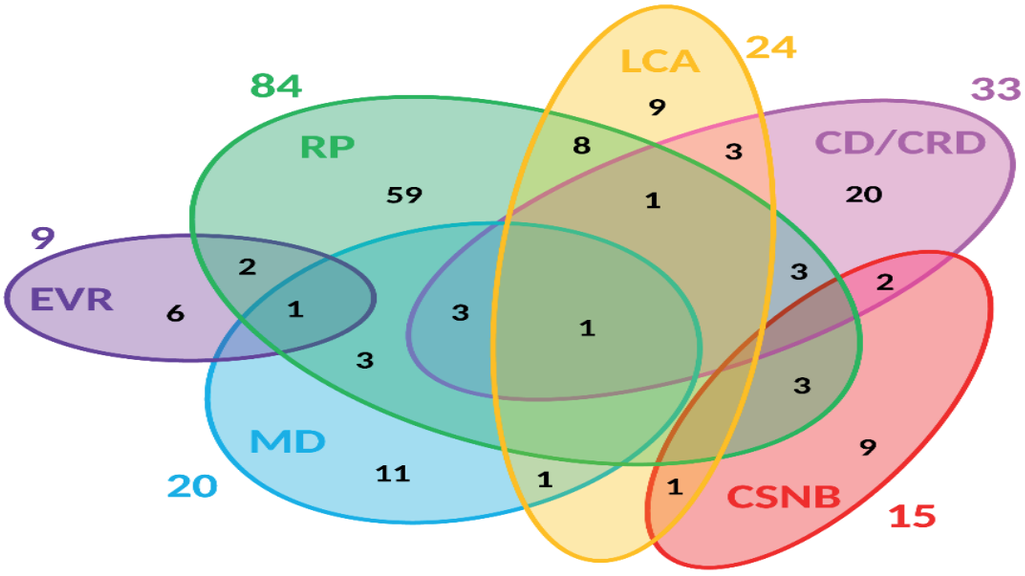


Figure 3. Genetic overlap between non-syndromic monogenic retinal diseases. Clinical diagnoses are indicated by colored circles. In the overlapping areas we provide the number of genes implicated in different phenotypes. Colored numbers outside the balloons represent the total number of genes associated with these phenotypes. RP: retinitis pigmentosa; CSNB: congenital stationary night blindness; LCA: Leber congenital amaurosis; CD/CRD: cone and cone-rod dystrophies; MD: macular degeneration; EVR: exudative vitreoretinopathies.

The changing landscape of retinal disease gene identification

The methods and tools available for gene identification have continuously evolved in the last three decades. The first retinal disease-associated gene identified was the ornithine aminotransferase (*OAT*) gene involved in gyrate atrophy. An enzymatic

defect of ornithine aminotransferase activity was measured in patient's cells in 1977 (Valle et al. 1977), and eleven years later the *OAT* gene was cloned and the first mutation identified (Mitchell et al. 1988). In 1990, mutations in the rhodopsin gene were identified in patients with autosomal dominant retinitis pigmentosa using a candidate gene approach (Dryja et al. 1990), after linkage analysis in a large Irish adRP family had pointed towards a genomic region encompassing this gene (McWilliam et al. 1989). In the same year, the choroideremia (*CHM*) gene was identified using a positional cloning approach by mapping deletions in patients with syndromic and non-syndromic choroideremia (Cremers et al. 1990).

The candidate gene approach, the search for IRD-associated variants in genes encoding proteins with known crucial functions in the retina has been very successful. The identification of IRD-associated genes through their genomic position (i.e., positional cloning) as determined by linkage analysis (see 2.8) has been used effectively, though this generally requires the availability of large families or a large set of families in which the same locus is involved. In the early years linkage analysis using polymorphic microsatellite markers was a labor-intensive method, but was lifted to a fast and genome-wide approach with the development of microarray technology allowing rapid genotyping of thousands of single nucleotide polymorphisms (SNPs) spread across the genome. SNP microarrays have also proven very valuable for homozygosity and identity-by-descent (IBD) mapping of recessive disease genes (see 2.5), not only in consanguineous families, but also in small families and single patients of non-consanguineous marriages (Collin et al. 2011).

We are witnessing a new era in disease gene identification with the introduction of next-generation sequencing, allowing the analysis of all genes in a defined linkage interval, all exons in the genome (whole exome sequencing - WES), or even the entire genomic sequence (whole genome sequencing - WGS). This also brings new challenges, such as data analysis and interpretation of genomic variants. Given the huge number of variants present in a patient's genome, positional information on where the causative gene may be localized (e.g., by linkage analysis and/or homozygosity mapping) remains very helpful to pinpoint the genetic defect.

Employing WGS, thousands of rare single nucleotide variants (SNVs) and structural variations (SVs) are found in every individual and it remains very challenging to identify the causal variant(s). A functional read-out is required to identify the culprit variant(s). Gene-specific mRNA analysis or genome-wide mRNA analysis (transcriptome analysis) may identify quantitative or structural defects in mRNAs.

Importance of molecular diagnostics

Receiving a molecular diagnosis becomes increasingly important with the development of (gene) therapy for IRDs (den Hollander et al. 2010; Roosing et al. 2014). Up to nine years ago, it was not possible to slow down, stabilize or treat the vision impairment in

patients with IRDs. This changed for a small group of patients with *RPE65* mutations, as gene augmentation was successfully and safely applied through subretinal injections of recombinant adeno-associated viruses (rAAVs) in Phase 1/2 trials (Bainbridge et al. 2008; Hauswirth et al. 2008; Maguire et al. 2008). rAAVs transduce the retinal pigment epithelium (RPE) cells, upon which the viruses are shuttled to the nucleus, and the rAAV vector remains a stable extrachromosomal element. In the meantime, many more patients have been treated in two centers in Philadelphia and one in London. Vision improvement was variable and in general modest and appears to be more effective in younger patients. A Phase 3 trial was conducted in one center using an improved rAAV vector which resulted in increased subjective and objective vision in the treated eye versus the untreated eye (Russell et al. 2017). Gene augmentation targeting photoreceptors and the RPE was also successfully performed in a Phase 1/2 trial in choroideremia patients (MacLaren et al. 2014; Edwards et al. 2016).

In addition, an oral 9-cis retinoid supplementation therapy seems effective in patients with *RPE65* and *LRAT* mutations (Koenekoop et al. 2014; Scholl et al. 2015). Several therapies that will be developed in the next years will be gene-, or even mutation-specific, emphasizing the importance for patients to receive a molecular diagnosis. An overview of all ongoing gene therapy trials can be found at http://clinicaltrials.gov.

To provide a more accurate prognosis, and to determine which forms of IRD would most likely benefit from (gene) therapy, patients should be thoroughly clinically examined using standardized protocols. Optical coherence tomography (OCT) studies have shown that certain forms of IRDs are likely less suitable for therapy, while in other forms photoreceptors remain viable for a prolonged period (Pasadhika et al. 2010).

IRDs are sometimes the first sign of syndromic disease, such as Senior-Loken syndrome that involves renal failure. Since the ocular phenotype precedes the manifestation of kidney abnormalities, there is often a delay in the diagnosis of nephronophthisis. This causes a risk for sudden death from fluid and electrolyte imbalance. Determining an early molecular diagnosis allows physicians to monitor patients carrying mutations in genes associated with a syndrome more closely, and provide better healthcare for kidney disease and other systemic features. Nephronophthisis patients that receive kidney transplants have excellent outcomes compared with the general pediatric transplant population.

Techniques

Sanger sequencing

Sanger sequencing is still the gold standard of DNA sequencing and mutation identification. Sanger sequencing is based on the incorporation of deoxynucleotides and fluorochrome-labeled dideoxynucleotides using DNA polymerase, the latter of

which abrogate the replication of a DNA fragment at random positions. In this way, a mixture of DNA fragments is synthesized and size-separated through capillary electrophoresis. The most widely used apparatus (Applied Biosystems) can analyze up to 96 samples in parallel. Sanger sequencing is preceded by PCR amplification of a DNA fragment of interest, and is the most widely used sequencing technique for a limited number of exons or amplicons. Its advantages are its accuracy, flexibility and speed. The costs are relatively low as long as the number of amplicons is limited, i.e., less than 20. If NGS-based techniques are available, genes with more exons or amplicons rather can be sequenced using NGS-based approaches (see below). Sanger sequencing however is the preferred method to study the segregation of variants in family members of the IRD proband.

Next generation sequencing

Although Sanger sequencing is considered as a gold standard for sequencing, it is not appropriate for the identification of disease-associated genes and variants involved in genetic heterogeneous disorders (Neveling et al. 2013; Siemiatkowska et al. 2014). In the last decade next generation sequencing (NGS) or massively parallel sequencing has been developed and continuously improved in terms of accuracy and throughput. This revolutionized all aspects of biological sciences and healthcare through the identification of an enormous number of genomic variants and their role in disease etiology (Levy et al. 2007; Myllykangas et al. 2011). These technological advancements sparked the identification of not only novel causative variants but also new causal genes for IRDs. Since the NGS-based discovery of the *TSPAN12* gene to underlie familial exudative vitreoretinopathy, 53 new IRD-associated genes have been identified using this technology (Broadgate et al. 2017).

Targeted next generation sequencing

With the emergence of high throughput sequencing technologies many human genomes have been completely sequenced. However, due to the relatively high costs and the complexity of data analysis, it is economically not feasible to screen many individuals with genetically heterogeneous disorders. Therefore, to overcome these issues often genomic regions of interest are selectively enriched and sequenced using high NGS, commonly known as targeted NGS or panel based sequencing (Tewhey et al. 2009).

Targeted approaches have several advantages over the holistic approach, i.e., lower costs in research settings, deeper coverage and the faster generation of data (Lin et al. 2012). Targeted approaches have been widely used to investigate novel variants in a large group of known genes associated with genetic disorders as well as to study

other specific genomic regions such as CpG islands and regulatory elements (Head et al. 2014). Therefore, this approach provided a better understanding of genetic etiology of heterogeneous disorders such as IRDs (Glockle et al. 2014).

To enrich targeted regions, various methods can be used depending on the aims of the study. The most widely used enrichment strategies include the hybridization approaches, which are multiplexed array-hybridization (NimbleGen) or in-solution hybridization (Agilent SureSelect Target Enrichment System, NimbleGen SeqCap EZ), highly multiplexed PCR (molecular inversion probes) and targeted circularization (Haloplex, Agilent Technologies) (Absalan and Ronaghi 2007; Jacob et al. 2007; Turner et al. 2009; Igartua et al. 2010; Teer et al. 2010; O'Roak et al. 2012). Of all these enrichment approaches, the molecular inversion probe-based multiplex PCR enrichment strategy has shown several advantages over other capture methods in terms of cost, ease of use, sensitivity and specificity. This can be employed for simultaneous analysis of hundreds of patients, with a drop in costs with an increasing number of analyzed cases (Mamanova et al. 2010; Cremers F.P.M, personal communication).

Whole exome sequencing

More than 20 different NGS platforms are commercially available. However, the most widely used platforms are Life Technologies systems (Ion Torrent, Proton and SOLiD 5500 xl), and the IIIumina platforms (e.g., NextSeq500, Miseq, Genome Analyzer IIx). Although all these platforms have a much higher throughput than conventional Sanger sequencing, they still differ significantly. Important differences concern the read length and number of reads produced in a specific run (e.g., SOLiD 5500XL: up to 75 bp (+35 bp) reads; HiSeq2000: up to 2 × 100 bp reads) (Mardis 2011).

As these technologies focus on the protein-coding elements of human genes, they rely on the enrichment of small fragments. This enrichment is hampered by low and high GC-containing sequences. The latter is frequently observed in 5’ regions of genes. A median coverage of 50x is considered acceptable but generally will leave up to 5% of coding regions not or poorly covered (Fukunaga et al. 2014). As described below, better coverage is achieved using newly emerging NGS platforms including Pacific BioSciences (RS II and Sequel) and Oxford Nanopore (Oxford Nanopore MK 1 MinION Oxford Nanopore PromethION) using single molecule real-time sequencing approach generate longer reads ranging between 8 to 200 kb (Goodwin et al. 2016; Levy and Myers 2016).

Whole genome sequencing: short read versus long read techniques: pros and cons

WES is not able to detect deep-intronic variants or SVs that might affect RNA splicing or transcription (Siemiatkowska et al. 2014). WGS is a more comprehensive method to detect genetic defects (Meienberg et al. 2016) as it enables identification of SNVs and SVs such as complex genomic rearrangements, large deletions and insertions (Chesworth et al. 2014; Belkadi et al. 2015; Knoppers et al. 2015).

The costs of WGS are continuously decreasing which makes the technology accessible to more laboratories while the analysis of huge amounts of genomic variant data remains challenging (Genome of the Netherlands, 2014). First, approximately 3 million variants are detected in each sample. Furthermore, determining the effect of regulatory and intronic variants is still complex, even using current available prediction programs (Siemiatkowska et al. 2014; Richards et al. 2015).

Generally, there are two different methods for performing WGS, long reads versus short reads. The sequencing read length depends on the purpose of the experiment and also the properties of the instruments. In regular short-read sequencing, the read length is between 25 and 100 bp, whereas 10 to 15 kb is usually sequenced in the long-read approach (Shendure et al. 2004; Whiteford et al. 2005; Bentley 2006).

Long-read sequencing is often performed when the aim is to identify SVs in the genome like CNVs and provides a high level of completeness in detecting these, especially for regions with repetitive nucleotides or like hairpin structure. For example, WES typically will not enrich for *RPGR ORF15* sequences, a hotspot for mutations causing X-linked RP and CRD, due to its repetitive nature.

The disadvantages of long-read sequencing, compared to short-read sequencing, are that it is ~5 times more expensive than short read sequencing (1,000$ versus 200$ per GB (Goodwin et al. 2016) and the 15% error rate in long read sequencing, mostly Indels (Carneiro et al. 2012). Reducing the error rate is crucial the use of these reads in, e.g., *de novo* genome assembly (Chin et al. 2013; Salmela et al. 2017). Also transposable elements complicate genome reconstruction due to their high sequence identity high copy number or complexity in genomic rearrangements. (McCoy et al. 2014).

Short-read sequencing is the preferred method for the detection of single nucleotide variations (SNVs) due to its deeper coverage, accuracy (Ajay et al. 2011), and low cost. For this purpose, the reads should align sufficient and uniquely with the reference sequence, which can be achieved in general with a 25-100 bp read length (Bentley 2006).

Homozygosity mapping

In the past decades, microsatellite markers and short tandem repeat polymorphisms were used to detect the disease locus using linkage analysis and homozygosity mapping. These multi-allelic markers were replaced by SNP-arrays that contain hundred thousands of biallelic markers, which can be analyzed rapidly (Purcell et al. 2007; Alkuraya 2013). WES and WGS data contain SNP data which can now be used to perform homozygosity mapping (Abu Safieh et al. 2010). This not only can pinpoint the culprit gene or variant in consanguineous families in which the offspring of first-cousin marriages on average show homozygosity in 10% of their genome (Woods et al. 2006) but also can be informative to identify the causal variant and sometimes a novel causal gene - in non-consanguineous families (Collin et al. 2008; Collin et al. 2011).

Copy number variant mapping

Copy number variants (CNVs) are one of the most significant SVs in the human genome that involve DNA fragments typically longer than 50 bp (Zarrei et al. 2015), whereas smaller elements are categorized as small insertions or deletions (indels) (Conrad and Hurles 2007). Generally, 4.8-9.5% of the human genome, based on the stringency of the map, contributes to CNVs (Zarrei et al. 2015). Therefore, every healthy individual has almost one thousand CNVs within the genome (Chen et al. 2011; Pirooznia et al. 2015).

Furthermore, CNVs play an important role in the human diversity, which are estimated to lead to a 1.2% difference from the reference human genome as well as in the disease susceptibility (Pang et al. 2010; Pirooznia et al. 2015). The phenotypic effect of the CNVs can vary from evolutionary changes to embryonic lethality, although the adaptive traits can be different in various environments (Beckmann et al. 2007; Buchanan and Scherer 2008). Pathogenic CNVs can be associated with inherited (monogenic) diseases, such as IRDs, as well as multifactorial diseases. WES analysis genetically solves 55-60% of these cases (Combs et al. 2013; Haer-Wigman et al. 2017). The hidden genetic variations may be unrecognized CNVs and deep-intronic variations, which can be identified by WGS or gene-specific locus sequencing. CNVs can explain 18% of previously unsolved cases (Eisenberger et al. 2013; Bujakowska et al. 2017). The copy number, content, and positional information are the three genomic features that should be considered in order to identify the CNVs precisely (Hehir-Kwa et al. 2015). Accordingly, CNV maps based on different ethnic groups of healthy populations were designed which can be utilized to accurately assess the variability of the human genome (Zarrei et al. 2015). The map consists of microscopic (>3 Mb) and submicroscopic variations (50 bp to 3 Mb) (Zarrei et al. 2015). The phenotype-first research approach has uncovered more deletions than duplications as gain-of-function

variants associate generally with milder phenotypes. The consequence of a milder phenotype is a lower selective pressure, which becomes visible when employing genotype-first research (Buchanan and Scherer 2008).

There are two different approaches for studying CNVs: array-based and NGS-based. Overall, NGS-based approaches have a higher sensitivity and resolution than array-based approaches and are able to generate more precise sequence-level breakpoint resolution (Pang et al. 2010; Pang et al. 2014). However, duplications are more likely to be detected by array-CGH than by SNP-based array or by NGS techniques (Pinto et al. 2011).

In addition, WES can be used for CNV analysis in order to combine the detection of small and large variants (Hehir-Kwa et al. 2015). Nevertheless, the orientation and location of duplicated sequences is complicated and also some SVs such as inversions cannot be identified by WES analysis (Newman et al. 2015). Consequently, WGS is a more accurate and reliable technique for studying CNVs. Furthermore, it gives us the opportunity to analyze not only the exonic regions, but also the non-coding elements of the genome such as promoters, untranslated regions and intronic sequences, as well as enhancers and insulators. It allows recognizing all different types of SVs in the human genome and is an ultimate goal of genetic testing in diagnostic laboratories, although it still requires a few years to become the standard diagnostic method (Conrad et al. 2010; Hehir-Kwa et al. 2015; Zarrei et al. 2015).

Linkage analysis

Although WES and WGS potentially can identify causal defects in single patients, in some families with multiple affected cases, the causal defect cannot be identified in the absence of a functional read-out. In this situation, linkage and haplotype analysis can narrow-down the search for the causal defect. Linkage analysis, following genome-wide SNP genotyping, can be performed to determine the chromosomal region that segregates with a trait. The logarithm of the odds (LOD)-score is the $\log_{10}$-ratio of the likelihood that the disease locus and a given genomic marker (e.g., SNP) are linked versus the likelihood that they are unlinked, and is generally used as an outcome measure in linkage calculations. In order to reach statistically significant locus assignment by genome-wide genotyping, a minimum LOD score of 3.3 has to be obtained, whereas a LOD-score of 1.86 is suggestive for linkage (Lander and Kruglyak 1995). The more individuals (both affected and unaffected) are genotyped in the linkage analysis, the higher the final LOD-score will be. Generally, linkage analysis is only performed if a LOD-score of >2 can be obtained with the available relatives, which can be calculated by a linkage simulation prior to the actual genotyping.

Issues that occasionally can interfere with linkage analysis are the occurrence of phenocopies (e.g., affected relatives with the same phenotype but a different (genetic) cause), or non-penetrance (e.g., the occurrence of individuals that carry the same causative mutation but do not, or hardly, display the clinical phenotype). Especially in some dominant retinal diseases, e.g., familial exudative vitreoretinopathy (Boonstra et al. 2009) and adRP caused by mutations in *PRPF31* (Al-Maghtheh et al. 1996), non-penetrance is frequently observed. The actual linkage calculations can be performed by freely available software programs, like LINKAGE, Allegro, Genehunter, or SimWalk2. Graphical user interfaces for linkage analysis software on Microsoft Windows based operating systems like easyLINKAGE (Hoffmann and Lindner 2005) or Alohamora (Ruschendorf and Nurnberg 2005) allow to use each of these programs, with manually adjustable settings for the mode of inheritance, ethnic origin of the family, disease prevalence, and penetrance, amongst others. A more detailed description on linkage analysis is provided in several textbooks (Terwillinger and Ott 1994; Nyholt 2008).

Next, we provide a review to interpret the causality of variants associated with inherited retinal disease and the registry of these variants.

***In vitro* RNA splice assays**

Variants that have a potential effect on RNA splicing can be tested directly by reverse-transcription PCR of mRNA extracted from accessible human tissues such as lymphoblasts or fibroblasts. As the mRNA derived from a gene with a protein-truncating mutation in the last exon or in the last 50-55 nucleotides of the penultimate exon will undergo nonsense-mediated decay, it is preferred to suppress nonsense-mediated decay using e.g., cycloheximide in the last phase of culturing cells.

The majority of IRD-associated genes are however not or poorly expressed in these non-ocular tissues. Alternatively, *in vitro* splice assays can be performed in human embryonic kidney (HEK) cells. Classically, the exonic or intronic segment to be analyzed were amplified from genomic DNA from a patient carrying the potential splicing defect and cloned into a minigene splice vector that contains a ubiquitously expressed promoter, a transcriptional start site and at least two exons of another gene that flank the cloning site. Minigenes carrying the wild-type sequence and minigenes containing the variant are transfected in parallel in HEK293T cells, and after 48hr, cells are collected for RNA extraction. RT-PCR using primers annealing to the flanking vector exons was performed to analyze the effect of the putative splice variant (Hicks et al. 2005; Movassat et al. 2014).

We recently detected that when using small minigenes that lack the proper genomic context, i.e., flanking exons and *cis*-acting elements that influence exon recognition, *in vitro* results do not always correlate with splicing defects observed in

patient cells and also are inadequate to rigorously test the effect of deep-intronic variants. We therefore designed an alternative strategy to generate sizeable multi-exonic splice vectors by employing bacterial artificial chromosome (BAC) DNA spanning the entire gene of interest. In this manner, vectors can be designed in which each exon is flanked by at least one downstream and upstream exon, and their corresponding introns and allowing variants to be tested in their genomic context without the influence of the splice site of flanking exons used in the original vector. BAC clones facilitated the generation of large, multi-exon wild-type splice vectors. This approach can readily be applied to generate midigenes from all human genes known to be involved in IRDs, or other diseases. In this way, deep-intronic, noncanonical splice site or even coding variants with unclear functional effects can be assessed for their effect on splicing, as BAC clones are available for this purpose (Osoegawa and de Jong 2004; Sangermano et al. 2018).

In vivo splice studies using stem cell technology

In vitro splice assays in HEK293T cells have one disadvantage, i.e., retina-specific splicing patterns may be missed. As shown for the consequences of a deep-intronic variant in *CEP290*, defective splicing (in this case pseudoexon insertion) can be more prominent in retina-like cells than in fibroblasts (Parfitt et al. 2016). We and others used stem cell technology to generate induced pluripotent stem cells (iPSCs) from patient's fibroblasts and subsequently differentiated iPSCs to photoreceptor precursor cells (PPCs) (Tucker et al. 2013; Yoshida et al. 2014; Lukovic et al. 2015; Parfitt et al. 2016; Sangermano et al. 2016). In this way we were able to reveal the causative effect of an apparently mild non-canonical splice site variant, c.5461-10T>C, which turned out to result in the skipping of exon 39 or exon 39 and 40, rendering this the most frequent severe *ABCA4* variant in Stargardt disease (Sangermano et al. 2016). PPCs also have been used to pinpoint the splicing defects due to many other deep-intronic variants (Albert et al. 2018).

Variant data interpretation

After using NGS to sequence a patient's exome or genome, an enormous amount of variants needs to be analyzed to find possibly disease-causing variants. When these filtering steps have been performed, often a number of putative pathogenic variants is still remaining. To find out which variant is most likely pathogenic, the American College of Medical Genetics (ACMG) guidelines can be used. When a likely pathogenic variant is found in a gene that is not known to cause disease yet, online tools like GeneMatcher (https://genematcher.org/) can be used.

American College of Medical Genetics

To prioritize novel rare variants according to their functional impact, the American College of Medical Genetics (ACMG) provided a system to classify variants along a five-tier gradient from benign, likely benign, uncertain significance, likely pathogenic to pathogenic (Richards et al. 2015). The system was developed by a collaboration of several experts in the field and made use of eleven classification protocols from different diagnostic groups. Furthermore, it was tested by classifying variants of which the pathogenicity was already known.

For each variant one should check, among others, if the variant or the gene is already reported to be involved in the disease, whether the prediction of a truncating variant is reliable, whether the variant segregates, what the allele frequency of the variant is and what *in silico* pathogenicity prediction programs predict for this variant. *In silico* prediction programs mainly focus on the effect of missense variants such as PolyPhen (Adzhubei et al. 2010), SIFT (Kumar et al. 2009) and MutationTaster (Schwarz et al. 2014) or on the effect on splicing such as Human Splicing Finder (Desmet et al. 2009), GeneSplicer (Pertea et al. 2001) and MaxEntScan (Yeo and Burge 2004). For each validation of a pathogenicity criteria a score is given: PVS (Very Strong), PS (Strong), PM (Moderate), and PP (Supporting).

When all the criteria are checked, the guidelines provide rules to combine the scores to the five-tire classification. For example, few Strong criteria need to be met to classify a variant as pathogenic, whereas more than three Moderate criteria solely classify a variant as likely pathogenic. A disadvantage of the ACMG guidelines is that any proof or indication against pathogenicity will overrule the pathogenicity classification. This means that one incorrect study outweighs multiple other studies.

The guidelines clearly show how complex the pathogenicity classification of a variant is and clearly mention that even when you followed this strict system, there still is a chance that the classification is incorrect due to unforeseen factors that could play a role (Richards et al. 2015).

GeneMatcher

Likely pathogenic variants that match the genetic inheritance pattern can be found in genes of which the function is unknown. This gene then becomes a candidate gene for the disease. Especially for candidate disease genes of patients with a rare phenotype, the online Web-tool GeneMatcher (https://genematcher.org/) was developed. This tool allows researchers and clinicians to upload a gene with information about the identified genetic variants and the phenotype found in the patient that carries this variant or combination of variants. When a new gene is submitted to GeneMatcher and this gene is already present in GeneMatcher, submitters of this gene automatically receive an email and can get in touch (Sobreira et al. 2015).

European Retinal Disease Consortium

Specifically, for candidate genes IRDs, the European Retinal Disease Consortium (ERDC) was set up. The ERDC consists of thirteen European research groups, one Canadian group and one group from the USA who collaborate to unravel the genetics of rare IRDs (http://www.erdc.info/).

Sequence variant registries and databases

In order to classify a variant to be pathogenic, several databases exist to provide variant information per gene. Especially databases that provide phenotypic information that accompanies variants can be very useful. ClinVar (Landrum et al. 2016) and Leiden Open source Variation Databases (LOVDs) (Fokkema et al. 2011) are examples of databases that provide this type of genotype-phenotype information. Most databases make use of the Human Genome Variation Society (HGVS) nomenclature (http://varnomen.hgvs.org/) to describe variants in a consistent manner.

ClinVar

One of the databases that aims at improving the translation of genetic information into clinical healthcare is ClinVar, located at http://www.ncbi.nlm.nih.gov/clinvar, which is part of the NCBI's Entrez system. A genetic variant with an accompanying phenotype can be uploaded, as long as the submission is based on clinical testing, research or literature curation. Each submitted variant gets a record ID, to which also the submitter and the phenotype of the patient are added, as well as the curation of the variant being benign or pathogenic, as described by the submitter. Results from functional studies, e.g., *in vitro* studies, are not accepted as evidence for pathogenicity. In 2015, ClinVar added a star system to improve the reliability of the pathogenicity classification:

- Zero stars: no assertion criteria were provided by the submitter
- One star: criteria are met by one submitter
- Two stars: multiple single submitters meet the criteria and their interpretation is similar
- Three stars: criteria are met and reviewed by an expert panel
- Four stars: practice guidelines were met, meaning on top of the met criteria and revision by an expert panel, a rating system as described on their website is used and external revision took place (https://www.ncbi.nlm.nih.gov/clinvar/docs/assertion_criteria/).

The classification of the variant's pathogenicity is thus based on the interpretation of the submitter and thereby not consistent, although ClinVar encourages the use of the ACMG guidelines are encouraged. Furthermore, the submitter is able to add 'Evidence': a description of how variants were called and some context about the variant and its inheritance. The strength of ClinVar is also its downside: on the one hand, because everyone can submit their data, it is difficult to judge the reliability of a

ClinVar record. On the other hand however, because everyone can submit their data, ClinVar contains over 158,000 variant interpretations (Landrum et al. 2016).

Leiden Open Variant Database

The Leiden Open Variation Database (LOVD) is in many ways similar to ClinVar and contains over 155,000 variants identified in over 22,000 genes (July 31st, 2017), including ~20 IRD-associated genes (Cremers et al. 2014). The database contains variant information related to human phenotypes. The main differences between LOVDs and ClinVar is that the variants in the LOVD are curated by an expert who is responsible to check all variant records for a certain gene before they are uploaded to the database. Moreover, the variants are uploaded per patient. The latter allows one to look at the frequency of a variant within a patient group. Furthermore, there is space for the phenotype description, family information and a reference needs to be given in order for the user to look up more information (Fokkema et al. 2011).

Comprehensive gene-specific databases were set up for the Norrie disease pseudo glioma (*NDP*) gene, associated with X-linked Norrie disease and vitreoretinopathy [http://www.medmolgen.uzh.ch/research/eyediseases/norriedisease/Norrinmutations.html], and *CEP290*, associated with LCA, early-onset RP, and several syndromic retinopathies [http://www.LOVD.nl/CEP290; https://cep290base.cmgg.be/index.php]. A large proportion of variants in Bardet-Biedl syndrome (BBS)-associated genes can also be accessed online [https://lovd.euro-wabb.org/status.php]. Comprehensive Leiden Open Variation Databases (LOVDs) have been developed for 9 genes implicated in Usher syndrome, [https://grenada.lumc.nl/LOVD2/Usher_montpellier/USHbases.html] (Baux et al. 2014). Only for 10 other IRD-associated genes (*ABCA4,* www.LOVD.nl/ABCA4, (Cornelis et al. 2017); *AIPL1,* http://www.LOVD.nl/AIPL1; *CHM,* https://grenada.lumc.nl/LOVD2/Usher_montpellier/home.php?select_db=CHM; *CRB1* (http://www.LOVD.nl/CRB1 (Bujakowska et al. 2012); *EYS,* http://www.LOVD.nl/EYS (Messchaert et al. 2018); *LCA5,* http://www.LOVD.nl/LCA5 (Mackay et al. 2013); *RDH5,* http://www.LOVD.nl/RDH5; *RPE65*, http://www.LOVD.nl/RPE65; *SEMA4A,* http://www.LOVD.nl/SEM4A4; *TULP1,* http://www.LOVD.nl/TULP1), comprehensive LOVD databases have thus far been created. LOVD databases for any gene can be searched for using GeneSymbol.LOVD.nl.

Human Gene Mutation Database

The Human Gene Mutation Database (HGMD) is another database that contains variants that are found in patients and are likely disease causing. All the variants in the database have been published in literature and are manually curated. A subset of the variant data, i.e., 141,000 variants, is freely available for academic institutions and non-

profit organizations, although registration is needed. The pathogenicity classification system of HGMD is also five-tier based:

- Disease causing mutations (DM), when the author of the paper reporting the variant has established that the reported variant is involved in the phenotype
- Possible/probable disease causing mutations (DM?), when the author, the curators or other literature indicates that there is doubt about the pathogenicity of the variant
- Disease associated polymorphism (DP), when a significant association has been reported between the polymorphism and the disease.
- Functional polymorphism (FP), when a functional effect of the variant has been shown in research, but no disease association has been reported so far.
- Disease-associated polymorphisms with supporting functional evidence (DFP), when the polymorphism is both associated with disease and has been shown to have a functional effect.

There is also a commercial version of HGMD which contains over 203,000 variants. The information in HGMD gives an overview per variant with minimal information, including a reference to the literature. Interestingly, in 2013, Cassa et al. published a study indicating that HGMD likely contains variants that are erroneously described to be pathogenic: 4.6% of the variants in HGMD have an allele frequency >0.01 and 3.5% of the variants even have an allele frequency of >0.05. Furthermore, 8.5% of a large set of pathologic variants tested, were found in asymptomatic individuals (Cassa et al. 2013). These erroneously pathogenic classified variants are also likely to be present in ClinVar and LOVD, as those also contain variants from literature. Also, Abouelhoda et al. categorized many HGMD variants as benign by the use of a Saudi Arabian dataset of healthy consanguineous individuals, which show many homozygous rare variants (Abouelhoda et al. 2016). HGMD now removed ~1,000 'Retired records', which were erroneously included into HGMD. It is unclear whether these were removed from the public or the commercial version of HGMD (Stenson et al. 2017).

Multiple studies have shown that variant databases often do not agree on the classification of variants or report classifications based on too little supportive evidence. In 2013, Vail et al. reviewed the pathogenicity classification of 2017 variants in *BRCA1* and *BRCA2* by five different variant databases, including ClinVar, LOVD and HGMD. From variants present in two out of five databases, over 30% of the variants that were classified as pathogenic in one database, were not classified as such in the second database and the same for ~50% of the variants classified as Variant of Unknown Significance in one out of two and ~90% of the variants classified as benign in one out of two. When variants were present in more than two databases, the

percentage of discrepancy in most cases increased even more. Out of ClinVar, LOVD and HGMD, the one for which the pathogenic classifications of a subset of the variants was most often based on appropriate literature 40% of the cases was LOVD (Vail et al. 2015).

Future outlook

The identification of SNVs and SVs in IRD patients will become more straightforward as sequencing technologies keep improving, both in terms of sensitivity and cost-effectiveness. Long-read WGS represents the most attractive technology. Data storage costs will become more important as well as variant data interpretation.

Novel candidate IRD-associated genes are identified continuously but an increasing number has only been found to be involved in a single IRD patient or family. Within the ERDC, 130 IRD candidate genes have been identified, a small subset of which (www.erdc.info) is publicly available (S. Roosing and F.P.M. Cremers, unpublished data). World-wide sharing of genotype data however would increase the pace to find additional families with variants in the same gene, thereby strengthening their candidacy.

Modeling the identified mutations in animal models also is important to provide proof for the causality of gene defects. The entrance of CRISPR/Cas into science has opened great opportunities for IRD research (Peng et al. 2017). Apart from the modeling of gene defects in animals, novel functional assays need to be developed to assess the effect of variants at the RNA and protein levels. As mentioned above, in the absence of patients' somatic cells that express the gene of interest, robust *in vitro* RNA splice assays can be set-up for every human gene. In case retina-specific splicing defects could play a role, photoreceptor precursor cells can be derived from iPSCs generated from blood cells or fibroblasts. This technology has been used successfully to analyze the effect of known coding and non-coding variants in IRDs. Whether it also can be successfully used to identify elusive causal variants in large genomic regions or in entire genomes, remains to be seen.

Only a few examples of digenic inheritance and modifier genes for IRDs have been reported (Kajiwara et al. 1994; Beales et al. 2003; Vithana et al. 2003; Badano et al. 2006; Venturini et al. 2012; Liu et al. 2017). Nevertheless, there are many examples of significant differences between phenotypes (e.g., age at onset) in IRD cases that carry the same mutation(s), both within and between families. Reduced penetrance of variants might explain several autosomal dominant conditions, but as yet we have little clues regarding the genetic and possibly non-genetic modifiers. To study the mechanism of variable expression and non-penetrance, large case/control cohorts are required and genome-wide analysis techniques such as WES and WGS.

Acknowledgements

The work of M.K. is supported by the Rotterdamse Stichting Blindenbelangen, the Stichting Blindenhulp, the Stichting tot Verbetering van het Lot der Blinden, and the Stichting Blinden-Penning (to F.P.M.C and S.R.). The work of Z.F. is supported by the Foundation Fighting Blindness USA Project Program Award grant no. PPA-0517-0717-RAD (to F.P.M.C. and S.R.). The work of M.K. and S.C. is supported by the RP Fighting Blindness, UK, grant no. GR591 (to F.P.M.C.). The work of S.C. is supported by the Fighting Blindness Ireland (to F.P.M.C. and S.R.).

References

Abouelhoda M, Faquih T, El-Kalioby M, Alkuraya FS. 2016. Revisiting the morbid genome of Mendelian disorders. *Genome Biol* **17**: 235.

Absalan F, Ronaghi M. 2007. Molecular inversion probe assay. *Methods Mol Biol* **396**: 315-330.

Abu Safieh L, Aldahmesh MA, Shamseldin H, Hashem M, Shaheen R, Alkuraya H, Al Hazzaa SA, Al-Rajhi A, Alkuraya FS. 2010. Clinical and molecular characterisation of Bardet-Biedl syndrome in consanguineous populations: the power of homozygosity mapping. *J Med Genet* **47**: 236-241.

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**: 248-249.

Ajay SS, Parker SC, Abaan HO, Fajardo KV, Margulies EH. 2011. Accurate and comprehensive sequencing of personal genomes. *Genome Res* **21**: 1498-1505.

Al-Maghtheh M, Vithana E, Tarttelin E, Jay M, Evans K, Moore T, Bhattacharya S, Inglehearn CF. 1996. Evidence for a major retinitis pigmentosa locus on 19q13.4 (RP11) and association with a unique bimodal expressivity phenotype. *Am J Hum Genet* **59**: 864-871.

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

Alkuraya FS. 2013. The application of next-generation sequencing in the autozygosity mapping of human recessive diseases. *Hum Genet* **132**: 1197-1211.

Badano JL, Leitch CC, Ansley SJ, May-Simera H, Lawson S, Lewis RA, Beales PL, Dietz HC, Fisher S, Katsanis N. 2006. Dissection of epistasis in oligogenic Bardet-Biedl syndrome. *Nature* **439**: 326-330.

Bainbridge JW, Smith AJ, Barker SS, Robbie S, Henderson R, Balaggan K, Viswanathan A, Holder GE, Stockman A, Tyler N et al. 2008. Effect of gene therapy on visual function in Leber's congenital amaurosis. *N Engl J Med* **358**: 2231-2239.

Baux D, Blanchet C, Hamel C, Meunier I, Larrieu L, Faugere V, Vache C, Castorina P, Puech B, Bonneau D et al. 2014. Enrichment of LOVD-USHbases with 152 USH2A genotypes defines an extensive mutational spectrum and highlights missense hotspots. *Hum Mutat* **35**: 1179-1186.

Beales PL, Badano JL, Ross AJ, Ansley SJ, Hoskins BE, Kirsten B, Mein CA, Froguel P, Scambler PJ, Lewis RA et al. 2003. Genetic interaction of BBS1 mutations with alleles at other BBS loci can result in non-Mendelian Bardet-Biedl syndrome. *Am J Hum Genet* **72**: 1187-1199.

Beckmann JS, Estivill X, Antonarakis SE. 2007. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat Rev Genet* **8**: 639-646.

Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, Shang L, Boisson B, Casanova JL, Abel L. 2015. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A* **112**: 5473-5478.

Bentley DR. 2006. Whole-genome re-sequencing. *Curr Opin Genet Dev* **16**: 545-552.

Berger W, Kloeckener-Gruissem B, Neidhardt J. 2010. The molecular basis of human retinal and vitreoretinal diseases. *Prog Retin Eye Res* **29**: 335-375.

Boonstra FN, van Nouhuys CE, Schuil J, de Wijs IJ, van der Donk KP, Nikopoulos K, Mukhopadhyay A, Scheffer H, Tilanus MA, Cremers FP et al. 2009. Clinical and molecular evaluation of probands and family members with familial exudative vitreoretinopathy. *Invest Ophthalmol Vis Sci* **50**: 4379-4385.

Broadgate S, Yu J, Downes SM, Halford S. 2017. Unravelling the genetics of inherited retinal dystrophies: Past, present and future. *Prog Retin Eye Res* **59**: 53-96.

Buchanan JA, Scherer SW. 2008. Contemplating effects of genomic structural variation. *Genet Med* **10**: 639-647.

Bujakowska K, Audo I, Mohand-Said S, Lancelot ME, Antonio A, Germain A, Leveillard T, Letexier M, Saraiva JP, Lonjou C et al. 2012. CRB1 mutations in inherited retinal dystrophies. *Hum Mutat* **33**: 306-315.

Bujakowska KM, Fernandez-Godino R, Place E, Consugar M, Navarro-Gomez D, White J, Bedoukian EC, Zhu X, Xie HM, Gai X et al. 2017. Copy-number variation is an important contributor to the genetic causality of inherited retinal degenerations. *Genet Med* **19**: 643-651.

Carneiro MO, Russ C, Ross MG, Gabriel SB, Nusbaum C, DePristo MA. 2012. Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC genomics* **13**: 375.

Cassa CA, Tong MY, Jordan DM. 2013. Large numbers of genetic variants considered to be pathogenic are common in asymptomatic individuals. *Hum Mutat* **34**: 1216-1220.

Chen W, Hayward C, Wright AF, Hicks AA, Vitart V, Knott S, Wild SH, Pramstaller PP, Wilson JF, Rudan I et al. 2011. Copy number variation across European populations. *PLoS One* **6**: e23087.

Chesworth BM, Hamilton CB, Walton DM, Benoit M, Blake TA, Bredy H, Burns C, Chan L, Frey E, Gillies G et al. 2014. Reliability and validity of two versions of the upper extremity functional index. *Physiother Can* **66**: 243-253.

Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE et al. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* **10**: 563-569.

Collin RW, Littink KW, Klevering BJ, van den Born LI, Koenekoop RK, Zonneveld MN, Blokland EA, Strom TM, Hoyng CB, den Hollander AI et al. 2008. Identification of a 2 Mb human ortholog of Drosophila eyes shut/spacemaker that is mutated in patients with retinitis pigmentosa. *Am J Hum Genet* **83**: 594-603.

Collin RW, van den Born LI, Klevering BJ, de Castro-Miro M, Littink KW, Arimadyo K, Azam M, Yazar V, Zonneveld MN, Paun CC et al. 2011. High-resolution homozygosity mapping is a powerful tool to detect novel mutations causative

of autosomal recessive RP in the Dutch population. *Invest Ophthalmol Vis Sci* **52**: 2227-2239.

Combs R, McAllister M, Payne K, Lowndes J, Devery S, Webster AR, Downes SM, Moore AT, Ramsden S, Black G et al. 2013. Understanding the impact of genetic testing for inherited retinal dystrophy. *Eur J Hum Genet* **21**: 1209-1213.

Conrad DF, Hurles ME. 2007. The population genetics of structural variation. *Nat Genet* **39**: S30-36.

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P et al. 2010. Origins and functional impact of copy number variation in the human genome. *Nature* **464**: 704-712.

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

Cremers FP, den Dunnen JT, Ajmal M, Hussain A, Preising MN, Daiger SP, Qamar R. 2014. Comprehensive registration of DNA sequence variants associated with inherited retinal diseases in Leiden Open Variation Databases. *Hum Mutat* **35**: 147-148.

Cremers FP, van de Pol DJ, van Kerkhoff LP, Wieringa B, Ropers HH. 1990. Cloning of a gene that is rearranged in patients with choroideraemia. *Nature* **347**: 674-677.

den Hollander AI, Black A, Bennett J, Cremers FP. 2010. Lighting a candle in the dark: advances in genetics and gene therapy of recessive retinal dystrophies. *J Clin Invest* **120**: 3042-3053.

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**: e67.

Dryja TP, McGee TL, Reichel E, Hahn LB, Cowley GS, Yandell DW, Sandberg MA, Berson EL. 1990. A point mutation of the rhodopsin gene in one form of retinitis pigmentosa. *Nature* **343**: 364-366.

Edwards TL, Jolly JK, Groppe M, Barnard AR, Cottriall CL, Tolmachova T, Black GC, Webster AR, Lotery AJ, Holder GE et al. 2016. Visual Acuity after Retinal Gene Therapy for Choroideremia. *N Engl J Med* **374**: 1996-1998.

Eisenberger T, Neuhaus C, Khan AO, Decker C, Preising MN, Friedburg C, Bieg A, Gliem M, Charbel Issa P, Holz FG et al. 2013. Increasing the yield in targeted next-generation sequencing by implicating CNV analysis, non-coding exons and the overall variant load: the example of retinal dystrophies. *PLoS One* **8**: e78496.

Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. 2011. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat* **32**: 557-563.

Fukunaga R, Matsumoto T, Aoyagi Y, Matsuda D, Tanaka S, Okadome J, Morisaki K, Maehara Y. 2014. Thoracic stent graft with distal fenestration for the superior mesenteric artery for treatment of thoracic aortic aneurysm. *Ann Vasc Dis* **7**: 152-155.

Genome of the Netherlands C. 2014. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet* **46**: 818-825.

Glockle N, Kohl S, Mohr J, Scheurenbrand T, Sprecher A, Weisschuh N, Bernd A, Rudolph G, Schubach M, Poloschek C et al. 2014. Panel-based next generation sequencing as a reliable and efficient technique to detect mutations in unselected patients with retinal dystrophies. *Eur J Hum Genet* **22**: 99-104.

Goodwin S, McPherson JD, McCombie WR. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* **17**: 333-351.

Haer-Wigman L, van Zelst-Stams WA, Pfundt R, van den Born LI, Klaver CC, Verheij JB, Hoyng CB, Breuning MH, Boon CJ, Kievit AJ et al. 2017. Diagnostic exome sequencing in 266 Dutch patients with visual impairment. *Eur J Hum Genet* **25**: 591-599.

Hauswirth WW, Aleman TS, Kaushal S, Cideciyan AV, Schwartz SB, Wang L, Conlon TJ, Boye SL, Flotte TR, Byrne BJ et al. 2008. Treatment of leber congenital amaurosis due to RPE65 mutations by ocular subretinal injection of adeno-associated virus gene vector: short-term results of a phase I trial. *Human gene therapy* **19**: 979-990.

Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, Ordoukhanian P. 2014. Library construction for next-generation sequencing: overviews and challenges. *Biotechniques* **56**: 61-64, 66, 68, passim.

Hehir-Kwa JY, Pfundt R, Veltman JA. 2015. Exome sequencing and whole genome sequencing for the detection of copy number variation. *Expert Rev Mol Diagn* **15**: 1023-1032.

Hicks MJ, Lam BJ, Hertel KJ. 2005. Analyzing mechanisms of alternative pre-mRNA splicing using in vitro splicing assays. *Methods (San Diego, Calif)* **37**: 306-313.

Hoffmann K, Lindner TH. 2005. easyLINKAGE-Plus--automated linkage analyses using large-scale SNP data. *Bioinformatics* **21**: 3565-3567.

Igartua C, Turner EH, Ng SB, Hodges E, Hannon GJ, Bhattacharjee A, Rieder MJ, Nickerson DA, Shendure J. 2010. Targeted enrichment of specific regions in the human genome by array hybridization. *Curr Protoc Hum Genet* **Chapter 18**: Unit 18 13.

Jacob CO, Reiff A, Armstrong DL, Myones BL, Silverman E, Klein-Gitelman M, McCurdy D, Wagner-Weiner L, Nocton JJ, Solomon A et al. 2007. Identification of novel susceptibility genes in childhood-onset systemic lupus erythematosus using a uniquely designed candidate gene pathway platform. *Arthritis Rheum* **56**: 4164-4173.

Kajiwara K, Berson EL, Dryja TP. 1994. Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. *Science* **264**: 1604-1608.

Knoppers BM, Zawati MH, Senecal K. 2015. Return of genetic testing results in the era of whole-genome sequencing. *Nat Rev Genet* **16**: 553-559.

Koenekoop RK, Sui R, Sallum J, van den Born LI, Ajlan R, Khan A, den Hollander AI, Cremers FP, Mendola JD, Bittner AK et al. 2014. Oral 9-cis retinoid for childhood blindness due to Leber congenital amaurosis caused by RPE65 or LRAT mutations: an open-label phase 1b trial. *Lancet* **384**: 1513-1520.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**: 1073-1081.

Lander E, Kruglyak L. 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* **11**: 241-247.

Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J et al. 2016. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* **44**: D862-868.

Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G et al. 2007. The diploid genome sequence of an individual human. *PLoS Biol* **5**: e254.

Levy SE, Myers RM. 2016. Advancements in Next-Generation Sequencing. *Annu Rev Genomics Hum Genet* **17**: 95-115.

Lin X, Tang W, Ahmad S, Lu J, Colby CC, Zhu J, Yu Q. 2012. Applications of targeted gene capture and next-generation sequencing technologies in studies of human deafness and other genetic disabilities. *Hearing research* **288**: 67-76.

Liu YP, Bosch DG, Siemiatkowska AM, Rendtorff ND, Boonstra FN, Moller C, Tranebjaerg L, Katsanis N, Cremers FP. 2017. Putative digenic inheritance of heterozygous RP1L1 and C2orf71 null mutations in syndromic retinal dystrophy. *Ophthalmic genetics* **38**: 127-132.

Lukovic D, Artero Castro A, Delgado AB, Bernal Mde L, Luna Pelaez N, Diez Lloret A, Perez Espejo R, Kamenarova K, Fernandez Sanchez L, Cuenca N et al. 2015. Human iPSC derived disease model of MERTK-associated retinitis pigmentosa. *Scientific reports* **5**: 12910.

Mackay DS, Borman AD, Sui R, van den Born LI, Berson EL, Ocaka LA, Davidson AE, Heckenlively JR, Branham K, Ren H et al. 2013. Screening of a large cohort of leber congenital amaurosis and retinitis pigmentosa patients identifies novel LCA5 mutations and new genotype-phenotype correlations. *Hum Mutat* **34**: 1537-1546.

MacLaren RE, Groppe M, Barnard AR, Cottriall CL, Tolmachova T, Seymour L, Clark KR, During MJ, Cremers FP, Black GC et al. 2014. Retinal gene therapy in patients with choroideremia: initial findings from a phase 1/2 clinical trial. *Lancet* **383**: 1129-1137.

Maguire AM, Simonelli F, Pierce EA, Pugh EN, Jr., Mingozzi F, Bennicelli J, Banfi S, Marshall KA, Testa F, Surace EM et al. 2008. Safety and efficacy of gene transfer for Leber's congenital amaurosis. *N Engl J Med* **358**: 2240-2248.

Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ. 2010. Target-enrichment strategies for next-generation sequencing. *Nat Methods* **7**: 111-118.

Mardis ER. 2011. A decade's perspective on DNA sequencing technology. *Nature* **470**: 198-203.

McCoy RC, Taylor RW, Blauwkamp TA, Kelley JL, Kertesz M, Pushkarev D, Petrov DA, Fiston-Lavier AS. 2014. Illumina TruSeq synthetic long-reads empower de novo assembly and resolve complex, highly-repetitive transposable elements. *PLoS One* **9**: e106689.

McWilliam P, Farrar GJ, Kenna P, Bradley DG, Humphries MM, Sharp EM, McConnell DJ, Lawler M, Sheils D, Ryan C et al. 1989. Autosomal dominant retinitis

pigmentosa (ADRP): localization of an ADRP gene to the long arm of chromosome 3. *Genomics* **5**: 619-622.

Meienberg J, Bruggmann R, Oexle K, Matyas G. 2016. Clinical sequencing: is WGS the better WES? *Hum Genet* **135**: 359-362.

Messchaert M, Haer-Wigman L, Khan MI, Cremers FPM, Collin RWJ. 2018. EYS mutation update: In silico assessment of 271 reported and 26 novel variants in patients with retinitis pigmentosa. *Hum Mutat* **39**: 177-186.

Mitchell GA, Brody LC, Looney J, Steel G, Suchanek M, Dowling C, Der Kaloustian V, Kaiser-Kupfer M, Valle D. 1988. An initiator codon mutation in ornithine-delta-aminotransferase causing gyrate atrophy of the choroid and retina. *J Clin Invest* **81**: 630-633.

Movassat M, Mueller WF, Hertel KJ. 2014. In vitro assay of pre-mRNA splicing in mammalian nuclear extract. *Methods Mol Biol* **1126**: 151-160.

Myllykangas S, Natsoulis G, Bell JM, Ji HP. 2011. Targeted sequencing library preparation by genomic DNA circularization. *BMC Biotechnol* **11**: 122.

Neveling K, den Hollander AI, Cremers FP, Collin RW. 2013. Identification and analysis of inherited retinal disease genes. *Methods Mol Biol* **935**: 3-23.

Newman S, Hermetz KE, Weckselblatt B, Rudd MK. 2015. Next-generation sequencing of duplication CNVs reveals that most are tandem and some create fusion genes at breakpoints. *Am J Hum Genet* **96**: 208-220.

Nyholt D. 2008. Statistical genetics: gene mapping through linkage and association. . In *Principles of linkage analysis* (ed. FM Neale BM, Medland SE, Posthuma D (eds)), pp. pp 113-134. Taylor & Francis Group, New York.

O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, Carvill G, Kumar A, Lee C, Ankenman K et al. 2012. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338**: 1619-1622.

Osoegawa K, de Jong PJ. 2004. BAC library construction. *Methods Mol Biol* **255**: 1-46.

Pang AW, MacDonald JR, Pinto D, Wei J, Rafiq MA, Conrad DF, Park H, Hurles ME, Lee C, Venter JC et al. 2010. Towards a comprehensive structural variation map of an individual human genome. *Genome Biol* **11**: R52.

Pang AW, Macdonald JR, Yuen RK, Hayes VM, Scherer SW. 2014. Performance of high-throughput sequencing for the discovery of genetic variation across the complete size spectrum. *G3 (Bethesda)* **4**: 63-65.

Parfitt DA, Lane A, Ramsden C, Jovanovic K, Coffey PJ, Hardcastle AJ, Cheetham ME. 2016. Using induced pluripotent stem cells to understand retinal ciliopathy disease mechanisms and develop therapies. *Biochem Soc Trans* **44**: 1245-1251.

Pasadhika S, Fishman GA, Stone EM, Lindeman M, Zelkha R, Lopez I, Koenekoop RK, Shahidi M. 2010. Differential macular morphology in patients with RPE65-, CEP290-, GUCY2D-, and AIPL1-related Leber congenital amaurosis. *Invest Ophthalmol Vis Sci* **51**: 2608-2614.

Peng YQ, Tang LS, Yoshida S, Zhou YD. 2017. Applications of CRISPR/Cas9 in retinal degenerative diseases. *International journal of ophthalmology* **10**: 646-651.

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

Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, Lionel AC, Thiruvahindrapuram B, Macdonald JR, Mills R et al. 2011. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat Biotechnol* **29**: 512-520.

Pirooznia M, Goes FS, Zandi PP. 2015. Whole-genome CNV analysis: advances in computational approaches. *Front Genet* **6**: 138.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**: 559-575.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E et al. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**: 405-424.

Roosing S, Thiadens AA, Hoyng CB, Klaver CC, den Hollander AI, Cremers FP. 2014. Causes and consequences of inherited cone disorders. *Prog Retin Eye Res* **42**: 1-26.

Ruschendorf F, Nurnberg P. 2005. ALOHOMORA: a tool for linkage analysis using 10K SNP array data. *Bioinformatics* **21**: 2123-2125.

Russell S, Bennett J, Wellman JA, Chung DC, Yu ZF, Tillman A, Wittes J, Pappas J, Elci O, McCague S et al. 2017. Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *Lancet* **390**: 849-860.

Salmela L, Walve R, Rivals E, Ukkonen E. 2017. Accurate self-correction of errors in long reads using de Bruijn graphs. *Bioinformatics* **33**: 799-806.

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

Sangermano R, Khan M, Cornelis SS, Richelle V, Albert S, Garanto A, Elmelik D, Qamar R, Lugtenberg D, van den Born LI et al. 2018. ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Res* **28**: 100-110.

Scholl HP, Moore AT, Koenekoop RK, Wen Y, Fishman GA, van den Born LI, Bittner A, Bowles K, Fletcher EC, Collison FT et al. 2015. Safety and Proof-of-Concept Study of Oral QLT091001 in Retinitis Pigmentosa Due to Inherited Deficiencies of Retinal Pigment Epithelial 65 Protein (RPE65) or Lecithin:Retinol Acyltransferase (LRAT). *PLoS One* **10**: e0143846.

Schwarz JM, Cooper DN, Schuelke M, Seelow D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* **11**: 361-362.

Shendure J, Mitra RD, Varma C, Church GM. 2004. Advanced sequencing technologies: methods and goals. *Nat Rev Genet* **5**: 335-344.

Siemiatkowska AM, Collin RW, den Hollander AI, Cremers FP. 2014. Genomic approaches for the discovery of genes mutated in inherited retinal degeneration. *Cold Spring Harb Perspect Med* **4**.

Sobreira N, Schiettecatte F, Valle D, Hamosh A. 2015. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum Mutat* **36**: 928-930.

Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, Hussain M, Phillips AD, Cooper DN. 2017. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet* **136**: 665-677.

Teer JK, Bonnycastle LL, Chines PS, Hansen NF, Aoyama N, Swift AJ, Abaan HO, Albert TJ, Program NCS, Margulies EH et al. 2010. Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. *Genome Res* **20**: 1420-1431.

Terwillinger D, Ott J. 1994. *Handbook for human genetic linkage.* Johns Hopkins University Press, Baltimore.

Tewhey R, Warner JB, Nakano M, Libby B, Medkova M, David PH, Kotsopoulos SK, Samuels ML, Hutchison JB, Larson JW et al. 2009. Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nat Biotechnol* **27**: 1025-1031.

Tucker BA, Mullins RF, Streb LM, Anfinson K, Eyestone ME, Kaalberg E, Riker MJ, Drack AV, Braun TA, Stone EM. 2013. Patient-specific iPSC-derived photoreceptor precursor cells as a means to investigate retinitis pigmentosa. *Elife* **2**: e00824.

Turner EH, Lee C, Ng SB, Nickerson DA, Shendure J. 2009. Massively parallel exon capture and library-free resequencing across 16 genomes. *Nat Methods* **6**: 315-316.

Vail PJ, Morris B, van Kan A, Burdett BC, Moyes K, Theisen A, Kerr ID, Wenstrup RJ, Eggington JM. 2015. Comparison of locus-specific databases for BRCA1 and BRCA2 variants reveals disparity in variant classification within and among databases. *J Community Genet* **6**: 351-359.

Valle D, Kaiser-Kupfer MI, Del Valle LA. 1977. Gyrate atrophy of the choroid and retina: deficiency of ornithine aminotransferase in transformed lymphocytes. *Proc Natl Acad Sci U S A* **74**: 5159-5161.

Venturini G, Rose AM, Shah AZ, Bhattacharya SS, Rivolta C. 2012. CNOT3 is a modifier of PRPF31 mutations in retinitis pigmentosa with incomplete penetrance. *PLoS genetics* **8**: e1003040.

Vithana EN, Abu-Safieh L, Pelosini L, Winchester E, Hornan D, Bird AC, Hunt DM, Bustin SA, Bhattacharya SS. 2003. Expression of PRPF31 mRNA in patients with autosomal dominant retinitis pigmentosa: a molecular clue for incomplete penetrance? *Invest Ophthalmol Vis Sci* **44**: 4204-4209.

Whiteford N, Haslam N, Weber G, Prugel-Bennett A, Essex JW, Roach PL, Bradley M, Neylon C. 2005. An analysis of the feasibility of short read sequencing. *Nucleic Acids Res* **33**: e171.

Woods CG, Cox J, Springell K, Hampshire DJ, Mohamed MD, McKibbin M, Stern R, Raymond FL, Sandford R, Malik Sharif S et al. 2006. Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. *Am J Hum Genet* **78**: 889-896.

Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* **11**: 377-394.

Yoshida T, Ozawa Y, Suzuki K, Yuki K, Ohyama M, Akamatsu W, Matsuzaki Y, Shimmura S, Mitani K, Tsubota K et al. 2014. The use of induced pluripotent stem cells to reveal pathogenic gene mutations and explore treatments for retinitis pigmentosa. *Molecular brain* **7**: 45.

Zarrei M, MacDonald JR, Merico D, Scherer SW. 2015. A copy number variation map of the human genome. *Nat Rev Genet* **16**: 172-183.

Chapter 1.3

ABCA4-associated Stargardt disease

ABCA4-associated Stargardt disease

Mubeen Khan and Frans P.M. Cremers

Adapted from: ***Klin Monatsbl Augenheilkd.*** 2020**; 237:** 267-274.

Abstract

Autosomal recessive Stargardt disease (STGD1) is associated with variants in the *ABCA4* gene. The phenotypes range from early-onset STGD1, clinically resembling severe cone-rod dystrophy, to intermediate STGD1 and late-onset STGD1. These different phenotypes can be correlated with different combinations of *ABCA4* variants which can be classified according to their degree of severity. A significant part of STGD1 cases, in particular late-onset STGD1 cases, were shown to carry only a single *ABCA4* variant. Recently, a frequent coding variant (p.Asn1868Ile) was found which, when in combination with a severe *ABCA4* variant, was generally associated with late-onset STGD1. In addition, an increasing number of rare deep-intronic variants were found some of which also are associated with late-onset STGD1. The effect of these and other variants on the *ABCA4* RNA could be tested using *in vitro* assays in human kidney cells using specially designed midigenes. Employing stem cells and photoreceptor progenitor cells that can be derived from patient skin or blood cells, allowed the assessment of retina-specific splicing defects. Expert clinical examination to distinguish STGD1 cases from other maculopathies, as well as in-depth genomics and transcriptomics data currently allow the identification of both mutant *ABCA4* alleles in >95% of the cases.

Spectrum of *ABCA4*-associated phenotypes

Stargardt disease due to variants in the ATP-binding cassette transporter type A4 (*ABCA4*) gene represents the most frequent juvenile macular dystrophy (Allikmets et al. 1997; Cremers et al. 1998). Moreover, with an estimated prevalence of 1/10,000, it is the most frequent inherited retinal disease due to variants in one gene and accounts for 7% of all retinal dystrophies (Blacharski 1988; Simonelli et al. 2000; Tanna et al. 2017). Despite being a monogenic disease, there is a substantial heterogeneity in age of onset and clinical representation of disease, which is mainly due to different combination of mutations (Valkenburg et al. 2019). The most severe phenotypes, early-onset STGD1, is observed before the age of 10. It also has been described as a severe autosomal recessive cone-rod dystrophy (arCRD). Intermediate or 'typical' STGD1 shows an onset between 10 and 45 yrs of age. It is characterized by reduced central visual acuity, delayed dark adaptation and formation of fish-tailed yellowish flecks in the retinal pigment epithelium (RPE), hence the alternative clinical name 'fundus flavimaculatus'. Many autofluorescence studies also revealed the macular atrophy. In late-onset STGD1 (onset >44 yrs), macular atrophy is accompanied by a ring of hyper autofluorescence just outside the atrophic region. Foveal sparing is often observed which explains why affected individuals have relatively good visual acuity until late in the disease process.

ABCA4 protein structure and function

ABCA4, also known as ABCR or rim protein, is a member of the ABCA subfamily (Allikmets et al. 1997). ABCA4 is a transmembrane transporter composed of 2,273 amino-acid (Bungert et al. 2001), organized in two tandem-arranged halves, each comprised of a transmembrane segment followed by a large exocytoplasmic (extracellular/lumen) domain, and a nucleotide binding domain in both halves of the protein (**Figure 1A**). *ABCA4* is almost exclusively expressed in the outer segments of rod and cone photoreceptor cells (Sun and Nathans 1997; Molday et al. 2000). Limited expression has also been found in keratinocytes, fibroblasts and in human and murine RPE cells (Wiley et al. 2016; Aukrust et al. 2016; Lenis et al. 2017). As depicted in **Figure 1B**, it plays a crucial role in the clearance of all-*trans*-retinal and potentially toxic by-products such as di-retinoid-pyridinium-ethanolamine (A2E) of the visual cycle from the inner leaflet of the outer segment disks to the photoreceptor cells cytoplasm upon which it is transferred to the RPE. Mutant ABCA4 protein can no longer remove the toxic substances efficiently from the photoreceptors cells hence leading to the accumulation of all-*trans*-retinal and its derivatives in the form of lipofuscin. This can block the effective diffusion of oxygen and carbon dioxide to and from the outer part of the neural retina, and can result in the death of RPE cells and the adjacent photoreceptors.

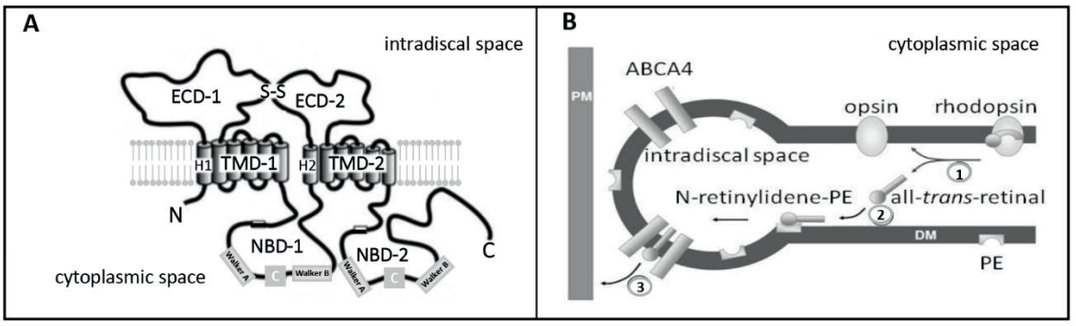


Figure 1. Protein domain structure and function of ABCA4. **A.** Topological model for the ABCA4 protein. (ECD) exocytoplasmic domain; (H1 or H2) first transmembrane segment; (TMD) transmembrane domain; (NBD) nucleotide binding domain. The two Walker motifs (A and B) and the C region which lies between them are the hallmark of ABC transporter superfamily members. **B.** Proposed function of the ABCA4 protein. 1. All-*trans*-retinal is released by the photoactivated rhodopsin. 2. All-*trans*-retinal enters the hydrophobic disc membrane (DM) and reacts to form a Schiff base with the amine of a phosphatidylethanolamine (PE), a phospholipid found in the DM outer segment inner leaflet. 3. All-*trans*-retinal or the N-retinylidene-PE complex is transported by ABCA4 to the cytoplasmic side. All-*trans*-retinal is then reduced by all-*trans*-retinol dehydrogenase and transported to the RPE (not shown). (PM) plasma membrane. Figure adapted with permission from Bungert et al. 2001.

STGD-like phenotypes due to variants in other genes

STGD1 is inherited in an autosomal recessive manner and only known to be caused by *ABCA4* mutations. However, there are other clinically overlapping phenotypes, inherited both in autosomal dominant (ad) and recessive (ar) manner. Mutations in *ELOVL4* and *PROM1* lead to adSTGD-like phenotype and are known as STGD3 (MIM# 600110) and STGD4 (MIM# 603786), respectively (Imani et al. 2018). Similarly, heterozygous mutations in *PRPH2, BEST1* and *CRX,* and bi-allelic variants in *CRB1, RAB28, RDH11* and *ROM1* are also known to cause STGD-like phenotypes (Apfelstedt-Sylla et al. 1995; Xie et al. 2014; Riveiro-Alvarez et al. 2015; Zaneveld et al. 2015; Zernant et al. 2017; Ma et al. 2019).

Genotype-phenotype correlation

A genotype-phenotype correlation model was established based on two observations. First, two severe *ABCA4* variants were associated with early-onset STGD1 (Cremers et al. 1998; Maugeri et al. 2000), and secondly, some frequent mild variants in the general population were only found in STGD1 cases in combination with severe variants (Maugeri et al. 1999). Thus, the combination of a mild and a severe variant or the presence of two moderately severe variants led to an intermediate STGD1 phenotype with moderate ABCA4 impairment (**Figure 2**). Early-onset STGD1 cases however, also described as arCRD and arRP, were proposed to be due to combinations of moderately severe and severe, or two severe variants, respectively. The monogenic nature of STGD1 and the typical features displayed by STGD1 cases has resulted in the

genotyping of a large number of cases. More than 1,000 variants have been associated with STGD1. Cornelis et al. (2017) collected *ABCA4* variants data of 3,928 reported STGD1 cases and performed *in silico* functional assessments on variants that did not result in a clear loss of ABCA4 function (stop, frameshift and canonical splice site variants) (Cornelis et al. 2017). Among a total of 849 different variants, the majority were missense, i.e., 471 variants, followed by 314 protein truncating variants (which includes large deletions), 49 noncanonical splice site (NCSS) variants, 5 synonymous and 10 deep-intronic (DI) variants. Five variants were categorized as 'mild' alleles, based on their frequent presence in control individuals and their underrepresentation in a homozygous state in STGD1 cases. Since then, 7 additional causal DI variants were reported (Bauwens et al. 2019, **Chapter 3**). To date, a total of 1,175 unique *ABCA4* variants have been identified in 6,675 cases (www.lovd.nl/ABCA4).

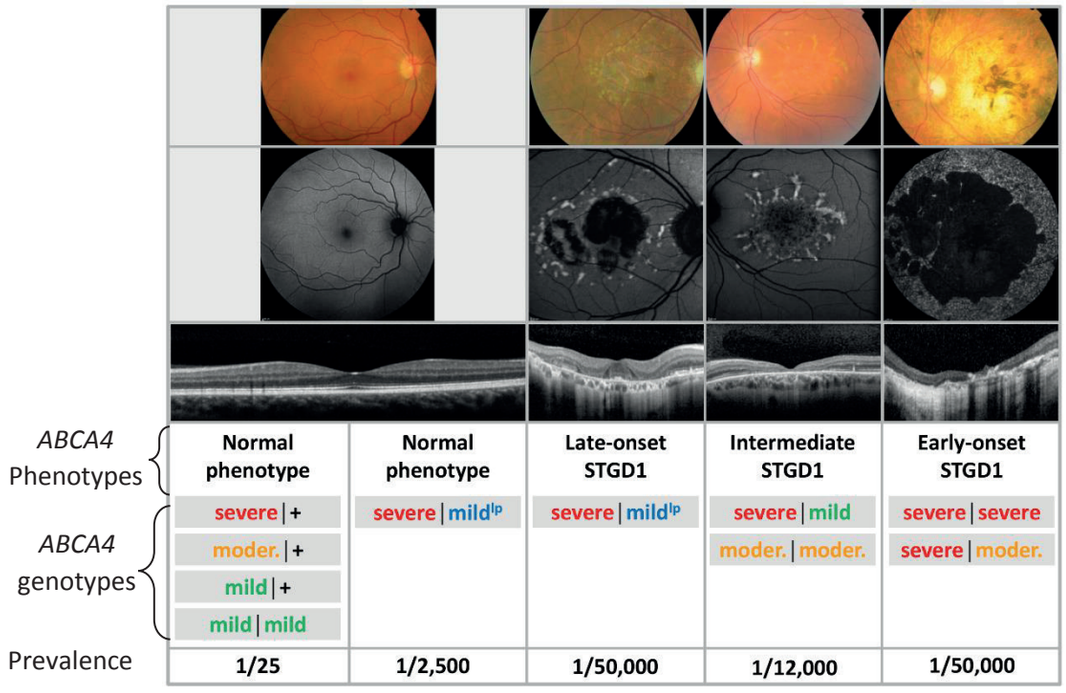


Figure 2. Genotype-phenotype correlation for *ABCA4*-associated STGD1. Upper panels contain fundus images, fundus autofluorescence images and optical coherence tomography of a healthy individual and STGD1 cases. Left panel: heterozygous carriers of any *ABCA4* variant and bi-allelic persons carrying two mild *ABCA4* variants do not show a STGD1 phenotype. Persons carrying a severe variant and c.5603A>T (p.Asn1868Ile) in ~95% of the cases have no STGD1 and in 5% of the cases show late-onset STGD1, often with foveal sparing (middle panel). Persons carrying a combination of a severe and (fully penetrant) mild variant show intermediate STGD1 and persons carrying two severe variants or a combination of a severe with a moderately severe variant show early-onset STGD1.

p.Asn1868Ile: a frequent low-penetrant *ABCA4* variant

The frequent c.5603A>T (p.Asn1868Ile) variant, present in 6.6% of alleles in the non-Finnish European database (Genome Aggregation Database - gnomAD,

http://gnomad.broadinstitute.org/), was found to be enriched in a few patient cohorts (Aguirre-Lamban et al. 2011; Schulz et al. 2017). This variant explained ~50% of mono-allelic and ~90% of late-onset STGD1 cases, when present in a compound heterozygous manner with a severe *ABCA4* variant (Zernant et al. 2017). Similarly, in a Dutch cohort of mono-allelic cases each carrying a severe *ABCA4* variant on one gene copy, p.Asn1868Ile was found to occur in *trans*, i.e., on the second gene copy, in ~40% of the probands (Runhart et al. 2018). It thereby occurs in ~10% of all STGD1 probands. In addition, it was shown that this particular variant, when present in *cis*, i.e., on the same gene copy, together with c.2588G>C (c.2588G>C; 5603A>T) exerts a mild effect on ABCA4 protein function, whereas a single c.2588G>C (p.[Gly863Ala,Gly863del]) variant was thought to be benign. Moreover, based on the p.Asn1868Ile frequency in the Netherlands, the sum frequency of all severe *ABCA4* variants in 23,000 Dutch controls, the number of expected and observed STGD1 cases in the Netherlands, we estimate the penetrance of the combination of a severe allele with p.Asn1868Ile to be ~5%. In families with at least two bi-allelic cases carrying p.Asn1868Ile, 7/11 (64%) of non-probands were affected. This difference in part may be explained by as yet unknown intronic variants present in STGD1 cases that are in *cis* with p.Asn1868Ile, thereby rendering them fully penetrant alleles. This very low penetrance may also depend on genetic modifiers in different genomic loci, or other non-genetic factors that can have a direct or indirect impact on *ABCA4* expression or ABCA4 function. Next to variant c.2588G>C, other variants have been found to be present in *cis* with c.5603A>T, most notable c.5461-10T>C, a frequent NCSS variant which results in skipping of exon 39 or exons 39/40, and a DI variant (c.769-784C>T) that results in the generation of a 162-nt pseudoexon (PE) in part of the mRNA (Sangermano et al. 2016; **Chapter 3**).

Mendelian and atypical inheritance patterns for STGD1

Most of the *ABCA4* variants show a classical autosomal recessive pattern of inheritance, with unaffected parents and unaffected offspring of STGD1 probands (**Figure 3A**). The relatively frequent occurrence of *ABCA4* variants in the general population (estimated sum carrier frequency: 0.04, excluding c.5603A>T) however can result in pseudo-dominant inheritance (**Figure 3B**) (Cremers et al. 1998; Shroyer et al. 2000; Klevering et al. 2004; Bax et al. 2015), or the occurrence of 2nd or 3rd degree relatives that are also affected by STGD1. The fact that two mild variants generally do not result in STGD1 means that unaffected parents in rare cases carry two (often different) mild variants and that different combinations of variants can be observed in affected offspring of the same family if the other parent is heterozygous carrier for a *ABCA4* variant (**Figure 3C**). Finally, due to the very low penetrance of at least one frequent variant, c.5603A>T (p.Asn1868Ile), siblings carrying the same combination of

ABCA4 variants may show very different expression of the disease, some of which do not show STGD1, even late in life (**Figure 3D**).

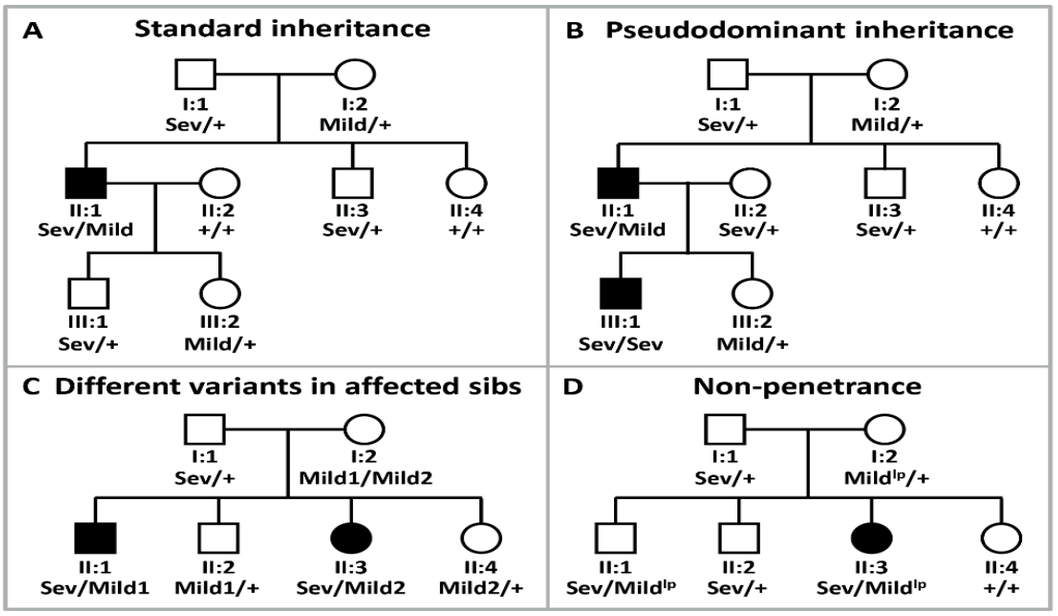


Figure 3. Pedigrees observed with STGD1 cases. A. Most STGD1 cases are found as single cases in pedigrees. **B.** In view of the high carrier frequency of ABCA4 variants in the general population (0.04), pseudo-dominant inheritance patterns are regularly observed. In this example spouse II:2 carries a severe variant and III:1 inherited severe variants from both parents. Two severe variants will result in early-onset STGD1 or severe cone-rod dystrophy, whereas a combination of a severe variant with a mild variant in most cases results in intermediate STGD1. **C.** In rare families, one of the unaffected parents carries two different mild variants and the offspring can inherit one of these two. **D.** Combinations of severe variants with a mild-low penetrant variant (c.5603A>T), in a few families resulted in affected and non-affected bi-allelic persons.

Splice site variants and splicing defects

As depicted in **Figure 4**, correct splicing of the pre-mRNA to the mRNA requires *cis* sequence elements, located in the pre-mRNA itself, and *trans* factors that can bind to these elements (Singh and Cooper 2012). Cleavage of the introns occurs mainly through the recognition of the highly conserved nucleotide sequences located at the 5' end of each exon (splice acceptor site; SAS) and at the 3' end of each exon (splice donor site; SDS) and a branch point site (BPS), a short conserved sequence containing an adenine residue between 10 and 50 nucleotides upstream of the SAS, necessary for the formation of the lariat structure (Gao et al. 2008). The SAS dinucleotide AG, located at the last two positions of the intron (i.e., positions -1 and -2 compared to the downstream exon) is 100% conserved. At positions -14 to -5 pyrimidines are preferred over purines. At -3 there is a preference for a cytosine and

the first exonic nucleotide often is a guanine. At the SDS, the first two nucleotides of the intron are GT (in RNA GU), with a few exceptions in which this is GC (e.g. exons 46 and 48 in *ABCA4*). The last nucleotide of the exon often is a guanine and the positions +3 to +6 in the DNA preferably are AAGT.

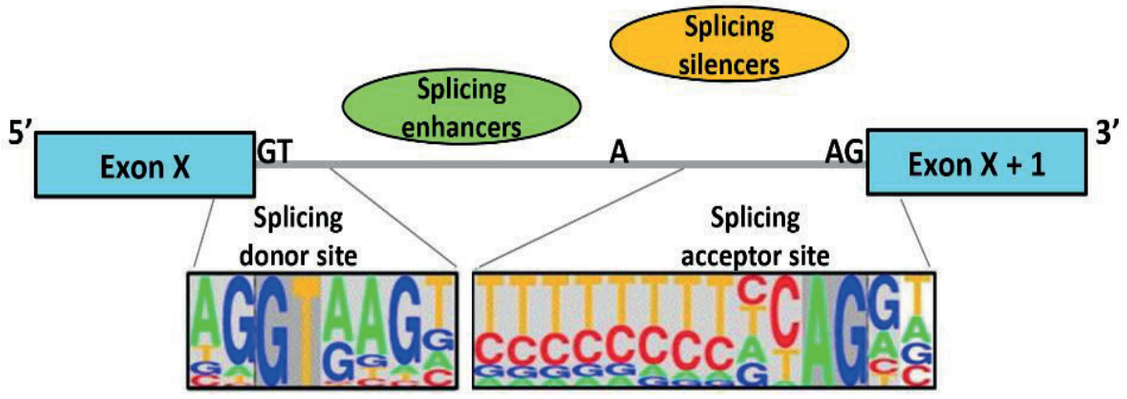


Figure 4. *Cis* sequences and *trans* elements required for correct splicing process.

Although splicing occurs via a strictly regulated process, there are alternative splicing events to support the diversity in proteins encoded by a single gene. There are several modes of alternative splicing, described as exon skipping, alternative 5' splice sites, alternative 3' splice sites, mutually exclusive exons, and intron retention (Ward and Cooper 2010). In approximately 95% of multi-exonic genes, alternative splicing occurs. This is needed to provide for the large protein diversity known in eukaryotic organisms (Black 2003). Mutations residing in the NCSS sequences affect the splice sites through similar alternative splicing events resulting in abnormal proteins.

These misplicing events have a large impact on protein function and have been reported for almost all known diseases (Wang and Cooper 2007). They may result in skipping of one or more exons, elongated or truncated exons. When introns are not very large, they can be retained between two exons as a result of a splicing variant. Alternatively, intronic segments can be erroneously recognized as coding sequences and retained in the mature transcript, thus generating PEs. An overview of the described misplicing events is provided in **Figure 5**.

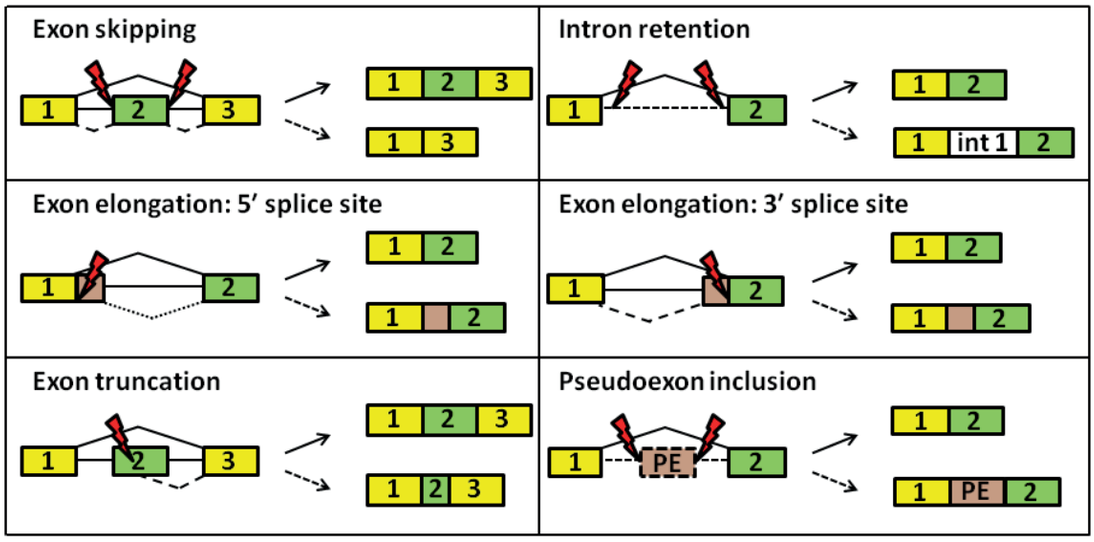


Figure 5. Splicing abnormalities. (int 1) intron 1; (PE) pseudoexon.

Comprehensive splice assays for *ABCA4* variants

If a gene is expressed in many different cell types, variants potentially affecting RNA splicing can be tested directly by reverse transcription-PCR of mRNA extracted from accessible human tissues such as lymphoblasts or fibroblasts. This is not the case for many retina-specific expressed genes including *ABCA4,* as they are very low or not expressed in non-ocular tissues. Alternatively, *in vitro* splice assays can be performed to assess the effect of these variants. In a regular minigene splice assay, the wild-type and mutant region of interest are amplified from genomic DNA of a patient carrying the potential splicing defect and cloned into a minigene splicing vector that contains a ubiquitously expressed promoter, a transcriptional start site, and at least two exons of another gene that flank the cloning site. But these minigenes do not contain the complete genomic context and the presence of strong splice donor and acceptor sites of the upstream and downstream exon, respectively, may not result in an accurate result. To overcome this issue, we amplified 29 sizeable (4.5 to 11.5 kb) wild-type genomic inserts from a bacterial artificial chromosome clone carrying *ABCA4*, and cloned them into a splicing vector (**Chapter 2**). These multi-exonic clones, coined 'midigenes', encompass all *ABCA4* exons (exons 2 to 48), flanked by *RHO* exons 3 and 5. These wild-type constructs were then used to assess the effect of all 47 reported *ABCA4* NCSS variants. All splicing defects were introduced into these midigenes and wild-type and variant-containing midigenes were tested in HEK293T cells. In this way, splicing defects were established for 44 of the 47 tested variants, of which 26 variants were deemed severe as they did not show any normal splicing product suggesting complete loss-of-function of the ABCA4 protein, and 18 variants showed a partial splicing defect (**Chapter 2**). **Figure 6** shows an example of such a splice assay. Variant

c.3607G>A was introduced into the wild-type splice midigene BA17 through mutagenesis and transfected into HEK293T cells (**Figure 6A**). Two days post transfection, RNA was isolated, reverse transcribed into cDNA and amplified employing PCR primers in exons 23 and 26. The result showed the skipping of *ABCA4* exon 24 in majority of the mutant mRNA (**Figure 6B**).

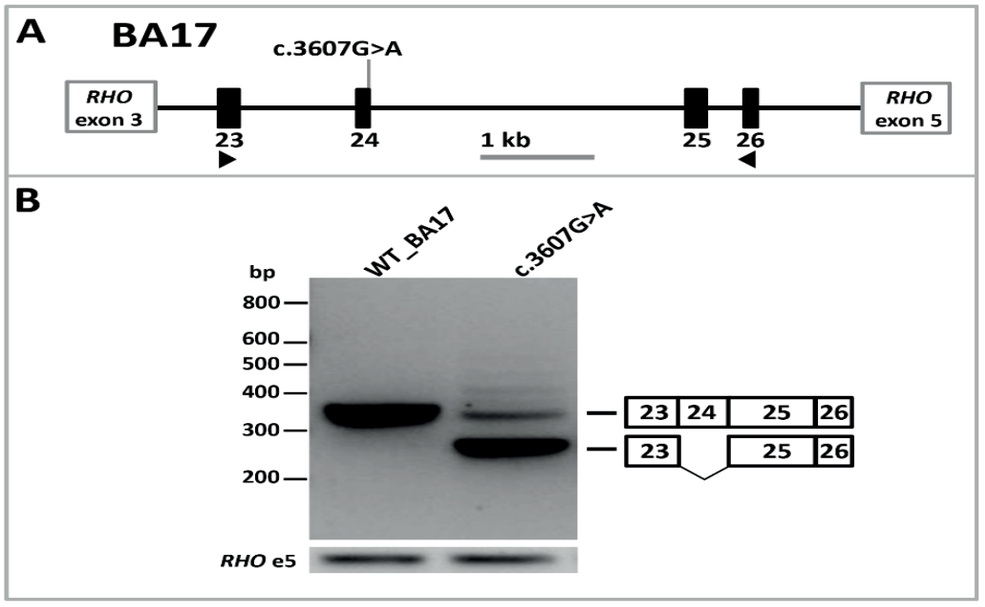


Figure 6. An example of *in* vitro splice assay. *In vitro* splice assay in HEK293T cells reveals exon 24 skipping due to a variant affecting the last position of exon 24. Arrow heads point to the location of forward and reverse primers to perform RT-PCR. *RHO* amplification is used for transfection comparison.

Identification and functional analysis of deep-intronic *ABCA4* variants

In previous studies, several DI variants were identified and for some, the effect on the *ABCA4* RNA, PE insertion, was revealed using patient-derived keratinocytes (Braun et al. 2013). These DI variants were subsequently found in several other studies (Zernant et al. 2014; Bauwens et al. 2015; Bax et al. 2015; Zaneveld et al. 2015; Schulz et al. 2017). From a cohort of 65 mono-allelic Dutch STGD1 cases, the p.Asn1868Ile variant was found as a likely causal 2nd allele in 25 cases. However, 40 patients remained genetically unsolved. Therefore, the entire 128-kb *ABCA4* gene was sequenced in these cases to search for DI variants. An average of three rare DI variants were identified per person in the *ABCA4* gene. Selected DI variants were introduced in the appropriate wild-type midigenes which were then transfected into HEK293T cells. RT-PCR analysis of 11 rare tested variants revealed PEs for five variants and partial exon 28 skipping for one variant (**Chapter 3**). PEs typically are between 50 and 200 bp in size and are mostly activated through the strengthening of weak (cryptic) splice donor or acceptor sites (**Figures 7A-B**) or through the strengthening of other splicing motifs (**Figure 7C**). Other

PEs were shown to be due to the recognition of retina-specific splicing motifs which are not present in HEK293T cells and could only be found in patient-derived photoreceptor precursor cell using stem cell technology (Albert et al. 2018). All causal DI variants result in protein truncations. Expert clinical and genetic studies for *ABCA4*-associated STGD1 enabled the genetic elucidation, i.e., the identification of variants in both *ABCA4* copies, in 292/300 (97%) cases.

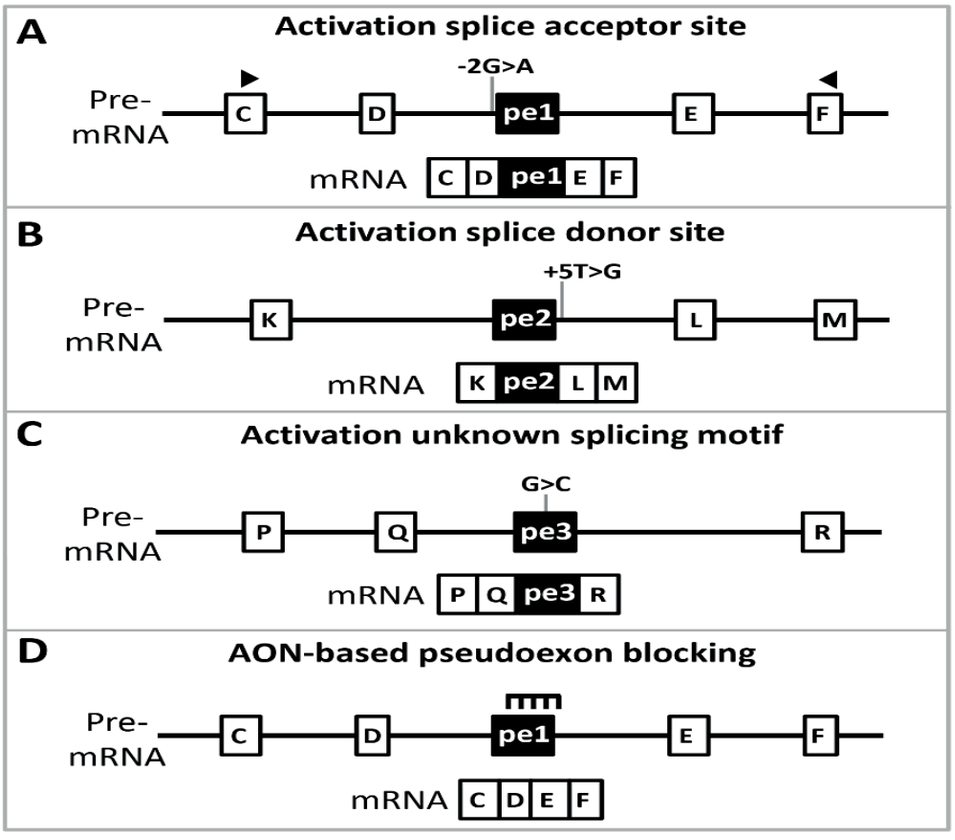


Figure 7. Pseudoexon insertions in the *ABCA4* mRNA and antisense oligonucleotide-based rescue of splicing defect. A. A new splice acceptor site is formed through a sequence alteration. In the presence of a functional splice donor site, a pseudoexon (PE) is formed which will be spliced into the mRNA. **B.** A cryptic splice donor site is strengthened through a nucleotide substitution. **C.** An exonic splicing enhancer motif is created through a G to C change. If this motif is located in a cryptic exon a PE is inserted into the mRNA. **D.** An antisense oligonucleotide that can bind to part of the PE can effectively block pe1 PE insertion.

Therapeutics

The ultimate aim of these studies is not only to identify the genetic causes of STGD1, but also to provide the basis for the development of potential therapies. Many different strategies are being employed to treat *ABCA4*-associated diseases. The most important ongoing clinical studies are mentioned below.

Antisense oligonucleotide-based therapy

Currently, we and others identified >25 DI variants that result in PE insertions into the mRNA. To correct these RNA splicing defects, antisense oligonucleotides (AONs) can been used. AONs are small (~15-25 nt) molecules which are able to bind pre-mRNA and redirect splicing, for instance by promoting PE skipping (Hammond and Wood 2011) (**Figure 7D**). AONs successfully redirected splicing, for several variants in *ABCA4* using HEK293T cells and patient-derived fibroblasts and photoreceptor precursors cells (Albert et al. 2018; Bauwens et al. 2019; **Chapter 3**). Moreover, AONs can be employed to rescue all splicing defects at the mRNA level, i.e., exon splicing, exon elongation and PE insertions. This could benefit a large number of STGD1 patients when there is an early detection of both disease-causing mutations. In recent studies, we have shown that 10% of all STGD1 probands carry DI mutations that potentially can be targeted by AONs (Bauwens et al. 2019; **Chapters 3** and **7**).

Drug Therapies

Currently many drug therapies are also available, which mainly aimed to suppress the disease progression by inhibiting the lipofuscin accumulation. So far three compounds have been developed (ALK-001, fenretinide and A1120), the effects of which are still under clinical investigation.

ALK-001 (C20-D3-vitamin A) is a deuterated form of vitamin A. It lowers the rate of vitamin A dimerization by the substitution of hydrogen atoms with deuterium isotopes. This results in less A2E/lipofuscin accumulation (Charbel Issa et al. 2015). Preclinical studies of ALK-001 in *Abca4* knockout mice showed no side effects on retinal function (Ma et al. 2011; Charbel Issa et al. 2015). ALK-001 passed clinical trial phase I for safety and proper biodistribution and it currently is in a phase II trial, for efficacy and long-term safety evaluation in a group of STGD1 patients (https://clinicaltrials.gov/; Id: NCT02402660).

Fenretinide (N-(4-hydroxyphenyl) retinamide) is a synthetic analogue of retinol which acts as a competitor for the binding to its carrier protein, retinol-binding protein 4 (RBP4). Due to its high affinity for the binding with RPB, it facilitates the removal of this protein, as the complex will be excreted via urine. Low concentrations of circulating RBP will reduce retinol concentration in the eye, resulting in decreased biosynthesis of A2E (Berni and Formelli 1992; Malpeli et al. 1996; Mata et al. 2013). A preclinical study in *Abca4* knockout mice recapitulated the *in vitro* results that showed a dose-dependent efficacy in the clearance of A2E (Weng et al. 1999; Radu et al. 2005). After the successful completion of phase I trial it entered in phase II for the treatment of geographic atrophy in patients with dry AMD (Mata et al. 2013). Despite the positive outcomes of phase II, some reports of possible adverse effects as a consequence of long-term exposure to the drug have been reported (Samuel et al. 2006).

A1120 (2-(4-(2-(trifluoromethyl)phenyl)piperidine-1-carboxamido)benzoic acid) is also a RBP4 antagonist. A preclinical study in *Abca4* knockout mice showed the efficacy of A1120 in reducing circulating levels of Rbp4, hence inhibiting the accumulation of A2E and other bisretinoids (Dobri et al. 2013).

Emixustat (emixustat hydrochloride) is another synthetic drug which modulates the visual cycle by inhibiting retinal pigment epithelium protein 65 (RPE65), a critical enzyme of this pathway. In turn reducing the availability of vitamin A derivatives (11-cis- and all-trans-retinal) to form precursors of A2E and related compounds. Phase III safety and efficacy studies are conducted by Acucela Inc. at multiple centers (https://clinicaltrials.gov/ct2/show/record/NCT03772665).

Zimura (avacincaptad pegol) is an anti-C5 complement factor inhibitor. The first STGD1 patient has been enrolled for Phase IIb clinical trials of Zimura conducted by Ophthotech (https://clinicaltrials.gov/ct2/show/NCT03364153).

Gene augmentation therapy

Gene augmentation therapy for *ABCA4*-associated STGD1 can be attractive as a relatively small increase in ABCA4 activity in retinal cells may halt disease progression, and the delayed degeneration of retinal cells in most of the *ABCA4*-associated phenotypes provides a reasonably large window for therapeutic intervention. Currently, two approaches are most widely used to deliver genetic material to the eye: viral-vector based gene delivery and non-viral gene delivery. The commonly used AAV-mediated gene delivery method is not very suitable for the *ABCA4* cDNA (size: 6.8 kb) due to a limited capacity of the vector (~4.7 kb). Therefore, some studies have been conducted using a recombinant AAV which has a AAV5 capsid (packaging capacity ~8.9 kb) to deliver the *ABCA4* in knockout mice but no significant results were observed (Allocca et al. 2008). Other studies described the delivery of two parts of the *ABCA4* coding sequence by the recombination of different AAVs (Auricchio et al. 2001; Hirsch et al. 2010; Lai et al. 2010).

An alternative approach is based on lentiviral-mediated gene delivery. These viruses are able to infect non-dividing cells and have a larger packaging capacity of ~8 kb (Thomas et al. 2003), but the disadvantage is their ability to randomly integrate in the genome. Still, lentivirus have been used to deliver human *ABCA4* cDNA under the expression of Bovine rod opsin promoter to *Abca4* knockout mice. One year follow-up studies showed the reduced accumulation of lipofuscin compared to the untreated eye (Kong et al. 2008). Based on positive outcomes of *in vitro* and animal studies, clinical trials for STGD1 were initiated using SAR422459, a lentivirus-based drug carrying the *ABCA4* gene. In the first clinical trial (NCT01367444) a patient has been treated subretinally at the Casey Eye Institute in Portland, USA. A phase I/II clinical trial was

also conducted in Paris, where 28 STGD1 cases were enrolled for StarGen, but no results have been published.

A non-viral DNA nanoparticles-based approach has been used for ocular, brain and lung gene delivery (Yurek et al. 2009; Koirala et al. 2011). These nanoparticles (~8 nm in diameter) have a packaging capacity of up to 20 kb (Fink et al. 2006). In line with this, Han and colleagues delivered the human *ABCA4* cDNA to *Abca4* knockout mice and persistent gene expression was observed up to eight months post-delivery (Han et al. 2012). Moreover, significant functional (dark adaptation by ERG) and structural (reduced fundus flecks) improvements were observed in treated animals.

Stem cell studies

A clinical trial based on stem cell therapy was started in 2011. This trial was conducted under the supervision of Advanced Cell Technology, Inc. (Marlborough, MA, NCT01345006). In this trial, RPE cells derived from human embryonic stem cells (hESCs) were used as a therapeutic approach. Twelve patients with advanced STGD1 phenotype were recruited for this study. Preclinical studies have already shown promising results as human embryonic stem cell-derived RPE cells injected into mouse models of retinal degeneration integrated with the native RPE improving visual function (Lu et al. 2009). Importantly, no evidence of tumor formation or spread to other areas of the body was observed, indicating that cells stably recognized and joined the original tissue.

In another study, hESCs-derived RPE cells were injected via subretinal injections in 18 individuals affected with STGD1 (n=9) and age-related macular degeneration (n=9) (Schwartz et al. 2016). An improved visual acuity was measured in more than half of the treated patients with no adverse cell therapy-related signs including hyperproliferation, tumorigenicity, or rejection-related inflammation. Although no concrete conclusions could be drawn due to relatively short follow-up time, lack of a formal control group, poor initial visual acuity, and a small number of patients. But these first-in-human safety studies have opened the door to future studies enrolling patients with less advanced disease.

Acknowledgments

We thank E.H. Runhart and C.B. Hoyng (Nijmegen) for providing fundus images. This work was supported by the RetinaUK, grant no. GR591 (to FPMC), the Rotterdamse Stichting Blindenbelangen, the Stichting Blindenhulp, and the Stichting tot Verbetering van het Lot der Blinden (to FPMC), and by the Landelijke Stichting voor Blinden en Slechtzienden, Macula Degeneratie fonds and the Stichting Blinden-Penning that contributed through Uitzicht 2016-12 (to FPMC).

References

Aguirre-Lamban J, Gonzalez-Aguilera JJ, Riveiro-Alvarez R, Cantalapiedra D, Avila-Fernandez A, Villaverde-Montero C, Corton M, Blanco-Kelly F, Garcia-Sandoval B, Ayuso C. 2011. Further associations between mutations and polymorphisms in the ABCA4 gene: clinical implication of allelic variants and their role as protector/risk factors. *Invest Ophthalmol Vis Sci* **52**: 6206-6212.

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

Allikmets R, Singh N, Sun H, Shroyer NF, 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.

Allocca M, Doria M, Petrillo M, Colella P, Garcia-Hoyos M, Gibbs D, Kim SR, Maguire A, Rex TS, Di Vicino U et al. 2008. Serotype-dependent packaging of large genes in adeno-associated viral vectors results in effective gene delivery in mice. *J Clin Invest* **118**: 1955-1964.

Apfelstedt-Sylla E, Theischen M, Ruther K, Wedemann H, Gal A, Zrenner E. 1995. Extensive intrafamilial and interfamilial phenotypic variation among patients with autosomal dominant retinal dystrophy and mutations in the human RDS/peripherin gene. *Br J Ophthalmol* **79**: 28-34.

Auricchio A, Kobinger G, Anand V, Hildinger M, O'Connor E, Maguire AM, Wilson JM, Bennett J. 2001. Exchange of surface proteins impacts on viral vector cellular specificity and transduction characteristics: the retina as a model. *Hum Mol Genet* **10**: 3075-3081.

Bauwens M, De Zaeytijd J, Weisschuh N, Kohl S, Meire F, Dahan K, Depasse F, De Jaegere S, De Ravel T, De Rademaeker M et al. 2015. An augmented ABCA4 screen targeting noncoding regions reveals a deep intronic founder variant in Belgian Stargardt patients. *Hum Mutat* **36**: 39-42.

Bauwens M, Garanto A, Sangermano R, Naessens S, Weisschuh N, De Zaeytijd J, Khan M, Sadler F, Balikova I, Van Cauwenbergh C et al. 2019. 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* doi:10.1038/s41436-018-0420-y.

Bax NM, Sangermano R, Roosing S, Thiadens AA, Hoefsloot LH, van den Born LI, Phan M, Klevering BJ, Westeneng-van Haaften C, Braun TA et al. 2015. Heterozygous deep-intronic variants and deletions in ABCA4 in persons with retinal dystrophies and one exonic ABCA4 variant. *Hum Mutat* **36**: 43-47.

Berni R, Formelli F. 1992. In vitro interaction of fenretinide with plasma retinol-binding protein and its functional consequences. *FEBS Lett* **308**: 43-45.

Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annual review of biochemistry* **72**: 291-336.

Braun TA, Mullins RF, Wagner AH, Andorf JL, Johnston RM, Bakall BB, Deluca AP, Fishman GA, Lam BL, Weleber RG et al. 2013. Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. *Hum Mol Genet* **22**: 5136-5145.

Bungert S, Molday LL, Molday RS. 2001. Membrane topology of the ATP binding cassette transporter ABCR and its relationship to ABC1 and related ABCA transporters: identification of N-linked glycosylation sites. *J Biol Chem* **276**: 23539-23546.

Charbel Issa P, Barnard AR, Herrmann P, Washington I, MacLaren RE. 2015. Rescue of the Stargardt phenotype in Abca4 knockout mice through inhibition of vitamin A dimerization. *Proc Natl Acad Sci U S A* **112**: 8415-8420.

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

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 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.

Dobri N, Qin Q, Kong J, Yamamoto K, Liu Z, Moiseyev G, Ma JX, Allikmets R, Sparrow JR, Petrukhin K. 2013. A1120, a nonretinoid RBP4 antagonist, inhibits formation of cytotoxic bisretinoids in the animal model of enhanced retinal lipofuscinogenesis. *Invest Ophthalmol Vis Sci* **54**: 85-95.

Fink TL, Klepcyk PJ, Oette SM, Gedeon CR, Hyatt SL, Kowalczyk TH, Moen RC, Cooper MJ. 2006. Plasmid size up to 20 kbp does not limit effective in vivo lung gene transfer using compacted DNA nanoparticles. *Gene Ther* **13**: 1048-1051.

Hammond SM, Wood MJ. 2011. Genetic therapies for RNA mis-splicing diseases. *Trends Genet* **27**: 196-205.

Han Z, Conley SM, Makkia RS, Cooper MJ, Naash MI. 2012. DNA nanoparticle-mediated ABCA4 delivery rescues Stargardt dystrophy in mice. *J Clin Invest* **122**: 3221-3226.

Hirsch ML, Agbandje-McKenna M, Samulski RJ. 2010. Little vector, big gene transduction: fragmented genome reassembly of adeno-associated virus. *Mol Ther* **18**: 6-8.

Imani S, Cheng J, Shasaltaneh MD, Wei C, Yang L, Fu S, Zou H, Khan MA, Zhang X, Chen H et al. 2018. Genetic identification and molecular modeling characterization reveal a novel PROM1 mutation in Stargardt4-like macular dystrophy. *Oncotarget* **9**: 122-141.

Klevering BJ, Yzer S, Rohrschneider K, Zonneveld M, Allikmets R, van den Born LI, Maugeri A, Hoyng CB, Cremers FP. 2004. Microarray-based mutation analysis of the ABCA4 (ABCR) gene in autosomal recessive cone-rod dystrophy and retinitis pigmentosa. *Eur J Hum Genet* **12**: 1024-1032.

Koirala A, Makkia RS, Cooper MJ, Naash MI. 2011. Nanoparticle-mediated gene transfer specific to retinal pigment epithelial cells. *Biomaterials* **32**: 9483-9493.

Kong J, Kim SR, 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.

Lai Y, Yue Y, Duan D. 2010. Evidence for the failure of adeno-associated virus serotype 5 to package a viral genome > or = 8.2 kb. *Mol Ther* **18**: 75-79.

Lu B, Malcuit C, Wang S, Girman S, Francis P, Lemieux L, Lanza R, Lund R. 2009. Long-term safety and function of RPE from human embryonic stem cells in preclinical models of macular degeneration. *Stem Cells* **27**: 2126-2135.

Ma CJ, Lee W, Stong N, Zernant J, Chang S, Goldstein D, Nagasaki T, Allikmets R. 2019. 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* doi:10.1101/mcs.a003624.

Ma L, Kaufman Y, Zhang J, Washington I. 2011. C20-D3-vitamin A slows lipofuscin accumulation and electrophysiological retinal degeneration in a mouse model of Stargardt disease. *J Biol Chem* **286**: 7966-7974.

Malpeli G, Folli C, Berni R. 1996. Retinoid binding to retinol-binding protein and the interference with the interaction with transthyretin. *Biochim Biophys Acta* **1294**: 48-54.

Mata NL, Lichter JB, Vogel R, Han Y, Bui TV, Singerman LJ. 2013. Investigation of oral fenretinide for treatment of geographic atrophy in age-related macular degeneration. *Retina* **33**: 498-507.

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**: 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 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.

PA B. 1988. Fundus flavimaculatus. In *Retinal dystrophies and degenerations*, (ed. e Newsome DA), pp. 135-159. Raven Press, New York.

Radu RA, Han Y, Bui TV, Nusinowitz S, Bok D, Lichter J, Widder K, Travis GH, Mata NL. 2005. 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* **46**: 4393-4401.

Riveiro-Alvarez R, Xie YA, Lopez-Martinez MA, Gambin T, Perez-Carro R, Avila-Fernandez A, Lopez-Molina MI, Zernant J, Jhangiani S, Muzny D et al. 2015. New mutations in the RAB28 gene in 2 Spanish families with cone-rod dystrophy. *JAMA Ophthalmol* **133**: 133-139.

Runhart EH, Sangermano R, Cornelis SS, Verheij J, Plomp AS, Boon CJF, Lugtenberg D, Roosing S, Bax NM, Blokland EAW et al. 2018. The Common ABCA4 Variant p.Asn1868Ile Shows Nonpenetrance and Variable Expression of Stargardt

Disease When Present in trans With Severe Variants. *Invest Ophthalmol Vis Sci* **59**: 3220-3231.

Samuel W, Kutty RK, Nagineni S, Vijayasarathy C, Chandraratna RA, Wiggert B. 2006. N-(4-hydroxyphenyl)retinamide induces apoptosis in human retinal pigment epithelial cells: retinoic acid receptors regulate apoptosis, reactive oxygen species generation, and the expression of heme oxygenase-1 and Gadd153. *J Cell Physiol* **209**: 854-865.

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

Sangermano R, Garanto A, Khan M, Runhart EH, Bauwens M, Bax NM, van den Born LI, Khan MI, Cornelis SS, Verheij J et al. 2019. Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. *Genet Med* doi:10.1038/s41436-018-0414-9.

Sangermano R, Khan M, Cornelis SS, Richelle V, Albert S, Garanto A, Elmelik D, Qamar R, Lugtenberg D, van den Born LI et al. 2018. ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Res* **28**: 100-110.

Schulz HL, Grassmann F, Kellner U, Spital G, Ruther K, Jagle H, Hufendiek K, Rating P, Huchzermeyer C, Baier MJ et al. 2017. Mutation Spectrum of the ABCA4 Gene in 335 Stargardt Disease Patients From a Multicenter German Cohort-Impact of Selected Deep Intronic Variants and Common SNPs. *Invest Ophthalmol Vis Sci* **58**: 394-403.

Shroyer NF, Lewis RA, Lupski JR. 2000. Complex inheritance of ABCR mutations in Stargardt disease: linkage disequilibrium, complex alleles, and pseudodominance. *Hum Genet* **106**: 244-248.

Simonelli F, Testa F, de Crecchio G, Rinaldi E, Hutchinson A, Atkinson A, Dean M, D'Urso M, Allikmets R. 2000. New ABCR mutations and clinical phenotype in Italian patients with Stargardt disease. *Invest Ophthalmol Vis Sci* **41**: 892-897.

Singh RK, Cooper TA. 2012. Pre-mRNA splicing in disease and therapeutics. *Trends in molecular medicine* **18**: 472-482.

Tanna P, Strauss RW, Fujinami K, Michaelides M. 2017. Stargardt disease: clinical features, molecular genetics, animal models and therapeutic options. *Br J Ophthalmol* **101**: 25-30.

Thomas CE, Ehrhardt A, Kay MA. 2003. Progress and problems with the use of viral vectors for gene therapy. *Nat Rev Genet* **4**: 346-358.

Ward AJ, Cooper TA. 2010. The pathobiology of splicing. *The Journal of pathology* **220**: 152-163.

Weng J, Mata NL, Azarian SM, Tzekov RT, Birch DG, Travis GH. 1999. Insights into the function of Rim protein in photoreceptors and etiology of Stargardt's disease from the phenotype in abcr knockout mice. *Cell* **98**: 13-23.

Xie YA, Lee W, Cai C, Gambin T, Noupuu K, Sujirakul T, Ayuso C, Jhangiani S, Muzny D, Boerwinkle E et al. 2014. New syndrome with retinitis pigmentosa is caused by nonsense mutations in retinol dehydrogenase RDH11. *Hum Mol Genet* **23**: 5774-5780.

Yurek DM, Fletcher AM, Smith GM, Seroogy KB, Ziady AG, Molter J, Kowalczyk TH, Padegimas L, Cooper MJ. 2009. Long-term transgene expression in the central nervous system using DNA nanoparticles. *Mol Ther* **17**: 641-650.

Zaneveld J, Siddiqui S, Li H, Wang X, Wang H, Wang K, Li H, Ren H, Lopez I, Dorfman A et al. 2015. Comprehensive analysis of patients with Stargardt macular dystrophy reveals new genotype-phenotype correlations and unexpected diagnostic revisions. *Genet Med* **17**: 262-270.

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**: 404-412.

Zernant J, Xie YA, Ayuso C, Riveiro-Alvarez R, Lopez-Martinez MA, Simonelli F, Testa F, Gorin MB, Strom SP, Bertelsen M et al. 2014. Analysis of the ABCA4 genomic locus in Stargardt disease. *Hum Mol Genet* **23**: 6797-6806

Chapter 1.4

Aim and outline of this thesis

Aim and outline of this thesis

The main purpose of this thesis is to resolve the missing heritability of *ABCA4*-associated retinal diseases through integrated genomic and transcriptomic analyses. Various genotyping methods were employed, which resulted not only in the identification of single nucleotide variants in coding regions of *ABCA4* but also in the detection of intronic variants and their functional characterization. Most of them lead to aberrant splicing due to pseudoexon (PE) inclusions. In addition, different structural variants were also identified in many probands. Finally, by using antisense oligonucleotides (AONs) we aimed to rescue *ABCA4* splicing defects at the RNA level, both in HEK293T cells and, when possible, in patient-derived photoreceptor precursor cells (PPCs).

Chapter 1 provides an introduction into the structure and function of the human retina, the genetics of IRDs, and an overview of *ABCA4* associated-STGD1 genetics and therapeutic strategies.

Chapter 2 describes the design and generation of a complete set of sizeable wild-type *ABCA4* expression constructs, so called midigenes. This allowed a comprehensive assessment of the effect of all 44 reported and three novel noncanonical splice-site (NCSS) *ABCA4* variants thus far identified.

Chapter 3 defines the targeted *ABCA4* locus sequencing in a cohort of 45 mono-allelic STGD1 patients to identify missing non-coding variants by Haloplex-based sequencing. Additionally, AONs were employed to correct the splicing defects.

Chapter 4 describes the sequence analysis of a cohort of 411 genetically unexplained STGD1 probands for all 50 exons and 14 previously identified DI variants by employing smMIPs-based sequencing.

Chapter 5 provides details on the functional characterization of all reported *ABCA4* intron 36 variants. Rescue studies were shown both in HEK293T cells and in PPCs.

Chapter 6 describes the pathogenic role of rare *ABCA4* NCSS variants and DI variants by employing midigene-based splice assays in HEK293T cells.

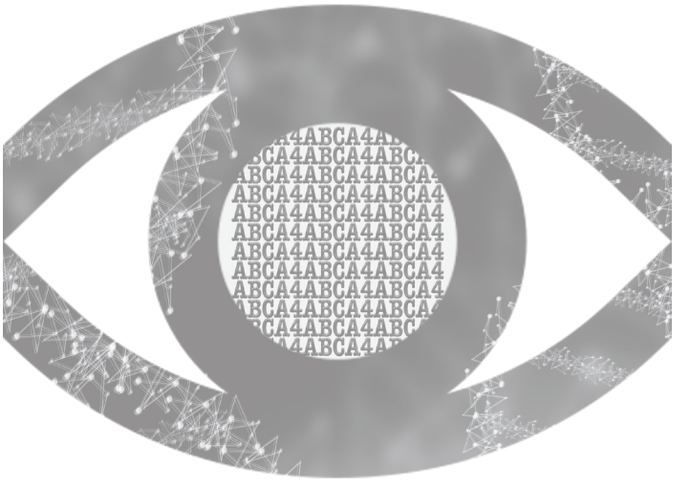
Chapter 7 describes the targeted complete *ABCA4* gene sequencing using smMIPs in a cohort of 1,054 mono-allelic and no-allele STGD1 patients to identify missing causal variants.

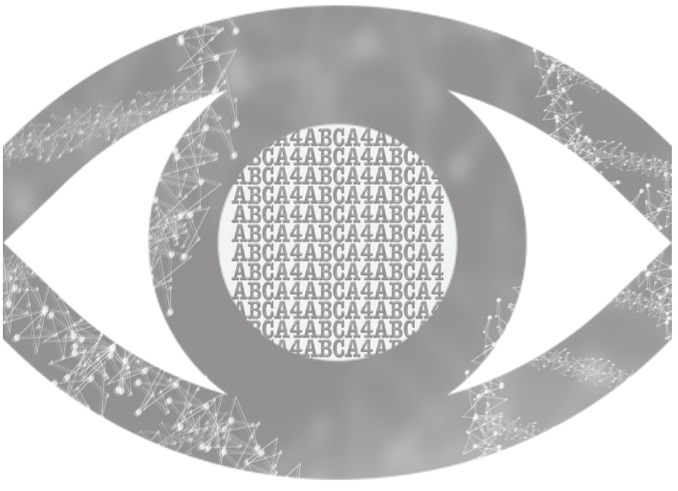
Chapter 8 describes a splice-site interdependency phenomenon. To explain this model the effect of NCSS variants in *ABCA4*, *DMD* and *TMC1* was studied.

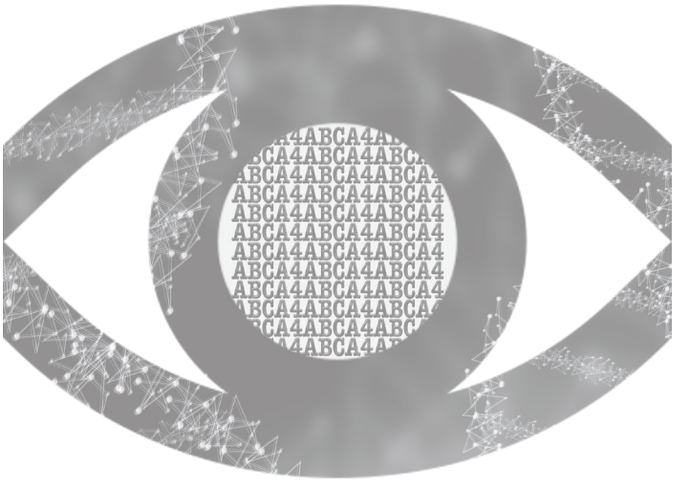
Finally, **chapters 9** and **10** provide the general discussion and summary, respectively.

Chapter 2

ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease







ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease

Riccardo Sangermano*, **Mubeen Khan***, Stéphanie S. Cornelis, Valerie Richelle, Silvia Albert, Alejandro Garanto, Duaa Elmelik, Raheel Qamar, Dorien Lugtenberg, L. Ingeborgh van den Born, Rob W.J. Collin, and Frans P.M. Cremers

Genome Res. 2018; **28:** 100-110.

*These authors contributed equally to this work

Abstract

Stargardt disease is caused by variants in the *ABCA4* gene, a significant part of which are noncanonical splice site (NCSS) variants. In case a gene of interest is not expressed in available somatic cells, small genomic fragments carrying potential disease-associated variants are tested for splice abnormalities using *in vitro* splice assays. We recently discovered that when using small minigenes lacking the proper genomic context, *in vitro* results do not correlate with splice defects observed in patient cells. We therefore devised a novel strategy in which a bacterial artificial chromosome was employed to generate midigenes, splice vectors of varying lengths (up to 11.7 kb) covering almost the entire *ABCA4* gene. These midigenes were used to analyze the effect of all 44 reported and 3 novel NCSS variants on *ABCA4* pre-mRNA splicing. Intriguingly, multi-exon skipping events were observed, as well as exon elongation and intron retention. The analysis of all reported NCSS variants in *ABCA4* allowed us to reveal the nature of aberrant splicing events and to classify the severity of these mutations based on the residual fraction of wild-type mRNA. Our strategy to generate large overlapping splice vectors carrying multiple exons, creating a toolbox for robust and high-throughput analysis of splice variants, can be applied to all human genes.

Introduction

Technological advancements have enabled the cost-effective detection of sequence variations in human genomes (Carss et al. 2017). Millions of single-nucleotide variants (SNVs) have been reported per individual, many thousands of which are rare and can be involved in rare monogenic disorders. Unfortunately, the enormous progress made in SNV detection has not been accompanied by the development of functional assays to investigate the effects of SNVs in a robust manner. Although it is very challenging to predict the effect of missense variants on protein function, characterizing putative splice variants is feasible. In the absence of patient cells, or if the gene of interest is not expressed in available somatic cells, *in vitro* splice assays are widely used. In these assays, small genomic fragments carrying potential disease-associated variants are tested for splice abnormalities. However, with the increasing detection of non-coding variants located within large introns, it is apparent that minigene assays are inadequate for testing the effects of deep-intronic (DI) variants without an intact genomic context, i.e., flanking exons and *cis*-acting elements that influence exon recognition.

The ATP-binding cassette transporter type A4 (*ABCA4*) is a 128-kb gene consisting of 50 exons and characterized by retina-specific expression (Allikmets et al. 1997). It encodes a multidomain transmembrane protein localized at the outer segments of photoreceptor cells (Sun and Nathans 1997; Molday et al. 2000). The estimated incidence of individuals with *ABCA4*-associated inherited retinal diseases is 1/10,000 (Blacharski 1988). *ABCA4* variants have been observed in the vast majority (95%) of patients with autosomal recessive (ar) Stargardt disease (STGD1), ~30% of patients with ar cone-rod dystrophy (CRD) and ~5% of patients with ar panretinal dystrophy (Cremers et al. 1998; Maugeri et al. 1999, 2000). The phenotypic heterogeneity observed in individuals with *ABCA4*-associated retinal dystrophies can be explained by a genotype-phenotype correlation model, in which the residual activity of the mutant ABCA4 protein determines the clinical outcome. Typical STGD1 patients are usually characterized by a combination of a severe and a mild variant or two moderately severe variants, while patients affected by more severe phenotypes such as CRD and panretinal dystrophy carry a severe and a moderately severe variant or two severe variants, respectively (Cremers et al. 1998; Maugeri et al. 1999, 2000).

In order to provide an accurate prognosis for people with *ABCA4*-associated disease, it is essential to assess the functional consequences of all variants. Thus far, functional studies have been limited to selected missense variants that were studied at the protein level (Sun et al. 2000; Suarez et al. 2002; Quazi and Molday

2013), and only a small number of splice site variants have been examined (Rivera et al. 2000; Shroyer et al. 2001; Wiszniewski et al. 2005; Schulz et al. 2017). While the high impact of protein-truncating *ABCA4* variants (nonsense, frameshift, canonical splice site variants) is clear, the effect of noncanonical splice site (NCSS) variants still remains largely unknown. These are either exonic or intronic variants located at the splice donor sites (SDSs) and splice acceptor sites (SASs). Intronic NCSS variants at SDSs are mostly situated at positions +3 to +6 downstream from exons, while intronic NCSS variants at splice SASs are located at positions −14 to −3 upstream of exons. Exonic *ABCA4* NCSS variants can only be found within the first or the last two nucleotides of an exon (Cornelis et al. 2017), which agrees with *in silico* predictions (Shapiro and Senapathy 1987). Very recently, all published *ABCA4* variants from 3,928 retinal dystrophy cases were collected in a Leiden Open Variation Database (www.LOVD.nl/ABCA4), and an *in silico* pathogenicity assessment was provided for the 913 unique variants (Cornelis et al. 2017). Forty-four NCSS variants were identified that together account for 10% of all published *ABCA4* alleles associated with STGD1 (Cornelis et al. 2017), and 41% of all *ABCA4* alleles in patients with STGD1 from the Radboudumc in Nijmegen (Bax N, Lambertus S, Cremers F.P.M, and Hoyng CB, unpublished data). *In silico* programs assessing the pathogenicity of NCSS variants cannot always predict the effect of pre-mRNA splicing. For instance, the second most frequent *ABCA4* variant, c.5461−10T>C, was predicted to be a mild variant, but it led to skipping exon 39, or exons 39 and 40, when tested in patient-derived photoreceptor precursor cells and a minigene assay (Sangermano et al. 2016), and in fibroblasts (Aukrust et al. 2017). It thereby could be classified as a severe *ABCA4* variant.

In view of the large number of NCSS variants in *ABCA4* with unknown functional effect, we designed a protocol to prepare sizeable multi-exon 'midigenes' in order to test the consequences of all *ABCA4* NCSS variants in their 'natural context,' resulting in a robust *in vitro* splice assay system. We generated a nearly complete library of overlapping wild-type (WT) midigenes from the *ABCA4* gene, which allowed us to systematically assess the effect of 47 NCSS variants on RNA splicing.

Results

A large minigene enables accurate RNA analysis of the noncanonical splice site variant c.5714+5G>A

Previous genotype-phenotype studies in a large family with individuals showing pseudodominant ar retinitis pigmentosa (RP) due to a homozygous c.4539+1G>T

ABCA4 splice variant, or ar CRD due to compound heterozygous c.4539+1G>T and c.5714+5G>A *ABCA4* variants, led to the hypothesis that the NCSS variant c.5714+5G>A had a moderately severe effect (Cremers et al. 1998). Intriguingly, HEK293T cells transfected with small WT and mutant minigenes that contained only exon 40 showed partial skipping of exon 40 for the WT construct and full exon 40 skipping due to the c.5714+5G>A variant (**Figure 1A**). Since our previous work (Cremers et al. 1998), and recent statistical analysis (Cornelis et al. 2017), suggested this variant had a moderate or mild effect on ABCA4 function, respectively, it seemed unlikely that the c.5714+5G>A variant would result in full exon 40 skipping. We noticed a similar artifact while studying the effect of another NCSS variant, c.5461−10T>C (Sangermano et al. 2016), so we hypothesized that the rhodopsin (*RHO*) exon 3 SDS and *RHO* exon 5 SAS are relatively strong (Human Splicing Finder scores of 87.5 and 90.7, respectively), resulting in erroneous exon 40 skipping. When we used larger minigenes containing exons 39 through 41, the erroneous exon 40 skipping in the WT construct was not observed anymore, and the mutant construct showed normal RNA product in addition to the exon 40 skipping product (**Figure 1B**), as expected for a moderately severe variant.

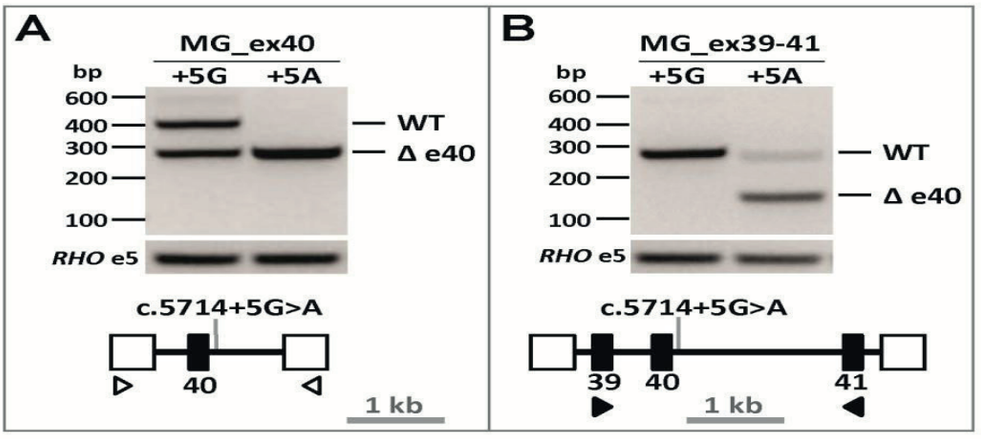


Figure 1. Large minigene enables accurate RNA analysis of the *ABCA4* noncanonical splice site variant c.5714+5G>A. A. Minigene containing the genomic region encompassing *ABCA4* exon 40 (MG_ex40, black rectangle) and flanked by rhodopsin (*RHO*) exons 3 and 5 (open boxes) was designed to investigate the effect of the noncanonical splice site variant c.5714+5G>A. Reverse-transcription polymerase chain reaction (RT-PCR) performed using primers (open arrowheads) targeting *RHO* exons 3 and 5 showed an expected 404-bp WT fragment and a 274-bp fragment corresponding to exon 40 skipping (Δ) in control minigene (+5G) and full exon 40 skipping in mutant minigene (+5A). **B.** Larger minigene containing the genomic region encompassing *ABCA4* exons 39–41 (MG_ex39–41) was designed to investigate the same variant. RT-PCR using primers (black arrowheads) targeting exons 39 and 41 showed a single 260-bp WT fragment in the control minigene, while in the mutant minigene fragments of 260 and 130 bp corresponding to WT and exon 40 skipping were found.

Generation of *ABCA4* wild-type midigene library

In view of these results and our goal to robustly test all reported NCSS variants identified in *ABCA4*, we generated multiple-exon splice vectors using DNA from a bacterial artificial chromosome (BAC) clone (CH17-325O16) spanning the entire *ABCA4* gene. In this way, a library of 29 overlapping WT midigenes (BA1–BA29) was generated (**Figure 2**), by PCR amplification of fragments that were cloned in a Gateway-adapted vector containing the *RHO* exons 3, 4, and 5 (pCI-NEO-*RHO*) with the recombination sequences located between *RHO* exons 3 and 4 and between exons 4 and 5. The insert sizes for these constructs ranged from 4.0 to 11.7 kb, with, on average, three *ABCA4* exons per construct. The maximal capacity of the vector is 12 kb. Hereafter, the pCI-NEO-*RHO* BA clones will be denoted as BA# clones.

ABCA4 exons 1 and 50 were not contained in these clones, as they lack a splice acceptor and donor site, respectively. Due to difficulties in long-range PCR amplification at the 3' end of the gene, we were also unable to amplify a genomic fragment containing exons 46–49. Instead, regular minigenes were generated for exons 48 (MG30) and 49 (MG31) using control genomic DNA as template. The midi- and minigenes together covered ~92% of the 128-kb *ABCA4* gene and 48 out of 50 exons (**Figure 2**).

In all constructs, each *ABCA4* exon was flanked by at least one upstream and one downstream *ABCA4* exon, with the exception of BA1, BA4, BA6, BA8, BA10, BA29, MG30, and MG31. For constructs BA4, BA6, BA8, and BA10, the limitation was due to the large sizes of intron 6 (15.4 kb) and 11 (14.4 kb). BA5 and BA9 did not contain any exon, as they were generated to cover the middle parts of introns 6 and 11, respectively.

Sequence analysis of the WT BA clones revealed many polymorphisms but no deleterious variants due to the PCR amplification of the insert when compared to the human genome reference GRCh37/hg19 assembly. Nevertheless, despite the use of a proofreading high-fidelity DNA polymerase, we found minor sequence variations (**Table S1**).

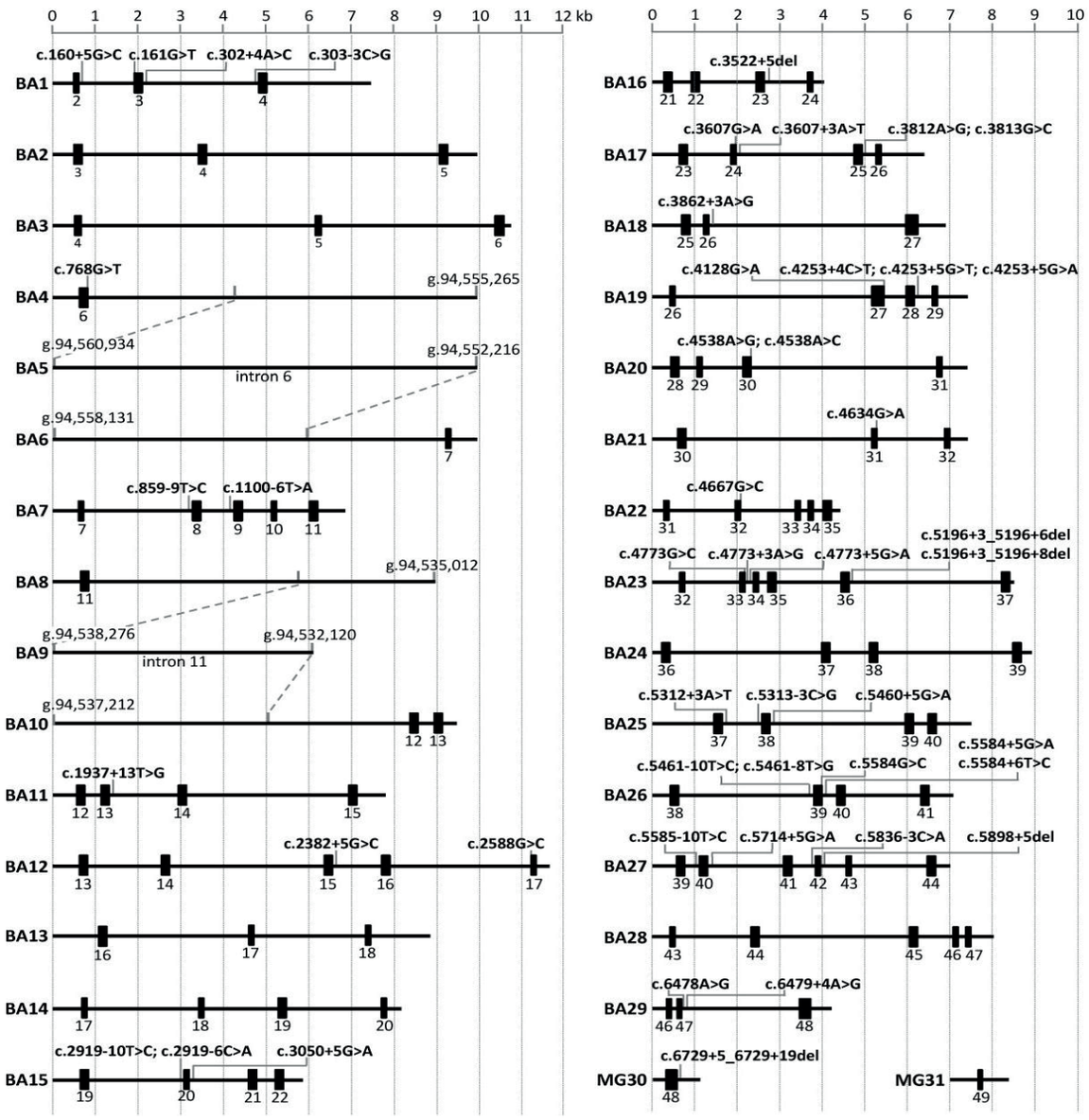


Figure 2. Overview of the wild-type midigene splice constructs of *ABCA4* and locations of all 47 noncanonical splice site variants. Exons are represented here as black rectangles. Employing bacterial artificial chromosome DNA and Gateway sequence-tagged primers, we amplified 29 genomic fragments and cloned these into the pCI-NEO-*RHO* Gateway-adapted vector (BA1–BA29). Genomic DNA from control persons (MG30 and MG31) and a patient carrying c.6729+5_6729+19del (MG30) were used to clone WT and mutant fragments.

Selection of *ABCA4* noncanonical splice site variants

The list of the 47 analyzed NCSS variants, their allelic frequencies, and predicted effect on splicing can be found in **Table S2**. The majority of these variants (70%, n = 33) are located in non-coding elements; 30% (n = 14) are located in coding regions. Highly mutated intronic positions at the SDS were at the +5 (n = 12) and +3 (n = 6) positions. We did not observe significant “hotspots” at the SASs, although positions −3 and −10 were most frequently mutated (n = 3 each).

Generation of mutant midigenes

Eighteen BA clones were used to insert 46 NCSS variants (**Figure 2** and **Figure S1**). Fifteen clones carry at least two flanking *ABCA4* exons surrounding the tested NCSS variant, thereby allowing RT-PCR using *ABCA4* primers instead of *RHO* primers. For three variants, this was not possible. c.160+5G>C and c.5312+3A>T were analyzed using a combination of a *RHO* exon 3 forward primer with *ABCA4* exons 4 or 40 reverse primers, respectively. As c.768G>T resided in BA4, RT-PCR could only be performed using *RHO* exons 3 and 5 primers.

Eleven BA clones and MG31 were not used for mutagenesis purposes in this study. For variant c.6729+5_+19del, we did not perform mutagenesis as the WT and mutant minigenes were made using genomic DNA from a patient carrying this variant in a heterozygous manner.

Splicing defects due to noncanonical splice site variants

We provided experimental evidence of splicing defects for 44 out of 47 NCSS variants. For the RT-PCR control, cDNA from 18 WT BA clones was used (**Figure 2**). Upon RT-PCR, 15 WT BA clones showed a single band corresponding to the expected splicing pattern, whereas BA1, BA12, and BA26 showed extra bands corresponding to alternative splice products (**Figure S2**). A comprehensive overview of all the splice assays for the 47 NCSS variants described in this study, as well as their sequencing analysis, can be found in **Figures S2** and **S3**. Representative results for nine variants are described below (**Figure 3** and **Figure S3**).

Variants c.160+5G>C, c.161G>T, c.302+4A>G, and c.303−3C>G were located in or near exons 2 and 4 and were introduced in BA1 (**Figures 2** and **3A–F**). Variant c.160+5G>C was predicted to significantly reduce the strength of the exon 2 SDS and likely cause exon skipping (**Table S2**). RT-PCR using a primer at the 5′ side of exon 2 therefore was not useful. Hence, we used primers in *RHO* exon 3 and *ABCA4* exon 4. The WT BA1 showed two main bands of 399 and 271 bp. Sanger sequence analysis showed that these were the expected mRNA fragment and a fragment in which exon 3 was skipped, respectively (**Figure 3A-B**). c.160+5G>C showed a complex splicing pattern, i.e., correct and three main aberrant splice products, that after pGEM-T cloning and Sanger sequencing revealed bands of 305, 285, and 271 bp, which represented the skipping of exon 2, a 14-nt elongation of exon 2 along with skipping of exon 3, and skipping of exon 3, respectively (**Figure 3A–C**). Exon 2 elongation can be explained the presence of a cryptic SDS 14-nt downstream from the actual splice site. The cryptic donor site is not shown in **Figure 3C** as it can only be seen in the genomic DNA sequence. RT-PCR spanning *ABCA4* exons 2–4 showed a WT 289-bp band for BA1. The introduction of variants c.161G>T and c.302+4A>C both showed a

unique 147-bp band corresponding to skipping of exon 3, while c.303−3C>G showed two bands of 291 and 149 bp, corresponding to a dinucleotide elongation of exon 4 and to the same elongation along with the skipping of exon 3, respectively (**Figure 3D–F**). The exon 4 elongation can be explained by the creation of a cryptic SAS due to the C>G change. The cryptic acceptor site is not shown in **Figure 3F** as it can only be seen in the genomic DNA sequence. The skipping of exon 3 in the WT BA1 clone probably was not an artifact, as we also observed exon 3 skipping in human retina RNA (**Figure S2**, **BA1**, **panel C**).

Variant c.768G>T is the second most frequent NCSS variant (allele frequency 0.106) in STGD1 cases in the Radboudumc, Nijmegen (Bax N, Lambertus S, Cremers F.P.M, and Hoyng C.B, unpublished data). Splice predictions showed a significant weakening of the exon 6 SDS (**Figure 3I**). For the WT and c.768G>T mutant BA4 clone, RT-PCR encompassing *RHO* exons 3 and 5 was performed, which yielded a WT band of 472 bp for BA4 and a 35-nt elongation of exon 6 for c.768G>T (**Figure 3G,H**), explained by the presence of a cryptic SDS 36-nt downstream (**Figure 3I**). The reading frame shift results in a predicted truncated protein, p.(Leu257Valfs*17). To validate this result '*ex vivo*', we generated photoreceptor precursor cells (PPCs) from somatic cells of a STGD1 patient carrying c.768G>T in a homozygous state. RT-PCR analysis of RNA from this patient and of normal human retina are shown in **Figure S4**. The PPC results are identical to those observed in the *in vitro* splice assay in HEK293T cells.

The novel variant c.1937+13T>G was not predicted to result in a splicing defect (Figure **3L**), but surprisingly, in 86% of mRNA products, a 12-nt elongation of exon 13 was observed for BA11 carrying c.1937+13T>G (**Figure 3J,K**). This led to a premature stop codon, resulting in p.(Phe647*) (**Figure 3L**).

For exon 30, the penultimate nucleotide substitutions c.4538A>G and c.4538A>C were introduced into BA20, and HEK293T cell transfection resulted in four fragments after pGEM-T cloning (**Figure 3M**). In addition to the correct mRNA (fragment 1), a 30-nt elongation due to the presence of a strong cryptic intronic SDS (fragment 2, **Figure 3M,O**), a deletion of 73 nt from the 3' end of exon 30 (fragment 3, **Figure 3N**), and a 30-nt elongation in combination with a 114-nt deletion from the 5' end of exon 30 (fragment 4, **Figures**. **S2**, **S3**) were found. Fragments 3 (227 bp) and 4 (216 bp) could not be resolved. The exon 30 deletions were the result of partially overlapping cryptic splice donor and acceptor sites, the noncanonical sequences (AG/GT) of which reside at exonic positions c.4465–4468 (**Figure S2**, **BA20**, **panel D**).

Six fragments were observed for c.5898+5del introduced in BA27, but we were only able to sequence three after pGEM-T cloning: bands of 717 and 780 bp corresponded to 107- nt and 170-nt elongations of exon 42 due to the use of cryptic

SDSs at positions c.5898+108 and +171, respectively, while a 1104-bp fragment corresponded to the entire retention of intron 42 (**Figure 3P,Q**).

Classification of noncanonical splice site variants according to the splicing defects

The percentage of correctly spliced *ABCA4* RNA product for NCSS variants was calculated by using a capillary electrophoresis system (**Figure 4**, and **Table S3**; **Figure S5**). Three variants, c.4634G>A, c.5196+3_5196+8del, and c.5585−10T>C, showed 100% correctly spliced *ABCA4* mRNAs. We could rule out pathogenicity for the last two variants but not for c.4634G>A p.(Ser1545Asn), as this variant is significantly enriched in >3000 Caucasian STGD1 patients compared to the non-Finnish ExAC population (Cornelis et al. 2017). Sixteen variants showed between 4.3% and 79.6% normal RNA and were tentatively classified as severe, moderately severe, and mild, while the remaining 26 NCSS variants did not show any normal RNA and were thus deemed severe variants. For c.2382+5G>C and c.2588G>C, no quantification was performed as the WT BA12 clone also showed exon skipping. Based on the Sanger sequence validation, we were able to annotate all the variants at the RNA level and predict their effect at the protein level (**Table 1**).

We correlated the predicted effect of the NCSS variants with the phenotypes in reported cases and, when the variants were not present in a homozygous state, with the severity of the second allele. For 25 of 47 variants, sufficient clinical data were available, and in all these cases, our predicted effect of the NCSS variants correlated with the observed phenotypes (**Table S4**).

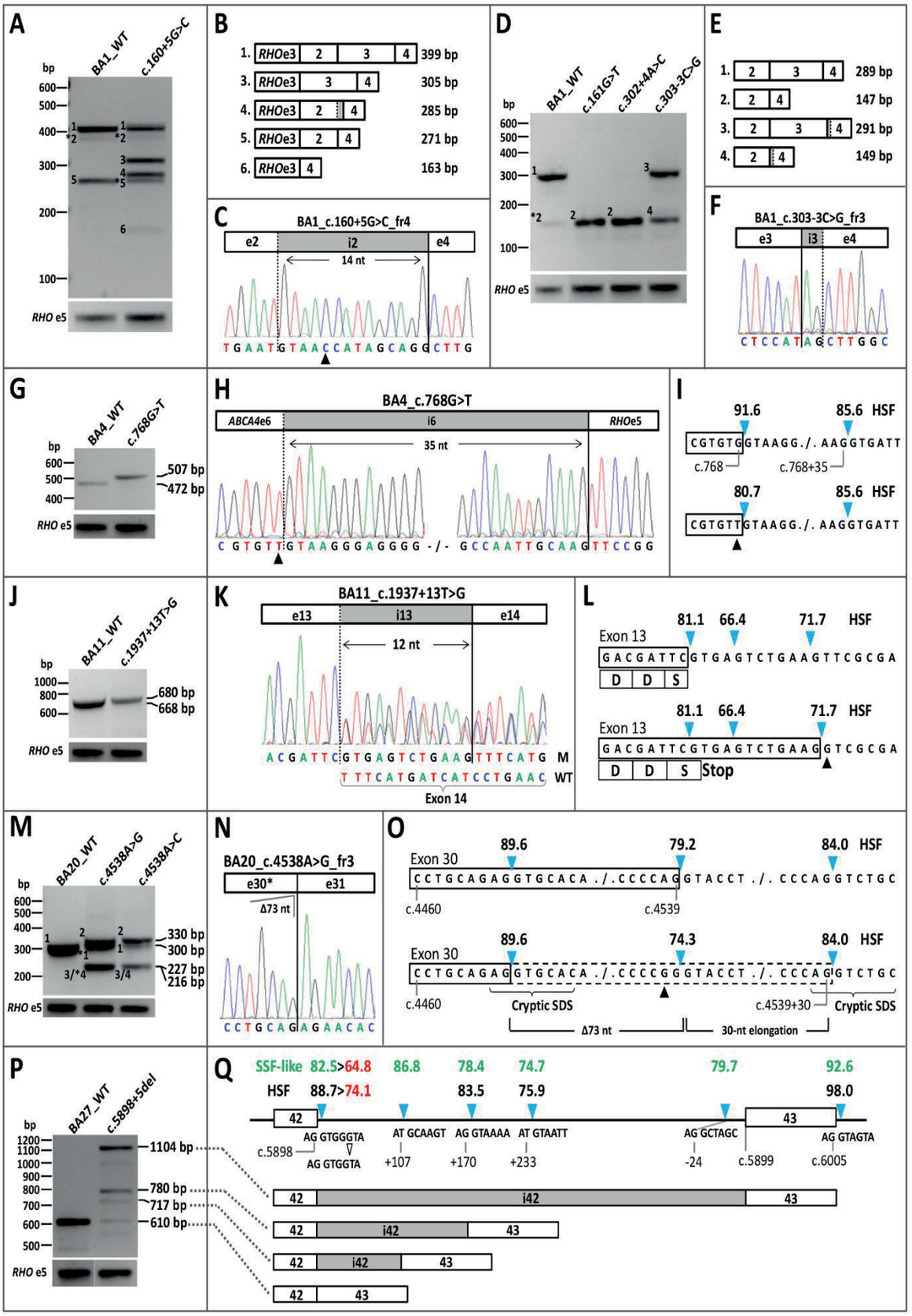

A
BA1_WT
c.160+5G>C
bp
600
500
400
300
200
100
RHO e5
B
1. RHOe3 2 3 4 399 bp
3. RHOe3 3 4 305 bp
4. RHOe3 2 4 285 bp
5. RHOe3 2 4 271 bp
6. RHOe3 4 163 bp
C
BA1_c.160+5G>C_fr4
e2
i2
e4
14 nt
TGAATGTAACCATAGCAGGCTTG
D
BA1_WT
c.161G>T
c.302+4A>C
c.303-3C>G
E
1. 2 3 4 289 bp
2. 2 4 147 bp
3. 2 3 4 291 bp
4. 2 4 149 bp
F
BA1_c.303-3C>G_fr3
e3
i3
e4
CTCCATAGCTTGGC
G
BA4_WT
c.768G>T
507 bp
472 bp
H
BA4_c.768G>T
ABCA4e6
i6
RHOe5
35 nt
CGTGTTGTAAGGGAGGGG-/-GCCAATTGCAAGTTCCGG
I
91.6
85.6 HSF
CGTGTGGTAAGG./.AAGGTGATT
c.768
c.768+35
80.7
85.6 HSF
CGTGTTGTAAGG./.AAGGTGATT
J
BA11_WT
c.1937+13T>G
1000
800
600
680 bp
668 bp
K
BA11_c.1937+13T>G
e13
i13
e14
12 nt
ACGATTCGTGAGTCTGAAGTTTCATG M
TTTCATGATCATCCTGAAC WT
Exon 14
L
81.1
66.4
71.7 HSF
Exon 13
GACGATTCGTGAGTCTGAAGTTCGCGA
D D S
81.1
66.4
71.7 HSF
Exon 13
GACGATTCGTGAGTCTGAAGGTCGCGA
D D S Stop
M
BA20_WT
c.4538A>G
c.4538A>C
330 bp
300 bp
227 bp
216 bp
3/*4
3/4
N
BA20_c.4538A>G_fr3
e30*
e31
Δ73 nt
CCTGCAGAGAACAC
O
89.6
79.2
84.0 HSF
Exon 30
CCTGCAGAGGTGCACA./.CCCCAGGTACCT./.CCCAGGTCTGC
c.4460
c.4539
89.6
74.3
84.0 HSF
Exon 30
CCTGCAGAGGTGCACA./.CCCCGGGTACCT./.CCCAGGTCTGC
c.4460
Cryptic SDS
c.4539+30
Cryptic SDS
Δ73 nt
30-nt elongation
P
BA27_WT
c.5898+5del
1200
1100
1000
900
800
700
600
500
1104 bp
780 bp
717 bp
610 bp
RHO e5
Q
SSF-like 82.5>64.8
86.8
78.4
74.7
79.7
92.6
HSF 88.7>74.1
83.5
75.9
98.0
42
43
AG GTGGGTA
AG GTGGTA
c.5898
AT GCAAGT
AG GTAAAA
AT GTAATT
AG GCTAGC
AG GTAGTA
+107
+170
+233
-24
c.5899
c.6005
42 i42 43
42 i42 43
42 i42 43
42 43

Figure 3. Overview of splicing defects due to nine noncanonical splice site variants in *ABCA4*. All wild-type (WT) and mutant midigenes were transfected in HEK293T cells and their RNA subjected to RT-PCR. **A, B.** RT-PCR for WT and c.160+5G>C mutant BA1 midigene showed complex splicing defects when using primers in *RHO* exon 3 and *ABCA4* exon 4. Five defects (fragments 2 through 6) were observed next to the WT fragment 1. Asterisks denote fragments for which we obtained no sequence information. The RT-PCR products showed skipping of exon 2, exon 3, or exons 2 and 3. **C.** Sequencing of pGEM-T-cloned fragment 4 revealed a 14-nt insertion due to the activation of a cryptic splice donor site (SDS) in intron 2 and the absence of exon 3. The triangle points to the +5G>C variant. **D, E**. RT-PCR products of WT and mutant BA1 containing NCSS variants residing at splice sites of exons 3 and 4. Primers in *ABCA4* exons 2 and 4 reveal exon 3 skipping and/or exon 4 elongation. Note that the WT vector BA1 shows exon 3 skipping (fragment 5), which is more pronounced when using a forward *RHO* exon 3 primer (**A**) compared to using a forward *ABCA4* exon 2 primer (**D**). **F.** Sanger sequence of gel-purified fragment 3 shows a 2-nt elongation at the 3′ end of exon 4 due to the activation of a cryptic splice acceptor site (SAS) in intron 3. Sanger sequencing results for all fragments without an asterisk are provided in **Figure S3**. **G, H**. RT-PCR and Sanger sequencing of RNA extracted from c.768G>T mutant BA4 midigene showed a 35-nt exon 6 elongation. **I.** Human Splice Finder (HSF) splice site scores (blue arrowheads) for WT and mutant sites. **J, K.** RT-PCR, gel analysis, and Sanger sequencing showed the WT product and a 12-nt exon 13 elongation due to c.1937+13T>G. M and WT denote mutant and WT splice products. **L.** Details of the exon elongation defect. As represented by the HSF prediction software, the scores of the normal and intronic SDSs are not changed due to this variant. Most of the RT-PCR product consists of the exon elongation which introduces a premature stop codon. **M–O**. Variants at the penultimate position of exon 30 (c.4538A>G and c.4538A>C) resulted in a correct mRNA (fragment 1), a 30-nt exon 30 elongation due to the presence of a strong cryptic intronic SDS (fragment 2), a deletion of 73 nt of the 3′ end of exon 30 (fragment 3), and a 30-nt exon 30 elongation in combination with a 114-nt deletion of the 5′ end of exon 30 (fragment 4). Fragments 3 (227 bp) and 4 (216 bp) could not be resolved. The exon 30 deletions are the result of a cryptic splice donor and acceptor sites that overlap at exonic positions c.4465–4468 (**Figure S2**, BA20). **P.** RT-PCR of BA27 carrying the c.5898+5del variant resulted in at least two exon 42 elongation products of 717 and 780 bp, as well as intron 42 retention. **Q.** The exon elongation products were due to the reduction of the strength of the exon 42 SDS, as shown with red numbers, in combination with the presence of nearby cryptic intronic SDSs with strong predicted scores. Only HSF and SpliceSiteFinder (SSF)-like scores are provided. Note that SSF-like is the only software that predicts the rare SDSs (positions c.5898+108 and c.5899−23) that contain the canonical GC sequence instead of GU. Intron 42 retention likely is due to the very strong SDS (HSF 98.0) flanking exon 43. The open arrowhead denotes the 1-bp deletion.

Table 1. *In vitro* tested noncanonical splice site variants and their observed RNA and predicted protein effects.

DNA variant	RNA effect	Protein effect	% correctly spliced mRNA	Classification based on RNA splice
c.160+5G>C	r.[=,67_160del, 161_302delins(14)]	p.[=,Ile23Alafs*24, His55Asnfs*63]	34,3	Moderate
c.161G>T	r.161_302del	p.(Cys54Serfs*14)	0	Severe
c.302+4A>C	r.161_302del	p.(Cys54Serfs*14)	0	Severe
c.303-3C>G	r.[161_302delinsag, 302_303insag]	p.[Cys54*,Leu102Alafs*14]	0	Severe
c.768G>T	r.768_769ins(35)	p.(Leu257Valfs*17)	0	Severe
c.859-9T>C	r.[=,859_1356del]	p.[=,Phe287_Arg452del]	75,7	Mild
c.1100-6T>A	r.1099_1100insgcag	p.(Thr367Serfs*6)	0	Severe
c.1937+13T>G	r.[1937_1938ins(12),=]	p.[Phe647*,=]	14#	Severe
c.2382+5G>C*	r.[=,2161_2382del]	p.[=,His721_Val794del]	47,9#	Moderate
c.2588G>C*	r.[2588g>c, 2588_2590delcag]	p.[Gly863Ala, Gly863del]	60#	Mild
c.2919-10T>C	r.[=,2919_3050del]	p.[=,Leu973_His1017delinsPhe]	61,1	Moderate
c.2919-6C>A	r.[=,2919_3050del]	p.[=,Leu973_His1017delinsPhe]	79,6	Mild
c.3050+5G>A	r.2919_3050del	p.(Leu973_His1017delinsPhe)	0	Severe
c.3522+5del	r.[=,3329_3522del]	p.[=,Arg1111Aspfs*7]	53,0	Moderate
c.3607G>A	r.3523_3607del	p.(Thr1176Metfs*2)	10,9	Severe
c.3607+3A>T	r.3523_3607del	p.(Thr1176Metfs*2)	0	Severe
c.3812A>G	r.3608_3813del	p.(Gly1203Aspfs*10)	0	Severe
c.3813G>C	r.3608_3813del	p.(Gly1203Aspfs*10)	0	Severe
c.3862+3A>G	r.[=,3814_3862del]	p.[=,Ile1272Valfs*101]	53,4	Moderate
c.4128G>A	r.4128_4129ins(12)	p.(Gln1376_Ile1377ins4)	0	Severe
c.4253+4C>T	r.4129_4253del	p.(Ile1377Hisfs*3)	7,8	Severe
c.4253+5G>A	r.4129_4253del	p.(Ile1377Hisfs*3)	0	Severe
c.4253+5G>T	r.[4129_4253del,=]	p.[Ile1377Hisfs*3,=]	5,4	Severe
c.4538A>G	r.[4539_4540ins(30), 4467_4539del]	p.[Arg1513_Arg1514ins10,Cys1490Glufs*12]	0$	Severe
c.4538A>C	r.[4539_4540ins(30),	p.[Pro1513_Arg1514ins10,Cys1490Glufs*12,Gl	4,3$	Severe
c.4634G>A	r.4634g>a	p.(Ser1545Asn)	100	Benign
c.4667G>C	r.4635_4667del	p.(Ser1545_Gln1555del)	0	Severe
c.4773G>C	r.[4668_5018del, 4668_4773del]	p.[Tyr1557_Val1673del, Tyr1557Alafs*18]	0	Severe
c.4773+3A>G	r.[4668_4773del,=]	p.[Tyr1557Alafs*18,=]	24,6	Severe
c.4773+5G>A	r.[4668_4773del,=,4668_5018del]	p.[Tyr1557Alafs*18, =,Tyr1557_Val1673del]	29,1	Severe
c.5196+3_5196+6	r.4849_5196del	p.(Val1617_Ile1732del)	0	Severe
c.5196+3_5196+8	r.=	p.(=)	100	Benign
c.5312+3A>T	r.5197_5312del	p.(Asn1734Glyfs*14)	0	Severe
c.5313-3C>G	r.5312_5313insag	p.(Trp1772Aspfs*7)	0	Severe

c.5460+5G>A	r.5313_5460del	p.(Trp1772Argfs*9)	0	Severe
c.5461-10T>C	r.5461_5714del	p.(Thr1821Aspfs*6)	0	Severe
c.5461-8T>G	r.5461_5714del	p.(Thr1821Aspfs*6)	0	Severe
c.5584G>C	r.5461_5714del	p.(Thr1821Aspfs*6)	0	Severe
c.5584+5G>A	r.[5461_5714del, 5461_5584del]	p.[Thr1821Aspfs*6,Thr1821Valfs*13]	0	Severe
c.5584+6T>C	r.[5461_5714del,5461_5584del]	p.[Thr1821Aspfs*6,Thr1821Valfs*13,Glu1863Le	0	Severe
c.5585-10T>C	r.=	p.(=)	100	Benign
c.5714+5G>A	r.[=, 5585_5714del]	p.[=, Glu1863Leufs*33]	39,8	Moderate
c.5836-3C>A	r.5835_5836ins(30)	p.(Lys1945_Ile1946Pheins10)	0	Severe
c.5898+5del	r.[5898_5899ins(494),	p.(Cys1967Valfs*24)	4,5	Severe
c.6478A>G	r.[6478a>g, 6387_6479del]	p.[Lys2160Glu, Ser2129_Lys2160delinsArg]	55,2	Moderate
c.6479+4A>G	r.6387_6479del	p.(Ser2129_Lys2160delinsArg)	0	Severe
c.6729+5_6729+19	r.6480_6729del	p.(Phe2161Cysfs*3)	0	Severe

The nomenclature of all variants is according to the Human Genome Variation Society recommendations for the description of the sequence variants. Complex effects at both the RNA and protein levels were included between square brackets. The products were listed according to their abundance based on capillary electrophoresis-based analysis (**Table S3**). For variants c.2382+5G>C and c.2588G>C, nomenclature does not include the skipping of *ABCA4* exon 15 (*). The c.2588G>C variant previously was classified as a mild allele (Maugeri et al. 1999). For variants c.1937+13T>G and c.2382+5G>C, the percentage of correctly spliced mRNA was determined by densitometry, and for variant c.2588G>C by direct analysis of Sanger sequence traces (#). For variants c.4538A>C and c.4538A>G, the percentage of correctly spliced mRNA was not entirely determined due to likely comigrating gel fragments of similar size ($). (=) Wild-type.

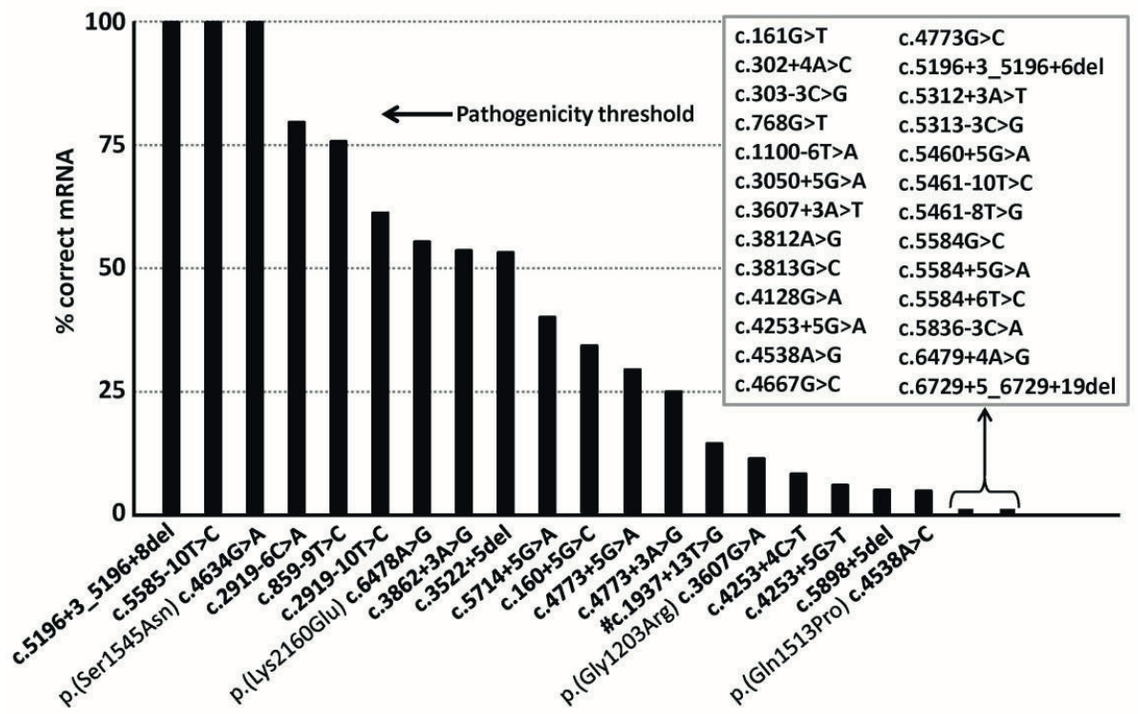


Figure 4. Percentages of remaining normal *ABCA4* transcripts due to noncanonical splice site variants based on capillary electrophoresis system analyses. The percentages of normal *ABCA4* transcript for 45 noncanonical splice site variants are represented by black bars. Nineteen variants showed varying fractions (4.3%–100%) of correct *ABCA4* mRNA. Four of them also result in missense changes as depicted, but only p.(Ser1545Asn) and p.(Lys2160Glu) are likely to have an effect on protein function, as significant amounts of ABCA4 protein will be produced. For the remaining 26 variants (in the square box), no residual RNA was observed. (#) For this variant, densitometric scanning was performed.

Discussion

Splice site variants cause 15%–20% of human genetic disease (Matlin et al. 2005), and an increasing number of DI variants have been identified through targeted gene sequencing and whole-genome sequencing (www.dbass.soton.ac.uk) (Buratti et al. 2011; Carss et al. 2017). Contrary to canonical splice site variants, which in about two-thirds of cases result in exon skipping (Cheung et al. 2017), the effect of NCSS variants can vary tremendously, depending on the sequence alteration itself and its sequence environment. In recent studies, large-scale analyses of variants in or near consensus splice site sequences were performed using minigene constructs containing small exons flanked by minimal intronic sequences (Cheung et al. 2017). The maximum insert size of the tested genomic segments measured 170 bp, as the analysis of spliced sequences was performed by short- read next-generation sequencing. Although this type of study enabled high-throughput analysis of randomly designed splice site sequences (Rosenberg et al. 2015) or naturally occurring NCSS variants (Cheung et al. 2017), the size constraint of the employed constructs represents a major limitation to fully assess the effect of NCSS variants. As shown here, it

is important to study NCSS variants in the context of their upstream or downstream introns and exons, as their effect often is influenced by other intronic *cis*-regulatory elements or affects exons located at considerable distances.

We have performed a systematic *in vitro* analysis of all reported and three novel NCSS variants in *ABCA4* that are associated with STGD1 by generating midigenes: large, multiple-exon splice vectors. Forty-four of 47 variants resulted in splicing defects. The most frequently observed defect, seen in 29 (62%) of the NCSS variants, was single or multiple-exon skipping. For eight (17%) variants, we observed the simultaneous occurrence of different splicing defects, e.g., exon skipping/elongation, skipping/retention, elongation/retention and truncation/elongation. Six NCSS variants (13%) resulted in exon elongation.

The majority of NCSS variants (n = 35; 75%) affected the SDS. This concurs with the observation that the majority (65%) of pseudoexon (PE) mRNA insertions associated with human diseases also concern the activation of SDSs (www.dbass.soton.ac.uk) (Buratti et al. 2011). The longer intronic consensus sequences of SASs (−3 to −14 for SAS vs. +3 to +6 for SDS) in which pyrimidine to pyrimidine substitutions have relatively little effect on the predicted splice strengths, may explain why variants that affect the SAS are generally less disruptive. The severe NCSS defect previously shown for c.5461−10T>C (exon 39 or exon 39 and 40 skipping) (Sangermano et al. 2016) prompted us to investigate the effect of two similar rare variants, i.e., c.2919−10T>C and c.5585−10T>C. We found that c.2919−10T>C resulted in exon skipping in 39% of the *ABCA4* transcripts, while c.5585−10T>C showed no splicing defects. The predicted reductions of the corresponding SAS splice scores all were minor. These contrasting results suggest that the effect of a NCSS variant also depends on the sequence context, e.g., the strengths of nearby splice acceptor and donor sites, the size of the flanking exons, and possibly exonic and intronic splicing enhancer (ESE, ISE) and silencer motifs (ESS, ISS).

The effect of still unknown sequence motifs on splicing is also illustrated by our findings for c.1937+13T>G. This is not a typical NCSS variant as it is located downstream from the exon 13 SDS consensus sequence. As illustrated in **Figure 3**, **J** through **L**, this change does not lower the strength of the normal exon 13 SDS, nor does it create a stronger predicted downstream SDS. Nevertheless, it resulted in a 12- nt elongation of exon 13 and introduced a premature stop codon in 86% of the transcripts, based on Sanger sequencing. The Alamut ESEfinder did not predict this variant to create a splicing enhancer (Cartegni et al. 2003; Smith et al. 2006). Follow-up studies are required to investigate the nature of this apparently unknown sequence motif that influences splicing. Importantly, this result also illustrates that the assessment of rare non-coding variants identified in diagnostic testing cannot fully rely on *in silico* predictions.

The generation of multiple-exon splice constructs in which most of the introduced NCSS variants were flanked by one or more upstream and downstream exons allowed us

to investigate the full spectrum of splicing defects. This is illustrated by two +5delG variants, c.3522+5del in **Figure S2** and c.5898+5del in **Figure 3**, **P** and **Q**, which not only showed multiple exon elongation but also retention of introns 23 and 42, respectively. The exon elongations could be explained by similarities shared by both introns, as they both carry multiple strong cryptic SDSs, small downstream introns (intron 23, 1081 bp; intron 42, 494 bp), and strong SDSs located at the 3'end of the following exon. The importance of the presence of the sequence context was also found for *NF1* as the extent of exons 36 and 37 skipping due to a variant in exon 37 was determined by the presence of exons 34–38 in the splice construct (Baralle et al. 2006).

The observation that NCSS variants with an effect on RNA splicing resulted in a maximum of 79.6% of normal RNA leads us to believe that there may be a threshold for alleles to behave as 'mild' alleles. In our experiments, that threshold is ~80% (**Figure 4**). In cases where there is more than 80% normal RNA, the variants will likely be benign and will not contribute to the phenotype in patients with STGD1. Similarly, there likely are thresholds for moderately severe and severe alleles. However, as the clinical data of patients carrying NCSS variants were not described consistently, we could not accurately assign these thresholds. A similar threshold model was recently described for the remaining normal CEP290 protein in patients with Leber congenital amaurosis and various syndromic phenotypes, as a result of different combinations of coding and DI variants (Drivas et al. 2015).

The 29 BA constructs contain large (4.0–11.7 kb) segments of *ABCA4*. We could not clone genomic fragments larger than 11.7 kb due to the size constraint (12 kb) of the Gateway cloning system. We therefore cloned introns 6 and 11 in different overlapping fragments and performed transcript analysis using *RHO* primers. Constructs containing the first and last exon were not generated as they lack a splice acceptor and donor site, respectively. A difficult region to amplify was the 3' part of the gene encompassing intron 48–exon 50. The absence of both flanking *ABCA4* exons triggered splicing artifacts due to the presence of a strong *RHO* splice donor and acceptor site, as we observed erroneous exon skipping in MG_ex40 (**Figure 1**).

Natural skipping of *ABCA4* exons 3, 15, and 39/40 was observed by others (Braun et al. 2013) and by us (**Figure S2**), suggesting that the midigenes mimic the *in vivo* situation. An accurate comparison of the skipping events for exons 3 and 39/40 between WT midigene and human retina is not possible, as they result in frameshifts and the corresponding mRNAs are subject to nonsense-mediated decay in the human retina.

Validating multiple splice products can be time-consuming as gel-purification and Sanger sequencing are hampered by contaminations from other fragments, thus requiring plasmid cloning and additional Sanger sequencing. A more general drawback of the *in vitro* splice assay is that, despite the observed splicing similarities between WT BA clones and human retina, splicing in HEK293T cells may not always mimic the splicing defects as they

might occur in the retina of patients. Other studies have indicated that indeed there are retina-specific splice factors (Murphy et al. 2016; Parfitt et al. 2016).

The use of bacterial artificial chromosome clones facilitated the generation of large, multiple-exon WT splice vectors. In an alternative approach, a 5.6-kb *BRCA1* minigene was constructed by cloning six consecutive fragments carrying exons 19–27 that, in the human genome, span 28.7 kb (Acedo et al. 2015). Our approach can readily be applied to generate midigenes from all human genes, as BAC clones are available for this purpose (https://bacpacresources.org/) (Osoegawa and de Jong 2004). These WT midigenes not only allow the accurate assessment of large splicing defects due to NCSS variants but can also be used to test DI variants hypothesized to result in PE insertions (**Chapter 3**).

We observed sizeable exon elongations in significantly occurring RNA products due to c.160+5G>C, c.768G>T, c.1937+13T>G, c.4538A>G, and c.4538A>C variants. Splice modulation using antisense oligonucleotides (AONs) may be an option to redirect the splicing back to the natural splice sites. AONs have already demonstrated their therapeutic potential for several eye disease-associated genes, such as *CEP290* (Collin et al. 2012; Gerard et al. 2012; Garanto et al. 2016; Parfitt et al. 2016), *USH2A* (Slijkerman et al. 2016), or *OPA1* (Bonifert et al. 2016), and have a great potential as therapeutic molecules for inherited retinal diseases.

In conclusion, we generated a complete set of sizeable WT *ABCA4* splice vectors, which allowed us to gain insight into the severity of the poorly explored class of NCSS *ABCA4* variants on a gene-wide scale. By assessing the functional consequences of putative splicing defects, STGD1 patients can obtain a more accurate molecular diagnosis and thereby become eligible for novel therapies. Our approach can be used for all human genes to effectively test putative splicing defects. Besides from NCSS variants, midigenes are also especially suited for testing rare DI variants and their effect on splicing in a reliable manner, as 90% of introns are <11 kb (Sakharkar et al. 2004) and thereby completely fit into midigenes, together with all cis-acting splice factor binding motifs and flanked by their natural exons.

Materials and Methods

STGD1 patients

Patient 1 is a 40-yr-old female who was diagnosed with typical STGD1 at the age of 16. Sanger sequence analysis of all 50 *ABCA4* exons and a MLPA analysis (MRC Holland P151 and P152) revealed heterozygous variants identified as two noncanonical splice site variants, the previously described c.859−9T>C and the novel c.303−3C>G. Patient 2 is a 37-yr-old male with STGD1 who was diagnosed at the age of 35 but had been suffering from reduced visual acuity since the age of 22. Sanger sequence analysis of all 50 *ABCA4* exons revealed the

coding variant c.70C>T [p.(Arg24Cys)] and a novel intronic variant 13-bp downstream from exon 13 (c.1937+13T>G). Patient 3 is a 53-yr-old female with STGD1 who was diagnosed at the age of 49. Sanger sequence analysis of all 50 *ABCA4* exons revealed a novel homozygous c.5461−8T>G variant. All three patients display extensive areas of retinal pigment epithelium atrophy throughout the posterior pole, extending beyond the vascular arcades. On electroretinography, the cone and rod responses were significantly and equally reduced in patients 2 and 3 at the age of diagnosis.

Generation of wild-type and mutant minigenes for exons 39–41, 40, 48, and 49

WT minigenes for exons 39–41, 40, 48, and 49 were generated by amplifying genomic DNA encompassing these exons from genomic DNA of an anonymous control person. Mutant minigenes for c.5714+5G>A and c.6729+5_+19del variants were generated by amplifying these regions from the genomic DNA of two patients carrying these *ABCA4* variants. Primer details can be found in **Tables S5** and **S6**. Minigenes were generated according to a previously described protocol (Sangermano et al. 2016).

Generation of *ABCA4* wild-type midigene library

The BACPAC Resource Center (BPRC) at the Children's Hospital Oakland Research Institute (Oakland, CA) provided the bacterial artificial chromosome clone, CH17-325O16 (insert g.94,434,639–94,670,492), containing the entire *ABCA4* gene. Using NucleoBond Xtra Midi EF (cat. no. 740420.250, Macherey-Nagel), BAC DNA was isolated and used as a template to generate large WT *ABCA4* fragments by PCR. PCR primers were designed using Primer3 software (http://bioinfo.ut.ee/primer3-0.4.0) and their specificity verified by using Human BLAT searches (https://genome.ucsc.edu/cgi-bin/hgBlat) (Kent 2002; **Table S6**). PCR reaction mixtures (50 µL) contained 0.5 µM of each forward and reverse primer, 0.2 mM dNTPs, Phusion High-Fidelity DNA Polymerase (cat. no. M0530L, New England Biolabs), 1× Phusion GC Buffer, 3% DMSO, and 2.5 ng BAC DNA. PCR reaction was performed by using the Veriti Thermal Cycler (Thermo Fisher Scientific). All PCRs were performed using the following thermocycling conditions: initial denaturation at 98°C for 30 sec; 15–20 cycles of denaturation at 98°C for 10 sec each, annealing at 57°C, and extension at 72°C (1 min/kb); with a final extension at 72°C for 15 min. The PCR-amplified products were analyzed on a 0.6% (w/v) agarose gel. Amplified bands were excised, and the NucleoSpin Gel and PCR cleanup kit (cat. no. 740609.250, Macherey-Nagel) was used to purify the DNA fragments, according to the manufacturer's protocol.

Two hundred nanograms of each purified fragment were cloned into the pDONR201 vector (Invitrogen) by using the Gateway BP Clonase Enzyme Mix (cat. no. 11789021, Thermo Fisher Scientific). The generated WT entry clones (denoted Bacterial Artificial Chromosome *ABCA4* [BA] clones) were digested with restriction enzymes and validated by

Sanger sequencing (**Table S7**), and each BA entry clone was transferred into the Gateway-adapted destination vector pCI-NEO-*RHO* as previously described (Sangermano et al. 2016).

Selection of candidate *ABCA4* noncanonical splice site variants

Of the 47 *ABCA4* NCSS variants investigated in this study, 43 were selected from the Leiden Open-source Variation Database (www.LOVD.nl/ABCA4) (Cornelis et al. 2017), and three were found in STGD1 patients (c.303−3C>G for patient 1, c.1937+13T>G for patient 2, c.5461−8T>G for patient 3) during routine diagnostic studies in the Radboudumc. c.2919−10T>C was a recently published novel NCSS variant (Schulz et al. 2017). Forty-six of the NCSS variants comply with the following three inclusion criteria: (1) located in intronic positions −14 to −3 (SAS), or +3 to +6 (SDS) or at the first or the last 2 nucleotides of exons; (2) a reduction in the strength of the mutant splice site compared to its WT reference when calculated by at least one of five algorithms (SpliceSiteFinder-like, MaxEntScan, NNSPLICE, GeneSplicer, Human Splicing Finder) (Shapiro and Senapathy 1987; Reese et al. 1997; Pertea et al. 2001; Yeo and Burge 2004; Desmet et al. 2009) via Alamut Visual software version 2.7 (Interactive Biosoftware; www.interactive- biosoftware.com); and (3) an allele frequency of ≤0.005 in the gnomAD browser (accessed May 31, 2017) (Lek et al. 2016). Variant c.1937+13T>G is not located in the SDS consensus sequence, and the five splice prediction programs did not show any reductions in canonical and downstream splice sites. It was selected for a splice assay as it was the only rare intronic *ABCA4* variant located outside the NCSS consensus sequences but within 20 bp of an exon junction that was identified in routine Sanger sequencing of ~300 STGD1 cases.

Generation of *ABCA4* mutant constructs

Site-directed mutagenesis was performed on WT BA entry clones as depicted in **Figure 2**. Mutagenesis primers were designed by using the online tool QuikChange Primer Design (http://www.genomics.agilent.com/primerDesignProgram.jsp?toggle=uploadNow&mutate =t rue&_requestid=1039517) and their GC content and temperature verified by using the online software (https://www.thermofisher.com/nl/en/home/brands/thermo-scientific/molecular- biology/molecular-biologylearning-center/molecular-biology-resource library/thermo scientificweb-tools/multiple-primer-analyzer.html) (**Table S8**).

The mutagenesis PCR reaction mixture (50 µL) contained 0.5 µM of each forward and reverse primer, 0.2 mM dNTPs, 1 U Phusion High-Fidelity DNA Polymerase (cat. no. M0530L, New England Biolabs), 1× Phusion HF Buffer, 2× Q-solution (cat. no. 1005485, Qiagen), and 20 ng WT entry clone. All PCRs were performed using the following thermocycling conditions: initial denaturation at 94°C for 5 min; 15 cycles of denaturation at 94°C for 30 sec each, annealing between 55°C and 75°C, and extension at 72°C (1 min/kb), with a final extension at 72°C for 20 min. PCR-amplified products were incubated with DpnI restriction enzyme (cat. no. R0176S, New England Biolabs) for 4.5 hrs, followed by heat inactivation at 65°C for 20

min to remove the WT plasmid DNA. Digested products were first run on a 0.6% (w/v) agarose gel to check for the presence of a unique linear fragment, corresponding to the mutant construct. Each of the NCSS variant mutants had their sequences validated (**Table S9**), and finally, each validated mutant construct was cloned into the Gateway-adapted destination vector pCI-NEO-RHO as previously described (Sangermano et al. 2016).

***In vitro* splice assay and transcriptional analysis**

To assess the effect of the NCSS variants on splicing, HEK293T cells (CRL-3216, ATCC) were transfected with WT and mutant constructs. Transfection, RNA isolation, and cDNA synthesis protocols were previously described (Sangermano et al. 2016). *ABCA4* transcript analysis was generally carried out by performing RT-PCR analysis using *ABCA4* exonic primers (**Table S10**). To test the effect of NCSS variants located in or near exons that could not be flanked by other *ABCA4* exons, e.g., due to the large size of some introns, primers were used from *RHO* exons 3 and 5. RT-PCR using primers targeting *RHO* exon 5 was used as a control for transfection efficiency (**Table S10**).

All reaction mixtures (50 µL) contained 0.2 µM of each primer pair, 1 U Taq DNA Polymerase (cat. no. 11647679001, Roche), 1×PCR buffer with MgCl2, 1 mM MgCl2, 0.2 mM dNTPs, 1× QSolution (cat. no. 1005485, Qiagen), and 50 ng cDNA. PCR conditions were a first denaturation step of 94°C for 5 min, followed by 35 cycles of melting (94°C for 30 sec each), annealing (58°C for 30 sec), and extension (72°C for 5 min) steps, with a final elongation step of 72°C for 2 min. The RT-PCR products were run on a 2% (w/v) agarose gel and the resulting bands were excised and purified with the NucleoSpin Gel & PCR cleanup kit (cat. no. 740609.250, Macherey-Nagel) according to the manufacturer's protocol. Finally, 100 ng of the purified nucleic acid were analyzed by Sanger sequencing in a 3100 or 3730 DNA Analyzer (Thermo Fisher Scientific). For complex RT-PCR band patterns, 4 µL of the RT-PCR products, containing between 1 and 2 µg of DNA, were further cloned via the pGEM-T Easy Vector System I (cat. no. A1360 Promega) according to the manufacturer's protocol.

To determine the fraction of correctly spliced product, quantitative analysis of the RT-PCR bands was performed using the Fragment Analyzer Auto Capillary Electrophoresis System (Advanced Analytical Technologies, Inc.). RT-PCR of three constructs with no splicing defects, 16 constructs with partial splicing defects, and three constructs with no correctly spliced RNAs were performed in duplicate. Human retina cDNA was used as the control, as several 'natural' exon skipping event in human retina have been reported previously (**Tables S3** and **S11**; **Figures S5** and **S6**; Braun et al. 2013). For variants c.1937+13G>T, c.2382+5G>C, and c.2588G>C, the fraction of correctly spliced product was assessed by densitometric analysis using Image J software (Schneider et al. 2012). The analysis was performed in duplicate and values were normalized for the size of the different fragments present in each RT-PCR product mixture (**Table S12**).

ABCA4 transcript analysis of patient-derived photoreceptor precursor cells

Fibroblasts from a patient carrying the noncanonical splice site variant c.768G>T in a homozygous state were successfully reprogrammed in iPSCs and further differentiated into photoreceptor precursor cells as previously described (Sangermano et al. 2016). Photoreceptor precursor cells were harvested, total RNA extracted, and RT-PCR performed by using a forward primer located in *ABCA4* exon 4 (5′-TCATGAATGCACCAGAGAGC-3′) and a reverse primer located in *ABCA4* exon 7 (5′-ACGGCTGTCTAGGAGTGTGG-3′). Gel resolution, excision, and purification of the bands and Sanger sequencing were performed as previously described (Sangermano et al. 2016).

Data access

The raw sequencing data for all WT splicing constructs (WT_BA), mutant splicing BA constructs (BA_c.), and RT-PCR fragments (RT_BA) from splice assays in HEK293T cells from this study have been submitted to the NCBI BioProject (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA417900).

Acknowledgments

We thank Erwin van Wijk for providing the Gateway-adapted minigene vector and Christian Gilissen for assistance in data processing. This work was supported by the FP7-PEOPLE-2012-ITN (Marie-Curie Actions) program EyeTN, agreement 317472 (to F.P.M.C.). This work was also supported by the Foundation Fighting Blindness USA, grant no. PPA-0517-0717-RAD (to F.P.M.C., R.W.J.C., S.A., and A.G.), RP Fighting Blindness, UK, grant no. GR591 (to F.P.M.C. and S.A.), the Macula Vision Research Foundation (to F.P.M.C.), the Rotterdamse Stichting Blindenbelangen, the Stichting Blindenhulp, the Stichting tot Verbetering van het Lot der Blinden (to F.P.M.C. and S.A.), and by the Landelijke Stichting voor Blinden en Slechtzienden, Macula Degeneratie fonds and the Stichting Blinden-Penning that contributed through Uitzicht (to F.P.M.C. and S.A.). None of the authors have any financial interest or conflicting interest to disclose.

Supplemental data

Supplemental data are not included in this thesis and they can be found online at the following link: http://genome.cshlp.org/content/suppl/2017/12/07/gr.226621.117.DC1

References

Acedo A, Hernández-Moro C, Curiel-García Á, Díez-Gómez B, Velasco EA. 2015. Functional classification of BRCA2 DNA variants by splicing assays in a large minigene with 9 exons. *Hum Mutat* **36:** 210-221.

Allikmets R, Singh N, Sun H, Shroyer NF, 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.

Aukrust I, Jansson RW, Bredrup C, Rusaas HE, Berland S, Jorgensen A, Haug MG, Rodahl E, Houge G, Knappskog PM. 2017. The intronic ABCA4 c.5461-10T>C variant, frequently seen in patients with Stargardt disease, causes splicing defects and reduced ABCA4 protein level. *Acta Ophthalmol* **95:** 240-246.

Baralle M, Skoko N, Knezevich A, De Conti L, Motti D, Bhuvanagiri M, Baralle D, Buratti E, Baralle FE. 2006. NF1 mRNA biogenesis: effect of the genomic milieu in splicing regulation of the NF1 exon 37 region. *FEBS Lett* **580**: 4449-4456.

Blacharski PA. 1998. Fundus flavimaculatus. In *Retinal dystrophies and degenerations* (ed. Newsome DA), pp. 135-159. Raven Press, New York.

Bonifert T, Gonzalez Menendez I, Battke F, Theurer Y, Synofzik M, Schols L, Wissinger B. 2016. Antisense oligonucleotide mediated splice correction of a deep intronic mutation in OPA1. *Mol Ther Nucleic Acids* **5:** e390.

Braun TA, Mullins RF, Wagner AH, Andorf JL, Johnston RM, Bakall BB, Deluca AP, Fishman GA, Lam BL, Weleber RG, et al. 2013. Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. *Hum Mol Genet* **22:** 5136-5145.

Buratti E, Chivers M, Hwang G, Vorechovsky I. 2011. DBASS3 and DBASS5: databases of aberrant 3'- and 5'-
splice sites. *Nucleic Acids Res* **39:** D86-91.

Carss KJ, Arno G, Erwood M, Stephens J, Sanchis-Juan A, Hull S, Megy K, Grozeva D, Dewhurst E, Malka S et al. 2017. Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. *Am J Hum Genet.* **100**: 334-342.

Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. 2003. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* **31:** 3568-3571.

Cheung R, Insigne KD, Yao D, Burghard CP, Jones EM, Goodman DB, Kosuri S. 2017. Large-scale screening of rare genetic variants in humans reveals frequent splicing disruptions. *BioRxiv* doi: http://dx.doi.org/10.1101/199927

Collin RW, den Hollander AI, van der Velde-Visser SD, Bennicelli J, Bennett J, Cremers FP. 2012. Antisense oligonucleotide (AON)-based therapy for leber congenital amaurosis caused by a frequent mutation in CEP290. *Mol Ther Nucleic Acids* **1:** e14.

Cornelis SS, Bax NM, Zernant J, Allikmets R, Fritsche LG, den Dunnen JT, Ajmal M, Hoyng CB, Cremers FP. 2017. In silico functional meta-analysis of 5,962 ABCA4 variants in 3,928 retinal dystrophy cases. *Hum Mutat* **38:** 400-408.

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, 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.

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:** e67.

Drivas TG, Wojno AP, Tucker BA, Stone EM, Bennett J. 2015. Basal exon skipping and genetic pleiotropy: a predictive model of disease pathogenesis. *Sci Transl Med* **7:** 291-297.

Garanto A, Chung DC, Duijkers L, Corral-Serrano JC, Messchaert M, Xiao R, Bennett J, Vandenberghe LH, Collin RW. 2016. In vitro and in vivo rescue of aberrant splicing in CEP290-associated LCA by antisense oligonucleotide delivery. *Hum Mol Genet* **25:** 2552-2563.

Gerard X, Perrault I, Hanein S, Silva E, Bigot K, Defoort-Delhemmes S, Rio M, Munnich A, Scherman D, Kaplan J, et al. 2012. AON-mediated exon skipping restores ciliation in fibroblasts harboring the common Leber congenital amaurosis CEP290 mutation. *Mol Ther Nucleic Acids* **1**: e29.

Kent WJ. 2002. BLAT—The BLAST-like Alignment Tool. *Genome Res* **12:** 656-664.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536:** 285-291.

Matlin AJ, Clark F, Smith CW. 2005. Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol* **6:** 386-398.

Maugeri A, van Driel MA, van de Pol DJ, Klevering BJ, van Haren FJ, Tijmes N, Bergen AA, Rohrschneider K, Blankenagel A, Pinckers AJ, 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.

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:** 960-966.

Molday LL, Rabin AR, Molday RS. 2000. ABCR expression in foveal cone photoreceptors and its role in Stargardt macular dystrophy. *Nat Genet* **25:** 257-258.

Murphy D, Cieply B, Carstens R, Ramamurthy V, Stoilov P. 2016. The Musashi 1 controls the splicing of photoreceptor-specific exons in the vertebrate retina. *PLoS Genet* **12:** e1006256.

Osoegawa K, de Jong PJ. 2004. BAC library construction. *Methods Mol Biol* **255:** 1-46.

Parfitt DA, Lane A, Ramsden CM, Carr AJ, Munro PM, Jovanovic K, Schwarz N, Kanuga N, Muthiah MN, Hull S, et al. 2016. Identification and correction of mechanisms underlying inherited blindness in human iPSC- derived optic cups. *Cell Stem Cell* **18:** 769-781.

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

Quazi F, Molday RS. 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.

Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. *J Comput Biol* **4:** 311- 323.

Rivera A, White K, Stohr H, Steiner K, Hemmrich N, Grimm T, Jurklies B, Lorenz B, Scholl HP, 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.

Rosenberg AB, Patwardhan RP, Shendure J, Seelig G. 2015. Learning the sequence determinants of alternative splicing from millions of random sequences. *Cell* **163:** 698-711.

Sakharkar MK, Chow VT, Kangueane P. 2004. Distributions of exons and introns in the human genome. 2004. *In Silico Biol.* **4**: 387-393.

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, et al. 2016. Photoreceptor

progenitor mRNA analysis reveals exon skipping resulting from the ABCA4 c.5461-10T-->C mutation in Stargardt disease. *Ophthalmology* **123:** 1375-1385.

Sangermano R, Garanto A, Khan M, Runhart EH, Bauwens M, Bax NM, van den Born LI, Khan MI, Cornelis SS, Verheij J et al. 2019. Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. *Genet Med* doi:10.1038/s41436-018-0414-9.

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9:**671-675.

Schulz HL, Grassmann F, Kellner U, Spital G, Ruther K, Jagle H, Hufendiek K, Rating P, Huchzermeyer C, Baier MJ, et al. 2017. Mutation spectrum of the ABCA4 gene in 335 Stargardt disease patients from a multicenter German cohort-impact of selected deep intronic variants and common SNPs. *Invest Ophthalmol Vis Sci* **58:** 394-403.

Shapiro MB, Senapathy P. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* **15:** 7155-7174.

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:** 2757-2761.

Slijkerman RW, Vache C, Dona M, Garcia-Garcia G, Claustres M, Hetterschijt L, Peters TA, Hartel BP, Pennings RJ, Millan JM, et al. 2016. Antisense oligonucleotide-based splice correction for USH2A-associated retinal degeneration caused by a frequent deep-intronic mutation. *Mol Ther Nucleic Acids* **5:** e381.

Smith PJ, Zhang C, Wang J, Chew SL, Zhang MQ, Krainer AR. 2006. An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum Mol Genet* **15:** 2490-2508.

Suarez T, Biswas SB, Biswas EE. 2002. Biochemical defects in retina-specific human ATP binding cassette

transporter nucleotide binding domain 1 mutants associated with macular degeneration. *J Biol Chem* **277:** 21759-21767.

Sun H, Nathans J. 1997. Stargardt's ABCR is localized to the disc membrane of retinal rod outer segments. *Nat Genet* **17:** 15-16.

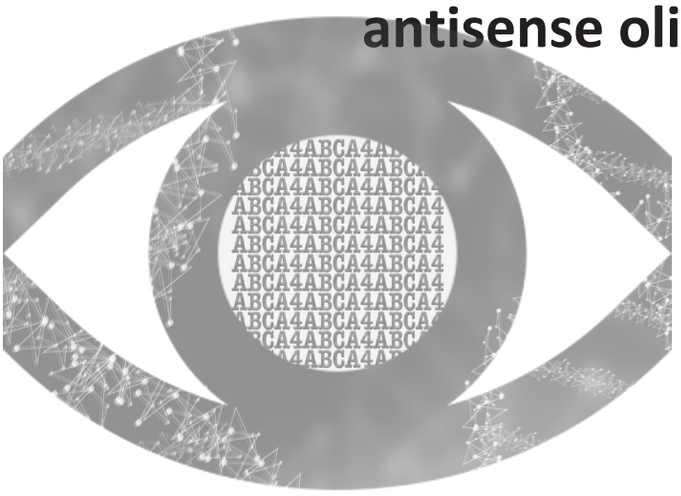
Sun H, Smallwood PM, Nathans J. 2000. Biochemical defects in ABCR protein variants associated with human retinopathies. *Nat Genet* **26:** 242-246.

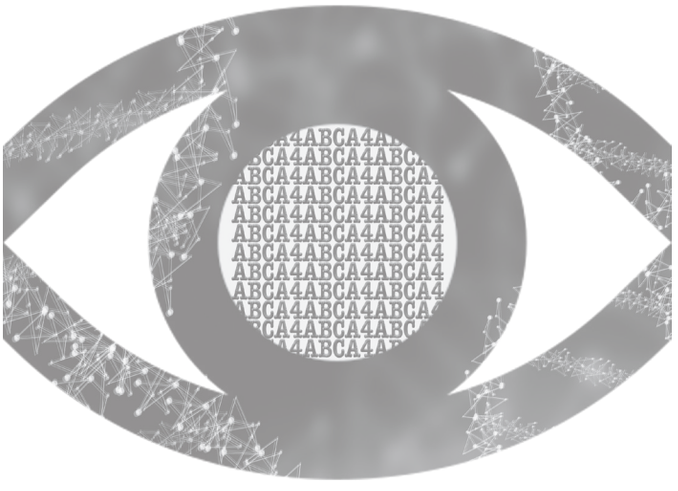
Wiszniewski W, Zaremba CM, Yatsenko AN, Jamrich M, Wensel TG, Lewis RA, Lupski JR. 2005. ABCA4 mutations causing mislocalization are found frequently in patients with severe retinal dystrophies. *Hum Mol Genet* **14:** 2769-2778.

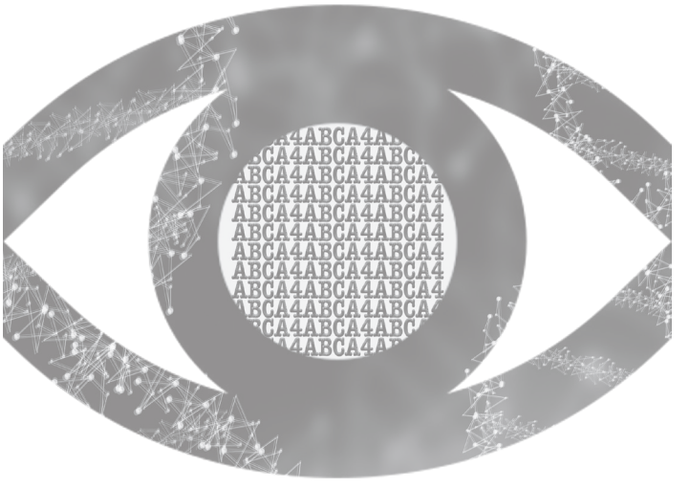
Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* **11:** 377-394.

Chapter 3

Deep-intronic *ABCA4* variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides







Deep-intronic *ABCA4* variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides

Riccardo Sangermano*, Alejandro Garanto*, **Mubeen Khan***, Esmee H. Runhart, Miriam Bauwens, Nathalie M. Bax, L. Ingeborgh van den Born, Muhammad Imran Khan, Stéphanie S. Cornelis, Joke B. G. M. Verheij, Jan-Willem R. Pott, Alberta A. H. J. Thiadens, Caroline C. W. Klaver, Bernard Puech, Isabelle Meunier, Sarah Naessens, Gavin Arno, Ana Fakin, Keren J. Carss, F. Lucy Raymond, Andrew R. Webster, Claire-Marie Dhaenens, Heidi Stöhr, Felix Grassmann, Bernhard H. F. Weber, Carel B. Hoyng, Elfride De Baere, Silvia Albert, Rob W. J. Collin# and Frans P. M. Cremers#

Genet Med. 2019; **21**: 1751-1760.

*These authors contributed equally to this work

#Senior authors contributed equally

Abstract

Purpose: Using exome sequencing, the underlying mutations in many persons with autosomal recessive diseases remain undetected. We explored autosomal recessive Stargardt disease (STGD1) as a model to identify the missing heritability.

Methods: Sequencing of *ABCA4* was performed in eight STGD1 cases with one variant and p.Asn1868Ile in *trans*, 25 cases with one variant, and three cases with no *ABCA4* variant. The effect of intronic variants was analyzed using *in vitro* splice assays in HEK293T cells and patient-derived fibroblasts. Antisense oligonucleotides were used to correct splicing defects.

Results: In 24 of the probands (67%), one known and five novel deep-intronic (DI) variants were found. The five novel variants resulted in mRNA pseudoexon inclusions, due to strengthening of cryptic splice sites or by disrupting a splicing silencer motif. Variant c.769-784C>T showed partial insertion of a pseudoexon and was found in *cis* with c.5603A>T (p.Asn1868Ile), so its causal role could not be fully established. Variant c.4253+43G>A resulted in partial skipping of exon 28. Remarkably, antisense oligonucleotides targeting the aberrant splice processes resulted in correction of all splicing defects.

Conclusions: Our data demonstrate the importance of assessing non-coding variants in genetic diseases, and show the great potential of splice modulation therapy for DI variants.

Introduction

In the last decade, next-generation sequencing (NGS) has enabled a rapid identification of novel disease genes (Ng et al. 2009; Nikopoulos et al. 2010; Carss et al. 2017), and transformed molecular diagnostic testing (Abu-Safieh et al. 2013; Beryozkin et al. 2015; Haer-Wigman et al. 2017). Thus far, the emphasis was on coding and flanking splice site sequences that harbor the majority of genetic defects. Herewith, the underlying defects could be identified in ~50% of individuals with genetically heterogeneous diseases, including inherited retinal diseases (IRDs) (Neveling et al. 2012; Neveling et al. 2013; Gilissen et al. 2014). Genome sequencing can identify the majority of non-coding sequence variants (Carss et al. 2017), but it will be very challenging to pinpoint the causal variants in the absence of assays that reveal the resulting mRNA and/or protein defects.

For diseases that are due to mutations in a single gene, the yield of mutations generally exceeds 50%. This is true for Stargardt disease (STGD1) (MIM 248200) which is caused by mutations in the gene encoding the ATP binding cassette type A4 (*ABCA4*) (MIM 601691) (Allikmets et al. 1997), as biallelic variants can be found in the majority of the cases (Zernant et al. 2014) and in a smaller proportion of other IRD subtypes (Cremers et al. 1998; Maugeri et al. 2000; Fakin et al. 2016). In ~30% of STGD1 cases however, the expected second *ABCA4* variant was lacking. A frequent coding variant, p.Asn1868Ile, was present in ~10% of STGD1 cases with a severe mutation in *trans* and who generally showed late-onset disease (Zernant et al. 2017; Runhart et al. 2018). The remaining mutations in *ABCA4* rarely consist of heterozygous deletions (Maugeri et al. 1999; Yatsenko et al. 2003; Bax et al. 2015), whereas DI *ABCA4* variants also account for some of the missing heritability in STGD1 (Braun et al. 2013; Zernant et al. 2014; Bauwens et al. 2015; Bax et al. 2015). By employing STGD1-derived keratinocytes, it was shown that some of these variants strengthen cryptic splice sites or create new splice sites, ultimately resulting in pseudoexon (PE) inclusions (Braun et al. 2013). Using STGD1-derived photoreceptor precursor cells (PPCs), we recently showed that two intronic variants in *ABCA4* (c.4539+2001G>A and c.4539+2028C>T) strengthen exonic enhancer elements (ESEs), resulting in the inclusion of a 345-nt PE (Albert et al. 2018).

PE inclusions due to DI variants are attractive targets for antisense oligonucleotide (AON)-based splicing correction (Hammond and Wood 2011). In the field of IRDs, these molecules have shown therapeutic potential for DI mutations in *CEP290* (Collin et al. 2012; Gerard et al. 2012; Han et al. 2012; Garanto et al. 2016; Parfitt et al. 2016), *USH2A* (Slijkerman et al. 2016) or *OPA1* (Bonifert et al. 2016), and a phase 1/2 clinical trial using AONs to treat *CEP290*-associated Leber congenital amaurosis is ongoing (NCT03140969).

In this study, we employed STGD1 as a model to assess the challenges to identify pathogenic non-coding variants. We sequenced the *ABCA4* locus in 36 STGD1 individuals with one *ABCA4* variant and p.Asn1868Ile in *trans* (n=8), one *ABCA4* variant (n=25), or no (n=3) *ABCA4* variants. Selected non-coding variants were tested by *in vitro* splice assays or using STGD1-derived fibroblasts. Subsequently, several AONs were designed and successfully applied to redirect erroneous splicing.

Results

Identification and selection of deep-intronic variants for splicing assay

Haloplex sequencing of eight STGD1 cases carrying one causal *ABCA4* variant and p.Asn1868Ile in *trans*, 22 STGD1 cases with only one, and two cases carrying no *ABCA4* variant, yielded 220 rare *ABCA4* variants (allele frequency (AF) ≤0.01 in control individuals), consisting of 85 independent novel variants (**Table S3**). In the four British STGD1 cases with one (n=3) or no (n=1) *ABCA4* variant analyzed by genome sequencing, eight novel rare (AF ≤0.01) independent *ABCA4* variants remained (**Table S3**).

Eleven non-coding variants were selected for splice assays (**Table S4**). Seven of these (c.768+7329A>G, c.858+526T>G, c.859-506G>C, c.1555-5008C>T, c.4539+1100A>G, c.4539+1106C>T, c.6148-421T>C) were selected as they were located at cryptic splice sites or created a new putative splice site with a relative strength of at least 75% of the maximal score in two of five algorithms, and showed an increased splice prediction score of at least 2%. Two variants, c.769-784C>T and c.4253+43G>A, belonging to three alleles (c.[769-784C>T;5603A>T], c.[4253+43G>A;6006-609T>A], c.[4253+43G>A;5603A>T]), were enriched among the mono-allelic cases, i.e. in seven and nine of 36 probands, respectively. The c.4253+43G>A variant was reported previously (Zernant et al. 2011). Finally, the c.1937+435C>G variant was selected because it was located in the genomic sequence of an alternate exon in intron 13 (Chr1: 94,527,737-94,527,644) (Braun et al. 2013).

Four novel deep-intronic variants result in PE generation through the strengthening of splice sites

To test the functional effect of the 11 selected non-coding variants, seven *ABCA4* wild-type (WT) midigene constructs previously described as BA4, BA6, BA7, BA9, BA11, BA19 and BA21 (**Chapter 2**) and the newly designed BA30 (**Table S5**), were mutagenized. Upon RT-PCR and Sanger sequence validation, four variants had no effect (**Figure S1**). Variant c.858+526T>G did not show a PE insertion, but did yield four weak smaller fragments (**Figure S1**), the relevance of which is unknown in the absence of sequence

results. This variant was found in *cis* with a complex allele containing a protein-truncating mutation (p.[Trp273*;Asn1868Ile]) in individual E-II:1, as determined by segregation analysis. The remaining six variants resulted in clear splicing defects (**Figure 1** and **Figure S2**) that allowed us to detect the second variant in 15/26 mono-allelic probands, two DI variants in 2/3 cases without *ABCA4* variants, and a complex allele (c.[769-784C>T;5603A>T]) in seven STGD1 cases carrying one DI variant. Considering that these variants or complex alleles are causal, segregation was as expected (**Table 1**). The remaining 12 unsolved STGD1 cases are listed in **Table S6**.

The DI variant c.769-784C>T, present in 7/36 STGD1 cases, is located in intron 6, and was predicted to increase the strength of an intronic splice acceptor site (SAS) at position g.94,549,775. In addition, a strong splice donor site (SDS) was situated 161 nt downstream. RT-PCR analysis using WT and mutant minigene constructs revealed that, compared to the WT construct, the mutant c.769-784C>T construct showed an additional band of 296 bp, containing a 162-nt PE (**Figure 1B** and **Figure S2**). Quantification of mutant mRNA in transfected HEK293T cells revealed that the PE fragment only accounted for ~8% of the total *ABCA4* pre-mRNA (**Table S7**), suggesting a mild effect at the RNA and thus protein level (r.[=, 768_769ins(162)]; p.[=,Leu257Aspfs*3]).

The DI variant c.859-506G>C was predicted to significantly strengthen a cryptic intronic SAS and is accompanied by a downstream SDS. This variant resulted in a 56-nt PE insertion (**Figure 1B** and **Figure S2**), which accounted for 75.5% of total splice products (**Table S7**), indicating the highly deleterious effect of this variant at the RNA and (predicted) protein level (r.[858_859ins(56),=]; p.[Phe287Thrfs*32,=]).

DI variants c.4539+1100A>G and c.4539+1106C>T were predicted to alter the same cryptic SDS located at position g.94,493,901. A very strong SAS was located at g.94,493,968. Besides the expected 68-nt PE fragment (PE30.1) due to the SAS at position c.4539+1033 and the SDS at position c.4539+1100, an additional 112-nt PE (PE30.2) carrying the same splice donor but another acceptor site located at g.94,494,012 was observed (**Figure 1B** and **Figure S2**). Interestingly, the relative ratios of these two PEs significantly differed between the midigene constructs carrying c.4539+1100A>G or c.4539+1106C>T. PE30.1 accounted for 60.6% and 93.1% of the total transcript in c.4539+1100A>G or c.4539+1106C>T respectively, while PE30.2 was present in 20.1% and 3.8% of the cDNA products, respectively (**Table S7**). As both c.4539+1100A>G and c.4539+1106C>T showed only a minor fraction of correctly spliced products (19.2% and 3.0%) and because PE30.1 and PE30.2 are both predicted to lead to a frameshift, these variants were both deemed to have a severe effect (r.[4539_4540ins(68), 4539_4540ins(112),=]; p.[Arg1514Glyfs*3,Arg1514Valfs*31,=]) (**Table 1** and **Table S8**).

c.1937+435C>G results in PE generation through the disruption of putative splice silencers

The c.1937+435C>G variant did not alter the strength of a cryptic splice site. Instead, it was predicted to disrupt an ESE (SC35) motif, and, more importantly, to inactivate two splice silencer motifs and to lower the predicted strength of a third one (**Figure S3**). This variant was located within the genomic sequence of a low-abundance *ABCA4* transcript that contains intron 13 sequences (Braun et al. 2013). RT-PCR analysis revealed a 134-nt PE insertion between positions c.1937+396 and c.1937+529 in mutant BA11 compared to WT (**Figure 1B** and **Figure S2**). Although the splicing defect caused by this variant disrupted the reading frame (r.[=,1937_1938ins(134)]; p.[=,Ser646Serfs*25]), the fraction of mutant transcript was low (8.6%) (**Table S7**).

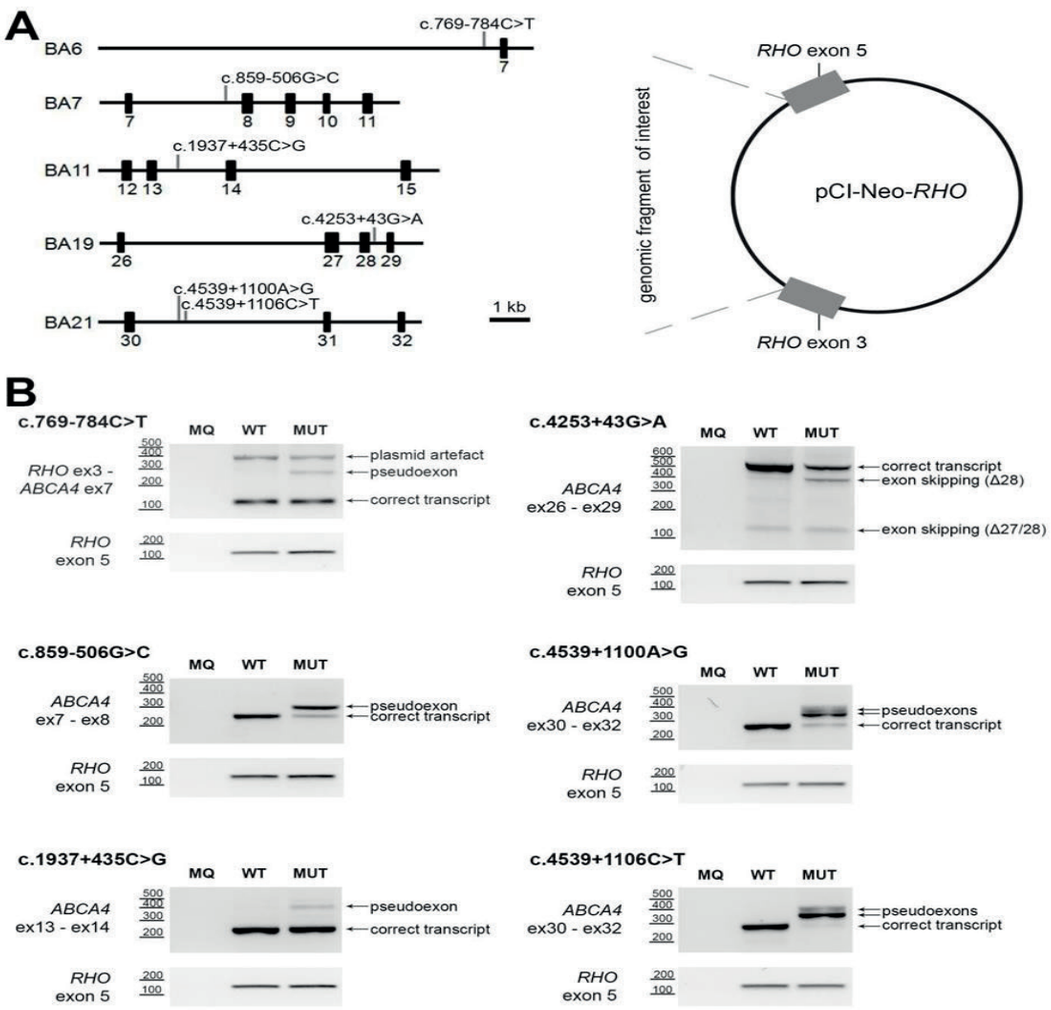


Figure 1. Generation and assessment of splicing defects using midigenes. A. Schematic representation of the five wild-type (WT) midigenes used that were cloned between exon 3 and 5 of Rhodopsin in pCI-Neo-*RHO*. Position of the mutations present in six mutant midigenes is indicated. **B.** Assessment of the splicing defects upon midigene transfection in HEK293T cells. Five PE inclusions and one exon skipping were detected in the mutant (MUT) constructs

compared to the WT. MQ stands for the negative control of the PCR reaction. Rhodopsin (*RHO)* amplification was used as a transfection and loading control.

Variant c.4253+43G>A results in exon skipping through the disruption of putative splice silencers

The c.4253+43G>A variant was present in eight mono-allelic STGD1 individuals and in one proband carrying a causal variant and p.Asn1868Ile in *trans*. It does not weaken the SDS of exon 28, nor does it affect any other cryptic splice sites *in silico*. It does affect the strength of several splicing enhancer or silencer motifs (**Figure S4**). The c.4253+43G>A variant was found in *cis* with the previously reported c.6006-609T>A (Zernant et al. 2014) in eight of nine probands, but c.6006-609T>A showed no effect in an *in vitro* splice assay (**Figure S2**). The c.4253+43G>A mutant yielded bands of 495 and 370 bp, the first corresponding to the correctly spliced product, and the second to a product in which exon 28 was skipped (**Figure 1B** and **Figure S2**). This 125-nt deletion is predicted to lead to a frameshift, and although both WT and mutant midigenes showed some skipping of exon 28, in the mutant construct this was significantly higher compared to WT (26.1% *vs*. 2.8%, **Table S7**). The fact that in all nine probands, a severe variant was found in *trans*, in combination with a late age of onset, suggests a mild effect of this variant (r.[=,4129_4253del(125)]; p.[=,Ile1377Hisfs*3]). An overview of the allele frequencies and of the splicing defects caused by all six variants can be found in **Table S8** and **Table S9**.

Table 1: Persons with STGD1 carrying two pathogenic *ABCA4*

Patient_ID	Gender	Age at onset (yrs)	Segregation confirmed	Allele1 DNA	Allele1 protein	Allele 2 DNA	Allele 2 protein
A-I:2	F	38	Yes	c.1822T>A	p.(Phe608Ile)	c.[769-784C>T; 5603A>T]	p.[=, Leu257Aspfs*3; Asn1868Ile][4]
B-I:2	F	10	Yes	c.768G>T	p.(Leu257Valfs*17)[1]	c.859-506G>C	p.[Phe287Thrfs*32, =][4]
C-I:2	F	48	Yes	c.768G>T	p.(Leu257Valfs*17)[1]	c.[769-784C>T; 5603A>T]	p.[=,Leu257Aspfs*3; Asn1868Ile][4]
D-III:1	M	45	Yes	c.4363T>C	p.(Cys1455Arg)	c.[4253+43G>A; 6006-609T>A]	p.[=, Ile1377Hisfs*3][4]
E-II:1	F	7	Yes	c.[818G>A; 5603A>T]	p.[Trp273*; Asn1868Ile]	c.4539+1100A>G	p.[Arg1514Valfs*31, Arg1514Glyfs*3, =][4]
F-II:2	F	20	Yes	c.1822T>A	p.(Phe608Ile)	c.[4253+43G>A; 6006-609T>A]	p.[=, Ile1377Hisfs*3][4]
G-I:2	F	51	Yes	c.1822T>A	p.(Phe608Ile)	c.[4253+43G>A; 6006-609T>A]	p.[=, Ile1377Hisfs*3][4]
H-I:2	F	53	Yes	c.[5461-10T>C; 5603A>T]	p.[Thr1821Valfs*13, Thr1821Aspfs*6; Asn1868Ile]	c.[4253+43G>A; 6006-609T>A]	p.[=, Ile1377Hisfs*3][4]
I-II:3	F	44	Yes	c.4577C>T	p.(Thr1526Met)	c.[769-784C>T; 5603A>T]	p.[=, Leu257Aspfs*3; Asn1868Ile][4]
J-II:3	M	61	Yes	c.768G>T	p.(Leu257Valfs*17)[1]	c.[4253+43G>A; 6006-609T>A]	p.[=, Ile1377Hisfs*3][4]
K-II:1	M	62	Yes	c.6155del	p.(Asn2052Thrfs*9)	c.[769-784C>T; 5603A>T]	p.[=, Leu257AspGln*; Asn1868Ile][4]
L-II:1	M	61	Yes	c.[5461-10T>C; 5603A>T]	p.[Thr1821Valfs*13, Thr1821Aspfs*6; Asn1868Ile]	c.[769-784C>T; 5603A>T]	p.[=, Leu257AspGln*; Asn1868Ile][4]
M-II:1	M	18	Yes	c.4539+2001G>A (V4)	p.[=, Arg1514Leufs*36][2]	c.[4253+43G>A; 5603A>T]	p.[=, Ile1377Hisfs*3; Asn1868Ile][4]
N-II:3	M	31	Yes	c.4773+1G>A	p.(?)	c.[4253+43G>A; 6006-609T>A]	p.[=, Ile1377Hisfs*3][4]
O-I:1	M	49	Yes	c.3113C>T	p.(Ala1038Val)	c.859-506G>C	p.[Phe287Thrfs*32, =][4]
P-II:3	F	4	Yes	c.5196+1137G>A (V1)	p.[Met1733Glufs*78, =][3]	c.859-506G>C	p.[Phe287Thrfs*32, =][4]
Q-II:1	M	69	Yes	c.4539+1G>T	p.(?)	c.[769-784C>T; 5603A>T]	p.[=, Leu257Aspfs*3; Asn1868Ile][4]
R-II:1	M	52	n.t.	c.4539+1G>T	p.(?)	c.[4253+43G>A; 6006-609T>A]	p.[=, Ile1377Hisfs*3][4]
S-II:1	F	64	n.t.	c.768G>T	p.(Leu257Valfs*17)[1]	c.[769-784C>T; 5603A>T]	p.[=, Leu257Aspfs*3; Asn1868Ile][4]
T-II:1	F	35	Yes	c.[2588G>C; 5603A>T]	p.[Gly863Ala, Gly863del; Asn1868Ile]	c.1937+435C>G	p.[=, Ser646Serfs*25][4]
U-II:1	M	9	n.t.	c.768G>T	p.(Leu257Valfs*17)[1]	c.1937+435C>G	p.[=, Ser646Serfs*25][4]
V-II:1	F	15	n.t.	c.[2588G>C; 5603A>T]	p.[Gly863Ala, Gly863del; Asn1868Ile]	c.4539+1100A>G	p.[Arg1514Valfs*31, Arg1514Glyfs*3, =][4]
W-II:1	n.a.	11*	n.t.	c.[1622T>C; 3113C>T]	p.[Leu541Pro; Ala1038Val]	c.4539+1106C>T	p.[Arg1514Valfs*31, Arg1514Glyfs*3][4]
X-II:1	n.a.	52*	n.t.	c.4469G>A	p.(Cys1490Tyr)	c.[4253+43G>A; 6006-609T>A]	p.[=, Ile1377Hisfs*3][4]

*Unknown age of onset. Current age; n.t., not tested; [1]Effect based on Sangermano et al. 2018; [2]Effect based on Albert et al. 2018; [3]Predicted based on Braun et al. 2013; [4]Predicted based on densitometry percentages in this study based on the method of which was described in Sangermano et al. 2018.

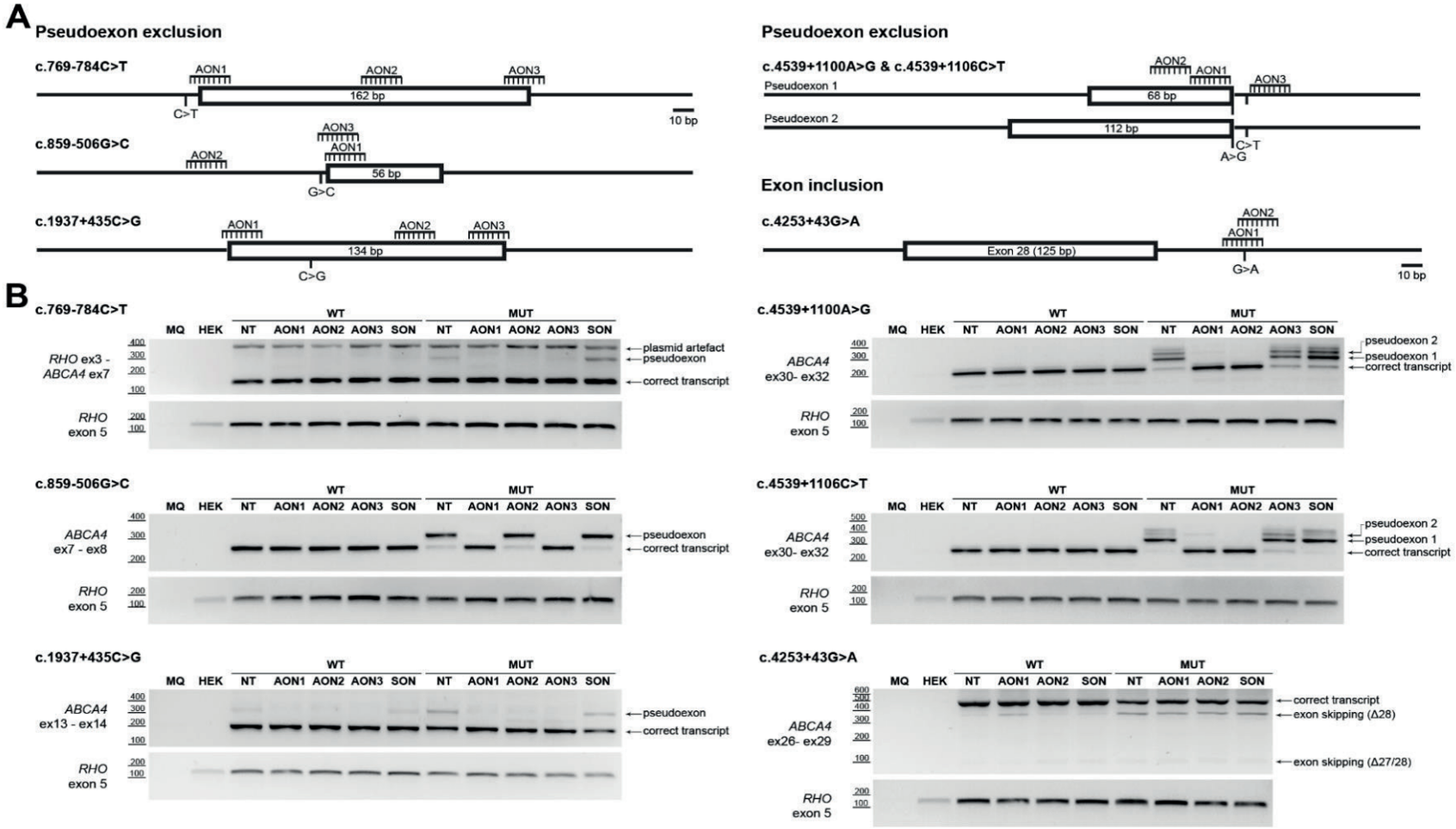

A
Pseudoexon exclusion
c.769-784C>T
AON1
AON2
AON3
162 bp
C>T
10 bp
c.859-506G>C
AON3
AON2
AON1
56 bp
G>C
c.1937+435C>G
AON1
AON2
AON3
134 bp
C>G
Pseudoexon exclusion
c.4539+1100A>G & c.4539+1106C>T
AON2
AON1
AON3
Pseudoexon 1
68 bp
Pseudoexon 2
112 bp
C>T
A>G
Exon inclusion
c.4253+43G>A
AON2
AON1
Exon 28 (125 bp)
G>A
10 bp
B
c.769-784C>T
WT
MUT
MQ HEK NT AON1 AON2 AON3 SON NT AON1 AON2 AON3 SON
RHO ex3 - ABCA4 ex7
plasmid artefact
pseudoexon
correct transcript
RHO exon 5
c.859-506G>C
ABCA4 ex7 - ex8
pseudoexon
correct transcript
RHO exon 5
c.1937+435C>G
ABCA4 ex13 - ex14
pseudoexon
correct transcript
RHO exon 5
c.4539+1100A>G
ABCA4 ex30- ex32
pseudoexon 2
pseudoexon 1
correct transcript
RHO exon 5
c.4539+1106C>T
ABCA4 ex30- ex32
pseudoexon 2
pseudoexon 1
correct transcript
RHO exon 5
c.4253+43G>A
MQ HEK NT AON1 AON2 SON NT AON1 AON2 SON
ABCA4 ex26- ex29
correct transcript
exon skipping (Δ28)
exon skipping (Δ27/28)
RHO exon 5

Figure 2. Antisense oligonucleotide rescue using midigenes in HEK293T cells. A. Position of the AONs and sizes of the pseudoexons (PEs) introduced by the mutations. **B.** Assessment of the splicing correction by RT-PCR upon AON delivery. Wild-type (WT) midigenes and mutant (MUT) midigenes were co-transfected with different AONs and a SON, except the non- transfected cells lane (NT). For each mutation introducing a PE, at least one AON was able to redirect splicing. In the case of exon 28 skipping, a 10% inclusion was achieved with one of the AONs. MQ denotes the negative control of the PCR reaction and HEK for the untransfected HEK293T. Rhodopsin (*RHO)* amplification was used as a transfection and loading control.

Single molecule molecular inversion probe-based sequence analysis of intronic *ABCA4* regions

Five out of six intronic *ABCA4* variants were analyzed for their recurrence in 412 genetically unsolved STGD1 cases from France (n=224) and Germany (n=188), by employing single molecule Molecular Inversion Probe (smMIPs) (Hiatt et al. 2013). Three variants were found in a heterozygous state in this cohort: c.769-784C>T (n=4), c.4253+43G>A (n=29), and c.4539+1106C>T (n=1) (**Table S8**), whereas variants c.859-506G>C and c.4539+1100A>G were not detected.

Antisense oligonucleotide-based exclusion of PEs

For five variants (c.769-784C>T, c.859-506G>C, c.1937+435C>G, c.4539+1100A>G and c.4539+1106C>T), we designed three different AONs each that aim to result in PE exclusion. For c.4253+43G>A, only two AONs were designed, aiming to block splice silencer motifs to reinforce inclusion of exon 28. A summary of the sequences and characteristics of the AONs is provided in **Table S2**. The effect of the AONs at the RNA level was assessed by RT-PCR and subsequently semi-quantified using capillary analysis (**Figure S5** and **Table S7**).

For c.769-784C>T, AON2 completely restored correct splicing, whereas AON3 had a partial effect (**Figure 2B**). When using AON1, also some splice redirection was observed, an additional faint band was present both in WT and mutant. Cloning and sequence analysis of this band revealed that the 5' 68 nt of the PE were skipped, while the 3' 94 nt remained. Given the position of AON1, this AON most likely blocks the SAS of the PE, triggering the use of a downstream SAS. For c.859-506G>C, two AONs (AON1 and AON3) corrected the splicing defect (**Figure 2B**). For c.1937+435C>G, all three AONs restored normal splicing (**Figure 2B**). In addition, we observed that the 134-nt PE could also naturally be included in the WT construct, as the band was also present in the non-treated and the SON-treated samples. Since the PE is also inserted when transfecting the WT construct, the variant apparently enhances PE recognition. Finally, for the two neighboring variants c.4539+1100A>G and c.4539+1106C>T, AON1 and

AON2 correctly restored splicing, while AON3 did not show any rescue (**Figure 2B**).

AON-based PE exclusion using STGD1-derived fibroblast cells

To investigate whether the AON-based splice corrections observed in the midigene assays would also occur in patient-derived cells, we generated fibroblast cell lines from a proband carrying c.[769-784C>T; 5603A>T] and c.1822T>A (p.Phe608Ile) (Patient A-I:2), and from a STGD1 proband carrying c.859-506G>C and a putative deletion on the other allele encompassing intron 7 (case DNA14-33085; not listed in **Table 1**). For both cases, the same PEs were detected as in the midigene assays (**Figure 3**), yet a larger amount of PE-containing mRNA was detected in fibroblasts from A-I:2 grown under NMD-suppressing conditions (18.2%) compared to HEK293T-transfected cells (8.6%). Considering that this STGD1 individual carries a missense variant (p.Phe608Ile), the amount of PE-containing mRNA from the c.769-784C>T allele likely is much higher than 18%. The same AONs that restored splicing in the midigenes also showed rescue in fibroblasts (**Figure 3**). For c.769-784C>T, AON2 was the most effective, correcting 100% of the transcripts, whereas AON3 also showed a strong rescue (**Figure 3** and **Figure S6**; **Table S10**). AON1 also resulted in PE skipping but induced an alternative splicing event. Interestingly, the insertion of this partial PE was also detected in control fibroblasts. Sequence analysis revealed that the upper band of the artefact corresponded to the partial PE, while the lower was a heteroduplex consisting of the WT and partial PE-containing transcripts. *ABCA4* mRNA was also barely detected in the fibroblasts carrying c.859-506G>C when they were not subjected to CHX treatment (**Figure 3**). After CHX incubation, a 56-nt PE was detected that, upon delivery of AON1 and AON3, was almost completely absent from the transcripts.

AON-based exon inclusion

To correct the splicing defect of c.4253+43G>A, AONs were designed to promote exon 28 inclusion. Two AONs were delivered to the midigene-transfected HEK293T cells. To assess whether the AON was inducing the inclusion of exon 28, the ratio between correct and Δexon 28 transcript was calculated with Fiji, or using the corrected peak area obtained in the capillary analysis (**Figure S5**). Using both methods, AON2 was shown to induce ~10% extra retention of exon 28 (**Figure 2**), whereas for AON1, splice redirection was less evident (**Table S7**).

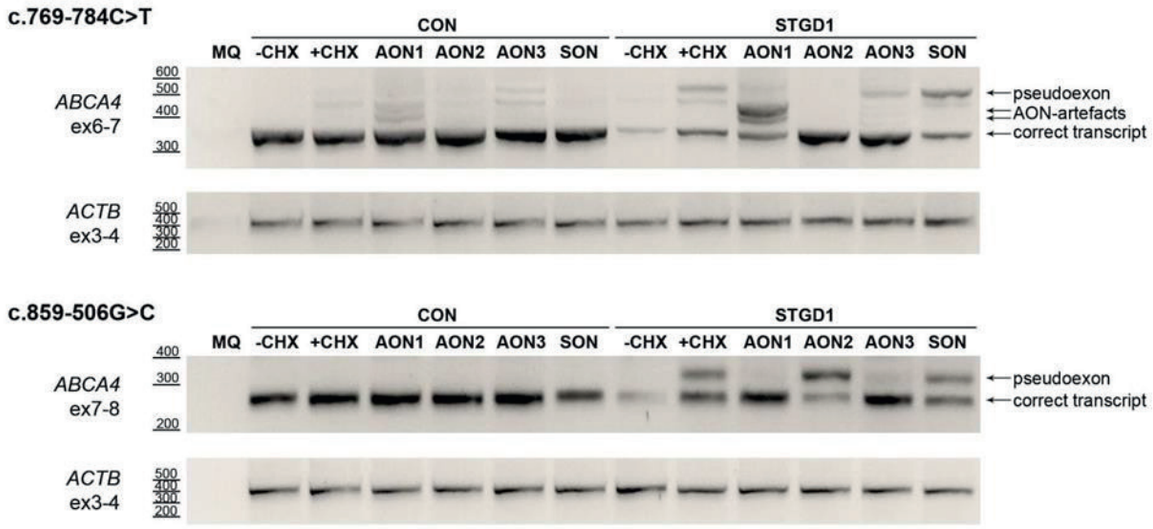


Figure 3. Antisense oligonucleotide rescue in fibroblasts derived from STGD1-affected individuals carrying c.769-784C>T or c.859-506G>C. Fibroblast cells from a control (CON) and a STGD1 individual were transfected with three AONs and a SON. In order to detect the pseudoexon cells were subjected to cycloheximide (+CHX) treatment. Only the non-transfected cells were not subjected to CHX treatment (-CHX). The MQ lane contains the negative control of the PCR reaction. Actin (*ACTB)* was used as a loading control.

Discussion

To identify missing non-coding variants in STGD1 cases, we sequenced the ABCA4 locus in 36 probands and identified one known and five novel intronic variants in 24 (67%) probands. Four novel DI variants (c.769-784C>T, c.859-506G>C, c.4539+1100A>G, c.4539+1106C>T) strengthen cryptic splice sites at non-canonical nucleotide positions and thereby result in PE insertions, contrary to most of the published DI splice site variants that create canonical splice sequences (Liquori et al. 2016; Vaz-Drago et al. 2017). Variant c.1937+435C>G disrupted a splicing silencer motif that is located within a low-expressed alternate exon, and resulted in a 134-nt PE formation (Braun et al. 2013). This PE contains a proper SAS but, similar to the alternate exon, no consensus SDS sequence. The c.4253+43G>A variant disrupted predicted splice silencers and created an ESE, and led to partial skipping of the upstream exon 28. Interestingly, we identified two probands (M-II:1 and P-II:3) that carried DI variants on both alleles.

All novel variants were predicted to result in protein truncation, but two variants, c.769-784C>T and c.4253+43G>A, found in seven and nine probands, respectively, only affect a small proportion of the mRNA. Testing of 412 other genetically unexplained STGD1 cases revealed 29 persons to carry c.4253+43G>A and

four cases to carry c.769-784C>T. AONs were able to (partially) correct the observed splicing defects for all variants in HEK293T cells and, for two variants, in STGD1-derived fibroblasts.

Fourteen first alleles in the 24 probands carry protein-truncating variants and are thus assumed to have a severe effect (**Table 1**). For other missense or complex alleles, the severity is less clear. Based on previous studies (Braun et al. 2013; Zhang et al. 2015; Cornelis et al. 2017; Zernant et al. 2017), our mRNA analysis of the DI variants, and the ages at onset of the STGD1 probands described here, we can hypothesize on the severity of the intronic variants. The c.859-506G>C, c.4539+1100A>G, and c.4539+1106C>T variants all show PE insertion in the majority of ABCA4 transcripts and can thus be considered severe variants. This correlates very well with an age at onset ≤10 years when found in trans with a severe variant and between 15-49 years when being in trans with a mild variant (**Table 1**).

For c.1937+435C>G, the picture is more complicated. The splicing defect was mild, and the two probands carrying this variant carry either a severe variant p.(Leu257Valfs*17) (U-II:1) or a mild allele p.[Gly863Ala, Gly863del; Asn1868Ile] (T-II:1) in trans. The ages of onset were 9 and 35 yrs, respectively, which argues for a severe nature of c.1937+435C>G in U-II:1 and a moderate-to-severe nature in T-II:1. This is discordant with the observed partial PE insertion in HEK293T cells. A similar observation was recently made while studying the effect of c.4539+2001G>A and c.4539+2028C>T.21 In STGD1-derived fibroblasts, mRNA analysis did not show any PE insertion. In STGD1-derived photoreceptor precursor cells however, we observed up to 15% (c.4539+2028C>T) and 25% (c.4539+2001G>A) of PE-containing mRNA (Albert et al. 2018). Based on their heterozygous presence in the corresponding STGD1 cases and their moderate-to-severe (c.4539+2028C>T) or severe (c.4539+2001G>A) nature, about two times these percentages were expected. We hypothesize that the photoreceptor precursor cells had not yet reached a differentiation stage that closely mimics the human retina situation. It is also of interest that, while c.1937+435C>G is predicted to disrupt ESEs, c.4539+2001G>A and c.4539+2028C>T are predicted to activate ESEs. Possibly, the action of these splicing motifs requires specific factors that are exclusively present in mature photoreceptors, which are absent in HEK293T cell and explain the small effect for c.1937+435C>G. Patient-derived photoreceptor-like cells may reveal a more complete splicing defect.

Recently, it was determined that c.2588G>C (p.[Gly863Ala,Gly863del]) only acts as a penetrant pathogenic variant when present in cis with c.5603A>T (p.Asn1868Ile) (Zernant et al. 2017). The common p.(Asn1868Ile) variant was found as a second *ABCA4* allele in up to 50% of mono-allelic STGD1 cases and is strongly associated with late-onset STGD1 (Zernant et al. 2017).Upon Haloplex-based *ABCA4* gene sequencing, we

found p.Asn1868Ile as a single variant in trans with other (potentially) causal variants in 25/65 (38%) of probands, who had an average age at onset of 42 years. In addition, we identified four biallelic but unaffected persons and calculated that this variant, when present in trans with a loss-of-function *ABCA4* variant, shows a penetrance <5% (Runhart et al. 2018). Based on the relatively high frequency of c.769-784C>T in non-Finnish European controls (AF 0.00426 in gnomAD [http://gnomad.broadinstitute.org/]), its partial effect on splicing, and the very late age of onset observed in compound heterozygotes carrying c.[769-784C>T; 5603A>T] (61.0 yrs; range: 38-69 yrs), we consider that c.769-784C>T without c.5603A>T in cis may not be causative. We therefore investigated the MAFs of the c.769-784C>T and c.5603A>T variants (single and in combination) in 250 control Dutch families (GoNL: http://www.nlgenome.nl/) (Genome of the Netherlands 2014). The single c.769-784C>T allele was found more often (6/998 alleles) than the c.[769-784C>T;5603A>T] allele (2/998 alleles). The allele frequency of c.[769-784C>T;5603A>T] in our STGD1 patient cohort (n=250), 0.014, is significantly higher than in our general population (Chi-square test: p=0.023; Bonferroni corrected). In two previous studies, the mean ages at onset for STGD1 patients carrying c.5603A>T as a single second allele in trans with severe *ABCA4* variants were 36 and 42 years (Zernant et al. 2017; Runhart et al. 2018). The mean age of onset for STGD1 patients carrying c.[769-784C>T;5603A>T] is higher than the age of onset of patients carrying the single c.5603A>T variant (Mann-Whithey U; p<0.05). Given its relatively small effect on mRNA splicing (18.2% PE inclusion in mRNA from patient carrying this variant in a heterozygous manner) and its invariable cis-configuration with c.5603A>T, we cannot unequivocally assign pathogenicity to c.769-784C>T. Possibly, patient-derived photoreceptor progenitor cells may shed further light on this matter.

Interestingly, c.4253+43G>A also shows a relatively high non-Finnish European AF (0.00598 in gnomAD). Although c.4253+43G>A, with one exception (M-II:1), was found in cis with c.6006-609T>A, we did not observe a splicing defect for the latter variant in vitro. This could be due to the lack of retina-specific splice factors in the tested cells or due to another missed DI variant, that acts in concert with c.4253+43G>A. In the one proband with c.4253+43G>A who lacked c.6006-609T>A, the c.5603A>T (p.Asn1868Ile) variant was found in cis. The age of onset of the cases carrying c.4253+43G>A ranged from 18 to 61 years (average 41 yrs). Interestingly, the earliest age at onset (18 yrs) was observed in the proband M-II:1 that carried c.[4253+43G>A;5603A>T].

The use of AONs to exclude PEs has been used for several IRD-associated genes (Collin et al. 2012; Gerard et al. 2012; Bax et al. 2015; Garanto et al. 2016; Parfitt et al. 2016; Slijkerman et al. 2016). However, in this study, several technical challenges were

observed. For the c.859-506G>C variant, the design of the AONs was limited to very small parts of the PE due to the repetitive nature of parts of the PE. In case of the c.4539+1100A>G and c.4539+1106C>T variants, we aimed to find effective AONs that would target the PE for both mutations. In addition, using midigenes for c.769-784C>T, we observed variability between replicates, probably due to differences in the transfection efficiencies or passage of the cells. Nevertheless, for each mutation, at least one effective AON was discovered.

Despite the significant yield of novel *ABCA4* mutations in this study, a second *ABCA4* variant was not detected in 11 monoallelic probands and one case with no *ABCA4* variants, which could be due to: i) the presence of heterozygous copy-number variations, as only three were tested using CNV analysis, ii) missed non-coding variants residing in non-covered sequences, iii) too stringent selection criteria when using splicing algorithms, iv) incomplete sensitivity of the midigene in vitro splice assay, and v) genocopies in view of the high AF of *ABCA4* variants (5-10%) in the general population (Maugeri et al. 1999a; Jaakson et al. 2003; Genome of the Netherlands 2014; Cornelis et al. 2017). Additional studies will be needed to further unravel the missing heritability of STGD1.

In conclusion, we identified DI variants in 24/36 (67%) of STGD1 cases with no *ABCA4* variants, one variant, or one causal variant in trans with c.5603A>T. Two alleles (c.[769-784C>T;5603A>T] and c.4253+43G>A) are frequent in our STGD1 cohort and appear to be associated with late-onset STGD1. Due to its small effect on *ABCA4* mRNA in HEK293T cells and patient-derived fibroblasts, and its invariable presence on the same allele as c.5603A>T, we could not establish causality for c.769-784C>T. Interestingly, all the observed splicing defects could be rescued with at least one of the tested AONs, which provides a basis for the development of new therapeutic strategies for individuals with STGD1 carrying these variants.

Materials and Methods

Samples

Twenty-five mono-allelic STGD1 individuals, eight cases carrying one causal *ABCA4* variant and p.Asn1868Ile in trans, and three probands with no ABCA4 variant were genotyped. Thirty-two probands were ascertained from seven different medical centers in the Netherlands and Germany while four were ascertained in the United Kingdom, as part of the UK 10K sequencing project (https://www.uk10k.org/). Patients had been clinically diagnosed with Stargardt disease (STGD1) or cone-rod dystrophy (CRD) by an experienced ophthalmologist in the field of ophthalmogenetics. Each person manifested at least one of the following features of STGD1: yellow-white

pisciform flecks in the retinal pigment epithelium of the posterior pole which were hyperautofluorescent on fundus autofluorescence imaging; and/or progressive atrophy of the macular retinal pigment epithelium. Age at onset for 22 STGD1 individuals and the current age of two STGD1 probands with intronic variants are given in **Table 1**. The age at onset for the remaining 12 STGD1 individuals is given in **Table S6**. Age of onset was defined as the age at which the initial symptoms were noted by the person. For 17 unrelated Dutch probands, genomic DNA of 45 family members was obtained using Oragene DNA saliva kits (DNA Genotek Inc. Ottawa, Canada) as described by the manufacturer, to confirm segregation of non-coding variants.

A cohort of 412 French and German STGD1 individuals with one or no *ABCA4* variant was genotyped by using smMIPs to assess the recurrence of causal non-coding variants described in this study. Written informed consent prior to participation in the study was obtained which adhered to the Declaration of Helsinki. This study was approved by the Medical Ethical Committee 2010-359 (Protocol nr. 2009-32; NL nr. 34152.078.10) and the Commissie Mensgebonden Onderzoek Arnhem-Nijmegen (Dossier no. 2015-1543; dossier code sRP4h).

***ABCA4* sequence analysis and selection of candidate splice variants**

For 22 mono-allelic cases, eight cases with one causal variant and c.5603A>T (p.Asn1868Ile) in trans, and two probands with no *ABCA4* variants, the entire 128-kb *ABCA4* gene (Chr1: 94,458,393 - 94,586,688) and approximately 117 kb proximal and 120 kb distal (total region: Chr1: 94,337,885 – 94,703,604) were enriched by using a custom Haloplex Target Enrichment kit (Agilent, Santa Clara, CA) followed by NGS. Probes to target the 365-kb region were designed with SureDesign (Agilent) and they covered >98% of the targeted regions; gaps in coverage corresponded with repeat-rich regions. Subsequently, samples were sequenced by using Illumina sequencing technology (MiSeq, Illumina, San Diego, CA, USA) to obtain 2x250 bp reads. Data analysis was performed with CLC Bio Software (CLC Bio Genomics Workbench, Qiagen, Hilden, Germany). Reads were aligned against a reference genome (GRCh37/hg 19 assembly) and variants were identified and annotated by a custom-designed workflow. Variants were only considered for further filtering if they were present in at least one forward and one reverse read, in a minimum of 20% of all reads and with a minimum coverage of five reads. The variants were further annotated by Alamut Batch v2.7 (Interactive Biosoftware). Prior to locus sequencing, the 41 STGD1 individuals were sequenced for variants in the coding and flanking splice site regions using Sanger sequencing and 29/41 cases also were analyzed for the presence of heterozygous CNVs using Multiplex Ligation Probe Amplification (MLPA). Probands carrying the p.Asn1868Ile common variant as a non-complex second allele (as based on segregation

analysis) were excluded from this group earlier and will be described elsewhere.

Three mono-allelic cases (GC20009, GC17027, GC21017) and one proband with no *ABCA4* variants (GC20637) underwent genome sequencing as part of the National Institute for Health Research (NIHR) BioResource- Rare Diseases study using the Illumina TruSeq DNA PCR-Free Sample preparation kit (Illumina, Inc.) and sequenced using an Illumina HiSeq 2500, generating minimum coverage of 15× for approximately 95% of the genome (Carss et al. 2017). Reads were aligned to the Genome Reference Consortium human genome build 37 (GRCh37) using Isaac Genome Alignment Software (version 01.14; Illumina, Inc.). Single nucleotide variants and small indels were identified using Isaac Variant Caller (version 2.0.17). Structural variants were identified using two independent algorithms: Isaac Copy Number Variant Caller (Canvas, Illumina), and Isaac Structural Variant Caller (Manta, Illumina) as previously described (Carss et al. 2017). In the first instance, coding variants with an AF <0.01 in control datasets including the NIHR BioResource Rare Diseases cohort and the Exome Aggregation Consortium (ExAC) database and passing standard quality filters were prioritized. A white list of previously reported pathogenic mutations in relevant genes normally excluded by these filters were investigated (e.g., higher AF *ABCA4* variants: c.2588G>C, c.5882G>A, c.5461-10T>C. An in-house curated database of known and candidate IRD-associated genes was used to prioritize candidate biallelic variants in simplex and recessive families.

In all 36 samples, after exclusion of known first alleles, candidate variants were selected according to two inclusion criteria: i) presence in the *ABCA4* gene region (Chr1: 94,458,393 - 94,586,688), and ii) a MAF less than or equal to 0.01 in Genome of the Netherlands (GoNL; http://www.nlgenome.nl) (Genome of the Netherlands 2014), UK 10K project (https://www.uk10k.org/) and non-Finnish European gnomAD (http://gnomad.broadinstitute.org/) population frequency databases.

In silico splice prediction for all variants was performed by using five algorithms (SpliceSiteFinder-like, MaxEntScan, NNSPLICE, GeneSplicer, Human Splicing Finder) (Shapiro and Senapathy 1987; Reese et al. 1997; Pertea et al. 2001; Yeo and Burge 2004; Desmet et al. 2009) via Alamut Visual software version 2.7, by comparing splicing scores for WT and variant nucleotides. Inclusion criteria for splicing assays were: 1) the presence of a variant splice score >75% of the scoring range for at least two prediction programs, 2) a difference between WT and variant splice score >2% of the scoring range in at least two algorithms identified in 1). A few variants did not meet these criteria but were found enriched in our STGD1 cases or were located in alternate exons previously described (Braun et al. 2013) and therefore were included in the splice assays.

Cell lines and culture conditions

Human Embryonic Kidney 293 cells (HEK293T) were purchased from ECACC (Catalogue no. 12022001; Salisbury, UK), which were tested for genomic integrity by STR PCR. HEK293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM), supplemented with 10% fetal calf serum (FCS), 1% penicillin-streptomycin (P/S) and 1% sodium pyruvate (NaPyr) at 37°C and 5% CO_2. Cells were split twice a week in a dilution 1:10. Fibroblast cell lines were generated in house from skin biopsies and cultured in DMEM supplemented with 20% FCS, 1% P/S and 1% NaPyr at 37°C and 5% CO_2. Cells were split twice a week in a dilution 1:3. For all cell lines Mycoplasma contamination tests were negative and were conducted every 2 months.

Midigene-based splice assay

The effect of 11 non-coding was assessed by midigene-based splicing assays employing seven WT BA clones previously described (**Chapter 2**) and the newly designed BA30. All constructs contained *ABCA4* genomic sequences from the bacterial artificial chromosome clone CH17-325O16 (insert g.94,434,639-94,670,492) and they served as templates to generate mutant constructs by site-directed mutagenesis, which subsequently were validated by Sanger sequencing (**Table S11** and **S12**). WT and mutant constructs were transfected in HEK293T cells and the extracted total RNA was subjected to reverse transcription (RT)-PCR (**Table S1**) as described previously (**Chapter 2**).

Single molecule molecular inversion probe-based sequence analysis of intronic *ABCA4* regions

To test the occurrence of five intronic *ABCA4* variants (c.769-784C>T, c.859-506G>C, c.4253+43G>A, c.4539+1100A>G, c.4539+1106C>T) in 412 genetically unsolved STGD1 cases from Lille, France (n = 224) and Regensburg, Germany (n = 188), we designed smMIPs based on a previously developed cost-effective sequencing protocol (**Table S13**) (Hiatt et al. 2013). STGD1 samples from a French cohort previously were prescreened using a mutation scanning technique, dHPLC or NGS (i.e., sequence analysis of inherited retinal disease-associated variants, including *ABCA4*). The STGD1 cases from a German cohort were previously sequenced using a custom-designed GeneChip CustomSeq Resequencing Array (RetChip) (Affymetrix, Santa Clara, CA, USA) or NGS on an ION Torrent semiconductor personal sequencing machine (Life Technologies, Darmstadt, Germany) upon multiplex-PCR amplification of the fragments with the STARGARDT MASTR kit (Multiplicom, Niel, Belgium) (Schulz et al. 2017).

The genomic positions of the intronic variants were obtained from the UCSC

genome browser (hg19; https://genome.ucsc.edu), and were used to design smMIPs using an in-house pipeline. For each intronic *ABCA4* variant, we designed two (overlapping) smMIPs, one on the plus strand and one on the minus strand. Each smMIP contains 78 nt and targets about 110 bp of genomic sequences based on the position of two annealing arms. The two annealing arms (denoted extension and ligation probes, together 40 nt) are connected using a common linker sequence of 38 nt which includes a 8-nt random tag that serves as the 'single molecule' identifier. The 8-nt random tags enable the recognition of individual capture events and thereby point to random sequencing errors. The probes were designed for each DNA strand separately to ensure coverage from the plus and minus strands. The 78-mer smMIPs were synthesized by Integrated DNA Technologies (IDT, Leuven, Belgium). All probes were pooled in equal amount (i.e., 1 µM), and thereafter used to prepare the sequencing libraries according to a protocol described elsewhere (O'Roak et al. 2012).

We sequenced 412 STGD1 cases by using bar-coded PCR primers in two series. Bar-coded libraries were pooled, and after purification were sequenced using Illumina NexSeq500 using two Mid output kits (maximum of 130 million reads). The paired end sequencing reads were combined into 152-nt forward-reverse-reads (fr-reads), and were mapped to the reference genome. Next, the sm tags were exploited to group fr-reads. The resultant fr-read groups were used to make a highly accurate consensus read referred to as single molecule consensus reads (smc-reads), which were used for variant calling. The variant calling and annotation was performed using the in-house pipeline. The smc-read average ranged from 14 to 1587, with an average coverage of 489x.

Antisense Oligonucleotides

For each DI variant causing a PE inclusion three AONs were designed (**Table S2**). The design of the therapeutic molecules was performed as described elsewhere (Aartsma-Rus 2012; Garanto and Collin 2018). Briefly, for each variant, the sequence of each PE plus the flanking regions (50 bp) were selected for further analyses. First the RNA structure and the exon splicing enhancers (ESEs) were predicted. Subsequently, AONs were designed taking several parameters into account (free energy, Tm, %GC, binding region, etc). For the c.4253+43G>A variant, two AONs were designed to bind directly on the mutant sequence. In addition, a sense oligonucleotide (SON) was synthesized. All oligonucleotides had a phosphorothioate backbone with a 2-*O*-methyl sugar modification (2OME/PS) and were synthesized by Eurogentec. All oligonucleotides (AONs and SON) were resuspended in PBS at a concentration of 100 µM (stock solution). The final concentration used in all the experiments was 0.5 µM.

In vitro rescue studies in HEK293T cells using midigenes and AONs

Around 400,000 HEK293T cells were seeded in each well of a 6-well plate. The next day, each well was transfected with 1.5 µg of either the WT or the mutant construct. In total, eleven wells were transfected with one of the midigene constructs, while one well was left untransfected (HEK293T control). Of note, variants c.4539+1100A>G and c.4539+1106C>T result in the insertion of the same PEs and therefore were inserted in the same WT construct and were treated with the same AONs. Twenty-four hours post-transfection, each well was trypsinised and subdivided in 5 wells of a 24-well plate. Once cells were attached, the cells of each well were transfected with the respective AON, the SON, or was left untransfected (NT). Transfections were performed using FuGENE® HD Transfection Reagent (Promega) using a ratio 3:1 (3 µl of reagent for each µg of DNA) for midigenes and 1:10 for AONs (1 µl of reagent for each 10 µl of AON). Forty-eight hours post-AON delivery, cells were harvested and RNA isolation was performed using NucleoSpin® RNA kit (Macherey-Nagel). Subsequently, 1 µg of RNA was used for cDNA synthesis using iSCRIPT cDNA synthesis kit (Bio-Rad). cDNA was diluted to 20 ng/µl by adding 30 µl of MQ and 50 ng (2.5 µl) was used for the PCR analysis. All PCR reactions were prepared at a final volume of 25 µl. The conditions for all PCR reactions were as follows: 2 min at 94°C, followed by 35 cycles of 30 s at 94°C, 30 s at 58°C and 2 min at 72°C, a final elongation step of 5 min at 72°C was performed. Subsequently, 10 µl of each reaction were loaded onto a 2%-agarose gel. Rhodopsin amplification was used as transfection and loading control. All experiments were performed in two independent replicates. Primer sequences can be found in the **Table S1**.

Rescue studies using antisense oligonucleotides in fibroblasts

Approximately 500,000 fibroblast cells were seeded in each well of a 6-well plate (6 wells per cell line). The next day, cells were transfected with either 0.5 µM of one of the AONs or SON, or transfected with empty liposomes (NT). Forty-four hours post-transfection medium was removed and medium containing cycloheximide (CHX) was added in order to block non-sense mediated decay (NMD). Four hours later cells were harvested and subjected to RNA isolation as previously described. For cDNA synthesis, 1 µg of RNA was used using Invitrogen™ SuperScript™ IV VILO™ Master Mix (Thermo Fisher Scientific). cDNA was diluted to 20 ng/µl by adding 30 µl of MQ. For each PCR reaction, 80 ng (4 µl) of cDNA were used, except for actin (50 ng). All PCR reactions were prepared at a final volume of 25 µl. The conditions for all *ABCA4* PCR reactions were as follows: 2 min at 94°C, followed by 40 cycles of 30 s at 94°C, 30 s at 58°C and 2 min at 72°C, a final elongation step of 5 min at 72°C was performed. For actin, the conditions were as follows: 2 min at 94°C, followed by 35 cycles of 30 s at 94°C, 30 s at

58°C and 30s at 72°C, a final elongation step of 2 min at 72°C. Subsequently, 20 µl of each *ABCA4* reaction were loaded onto a 2%-agarose gel. Ten microliter of the amplified actin product was used as loading control. All experiments were performed in two independent replicates. Primer sequences are indicated in **Table S1**.

Capillary analysis

In order to quantify the ratios between correct and aberrant transcripts we loaded the RT-PCR samples onto a Fragment Analyzer Auto Capillary Electrophoresis System (Advanced Analytical Technologies, Inc.). Midigene-derived RT-PCR products were diluted 1:10 whereas fibroblast-derived RT-PCRs were diluted 1:4. In both cases 2 µl was used for capillary analysis. Analysis of the peaks was performed using the corrected peak area parameter and only the peaks corresponding to the expected bands were taken into account.

Acknowledgments

We thank Ellen Blokland, Lonneke Duijkers, Duaa Elmelik, Anita Hoogendoorn, Marlie Jacobs-Camps, Saskia van der Velde-Visser and Marijke Zonneveld-Vrieling for technical assistance. We thank Sabine Defoort, Hélène Dollfus, Isabelle Drumare, Christian P. Hamel, Karsten Hufendiek, Cord Huchzermeyer, Herbert Jägle, Ulrich Kellner, Philipp Rating, Klaus Rüther, Eric Souied, Georg Spital and Xavier Zanlonghi for their cooperation and ascertaining STGD1 cases.

Supplemental data

Supplemental Figures S1, S2 and Supplemental Tables S1 till S12 are not included in this thesis and they can be found online at the following link.
https://static-content.springer.com/esm/art%3A10.1038%2Fs41436-018-0414
9/MediaObjects/41436_2018_414_MOESM1_ESM.pd

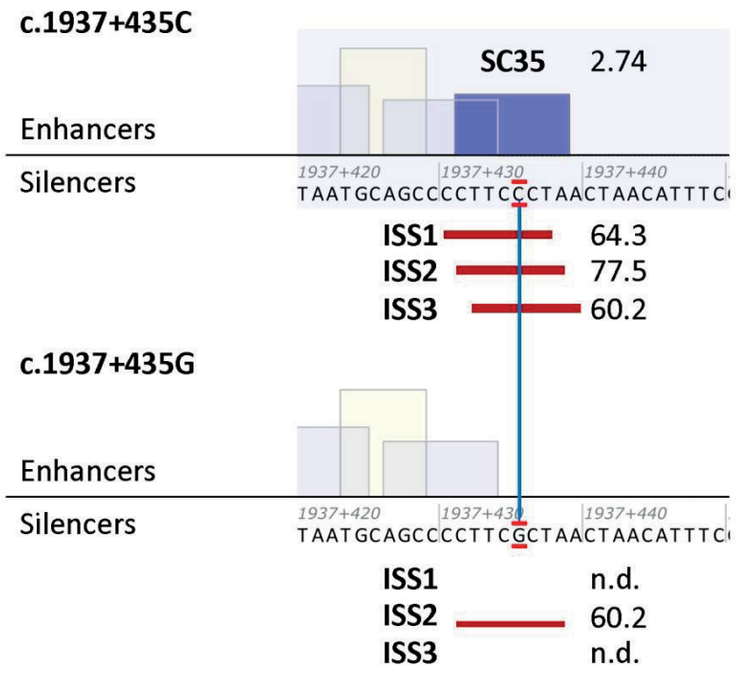


Figure S3. Enhancer and silencer motif variation caused by c.1937+435C>G. The c.1937+435C>G variant disrupts one SC35 enhancer and two intronic splice silencers (ISS1 and ISS3), and reduces the strength of ISS2. n.d., not detected.

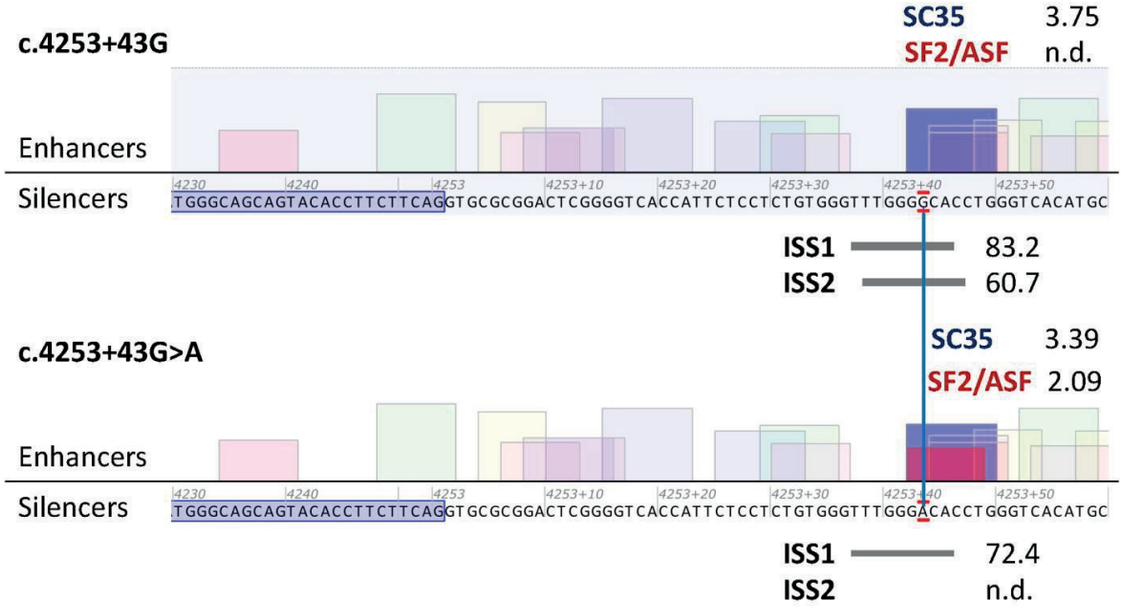


Figure S4. Enhancer and silencer motif variation caused by c.4253+43G>A. The c.4253+435G>C variant weakens one SC35 and creates one SF2/ASF enhancer. It also weakens one intronic splice silencers (ISS1) and disrupts ISS2. n.d., not

References

Aartsma-Rus A. 2012. Overview on AON design. *Methods Mol Biol* **867**: 117-129.

Abu-Safieh L, Alrashed M et al. 2013. Autozygome-guided exome sequencing in retinal dystrophy patients reveals pathogenetic mutations and novel candidate disease genes. *Genome Res* **23**: 236-247.

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

Allikmets R, Singh N et al. 1997. A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. *Nat Genet* **15**: 236-246.

Bauwens M, De Zaeytijd J et al. 2015. An augmented *ABCA4* screen targeting noncoding regions reveals a deep intronic founder variant in Belgian Stargardt patients. *Hum Mutat* **36**: 39-42.

Bax NM, Sangermano R et al. 2015. Heterozygous deep-intronic variants and deletions in *ABCA4* in persons with retinal dystrophies and one exonic *ABCA4* variant. *Hum Mutat* **36**: 43-47.

Beryozkin A, Shevah E et al. 2015. Whole Exome Sequencing Reveals Mutations in Known Retinal Disease Genes in 33 out of 68 Israeli Families with Inherited Retinopathies. *Sci Rep* **5**: 13187.

Bonifert T, Gonzalez Menendez I et al. 2016. Antisense Oligonucleotide Mediated Splice Correction of a Deep Intronic Mutation in OPA1. *Mol Ther Nucleic Acids* **5**: e390.

Braun TA, Mullins RF et al. 2013. Non-exomic and synonymous variants in *ABCA4* are an important cause of Stargardt disease. *Hum Mol Genet* **22**: 5136-5145.

Carss KJ, Arno G et al. 2017. Comprehensive Rare Variant Analysis via Whole-Genome Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease. *Am J Hum Genet* **100**: 75-90.

Collin RW, den Hollander AI et al. 2012. Antisense Oligonucleotide (AON)-based Therapy for Leber Congenital Amaurosis Caused by a Frequent Mutation in CEP290. *Mol Ther Nucleic Acids* **1**: e14.

Cornelis SS, Bax NM et al. 2017. In Silico Functional Meta-Analysis of 5,962 *ABCA4* Variants in 3,928 Retinal Dystrophy Cases. *Hum Mutat* **38**: 400-408.

Cremers FP, van de Pol DJ 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.

Desmet FO, Hamroun D et al. 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* **37**: e67.

Fakin A, Robson AG et al. 2016. Phenotype and Progression of Retinal Degeneration Associated With Nullizigosity of *ABCA4*. *Invest Ophthalmol Vis Sci* **57**: 4668-4678.

Garanto A, Chung DC et al. 2016. In vitro and in vivo rescue of aberrant splicing in CEP290-associated LCA by antisense oligonucleotide delivery. *Hum Mol Genet* **25**: 2552-2563.

Garanto A, Collin RWJ. 2018. Design and In Vitro Use of Antisense Oligonucleotides to Correct Pre-mRNA Splicing Defects in Inherited Retinal Dystrophies. *Methods Mol Biol* **1715**: 61-78.

Genome of the Netherlands C. 2014. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet* **46**: 818-825.

Gerard X, Perrault I et al. 2012. AON-mediated Exon Skipping Restores Ciliation in Fibroblasts

Harboring the Common Leber Congenital Amaurosis CEP290 Mutation. *Mol Ther Nucleic Acids* **1**: e29.

Gilissen C, Hehir-Kwa JY et al. 2014. Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511**: 344-347.

Haer-Wigman L, van Zelst-Stams WA et al. 2017. Diagnostic exome sequencing in 266 Dutch patients with visual impairment. *Eur J Hum Genet* **25**: 591-599.

Hammond SM, Wood MJ. 2011. Genetic therapies for RNA mis-splicing diseases. *Trends Genet* **27**: 196-205.

Han Z, Conley SM et al. 2012. DNA nanoparticle-mediated *ABCA4* delivery rescues Stargardt dystrophy in mice. *J Clin Invest* **122**: 3221-3226.

Hiatt JB, Pritchard CC et al. 2013. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. *Genome Res* **23**: 843-854.

Jaakson K, Zernant J et al. 2003. Genotyping microarray (gene chip) for the ABCR (*ABCA4*) gene. *Hum Mutat* **22**: 395-403.

Liquori A, Vache C et al. 2016. Whole USH2A Gene Sequencing Identifies Several New Deep Intronic Mutations. *Hum Mutat* **37**: 184-193.

Maugeri A, Klevering BJ et al. 2000. Mutations in the *ABCA4* (ABCR) gene are the major cause of autosomal recessive cone-rod dystrophy. *Am J Hum Genet* **67**: 960-966.

Maugeri A, van Driel MA et al. 1999a. 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.

Maugeri A, van Driel MA et al. 1999b. 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. *American Journal of Human Genetics* **64**: 1024-1035.

Neveling K, Collin RW et al. 2012. Next-generation genetic testing for retinitis pigmentosa. *Hum Mutat* **33**: 963-972.

Neveling K, Feenstra I et al. 2013. A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. *Hum Mutat* **34**: 1721-1726.

Ng SB, Turner EH et al. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**: 272-276.

Nikopoulos K, Gilissen C et al. 2010. Next-generation sequencing of a 40 Mb linkage interval reveals TSPAN12 mutations in patients with familial exudative vitreoretinopathy. *Am J Hum Genet* **86**: 240-247.

O'Roak BJ, Vives L et al. 2012. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338**: 1619-1622.

Parfitt DA, Lane A et al. 2016. Identification and Correction of Mechanisms Underlying Inherited Blindness in Human iPSC-Derived Optic Cups. *Cell Stem Cell* **18**: 769-781.

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

Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. *J Comput Biol* **4**: 311-323.

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

Sangermano R, Khan M et al. 2018. *ABCA4* midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Res* **28**: 100-110.

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

Shapiro MB, Senapathy P. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* **15**: 7155-7174.

Slijkerman RW, Vache C et al. 2016. Antisense Oligonucleotide-based Splice Correction for USH2A-associated Retinal Degeneration Caused by a Frequent Deep-intronic Mutation. *Mol Ther Nucleic Acids* **5**: e381.

Vaz-Drago R, Custodio N, Carmo-Fonseca M. 2017. Deep intronic mutations and human disease. *Hum Genet* **136**: 1093-1111.

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

Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* **11**: 377-394.

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

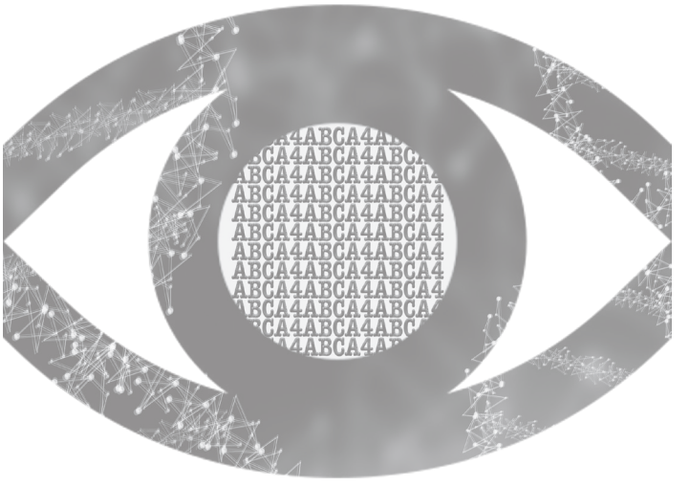
Zernant J, Schubert C et al. 2011. Analysis of the *ABCA4* gene by next-generation sequencing. *Invest Ophthalmol Vis Sci* **52**: 8479-8487.

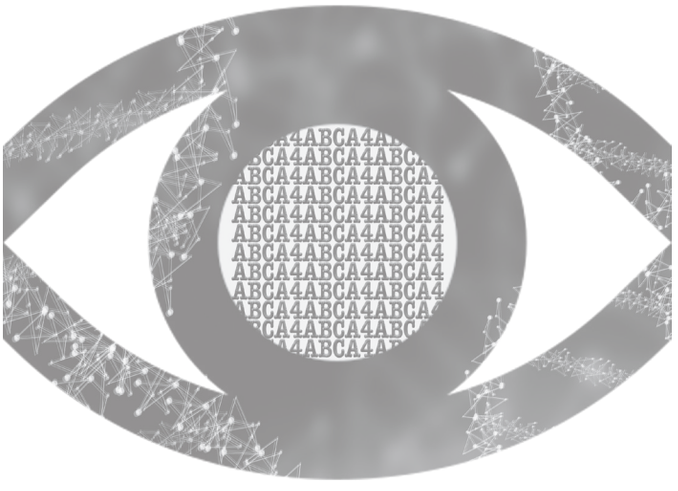
Zernant J, Xie YA et al. 2014. Analysis of the *ABCA4* genomic locus in Stargardt disease. *Hum Mol Genet* **23**: 6797-6806.

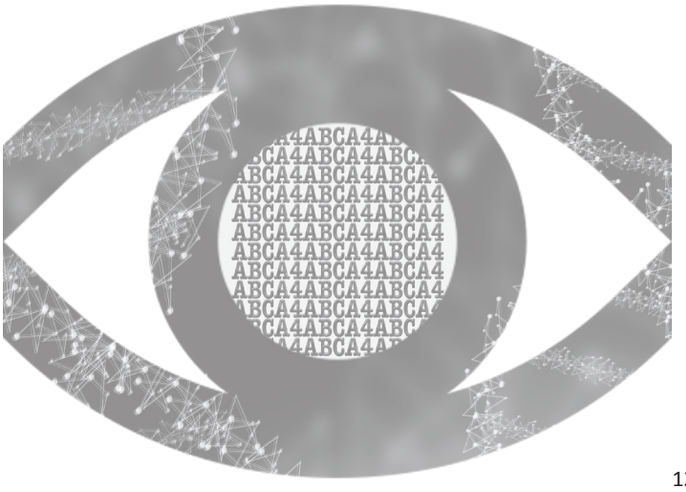
Zhang N, Tsybovsky Y et al. 2015. Protein misfolding and the pathogenesis of *ABCA4*-associated retinal degenerations. *Hum Mol Genet* **24**: 3220-3237.

Chapter 4

Cost-effective molecular inversion probe-based *ABCA4* sequencing reveals deep-intronic variants in Stargardt disease







Cost-effective molecular inversion probe-based *ABCA4* sequencing reveals deep-intronic variants in Stargardt disease

Mubeen Khan, Stéphanie S. Cornelis, Muhammad Imran Khan, Duaa Elmelik, Eline Manders, Sem Bakker, Ronny Derks, Kornelia Neveling, Maartje van de Vorst, Christian Gilissen, Isabelle Meunier, Sabine Defoort, Bernard Puech, Aurore Devos, Heidi L. Schulz, Heidi Stöhr, Felix Grassmann, Bernhard H.F. Weber, Claire-Marie Dhaenens and Frans P.M. Cremers

Hum Mutat. 2019; **40**: 1749-1759.

Abstract

Purpose: Stargardt disease (STGD1) is caused by biallelic mutations in *ABCA4*, but many patients are genetically unsolved due to insensitive mutation-scanning methods. We aimed to develop a cost-effective sequencing method for *ABCA4* exons and regions carrying known causal deep-intronic (DI) variants.

Methods: Fifty exons and 12 regions containing 14 DI variants of *ABCA4* were sequenced using double-tiled single molecule Molecular Inversion Probe (smMIP)-based next generation sequencing. DNAs of 16 STGD1 cases carrying 29 *ABCA4* alleles and of four healthy persons were sequenced using 483 smMIPs. Thereafter, DNAs of 411 STGD1 cases with one or no *ABCA4* variant were sequenced. The effect of novel non-coding variants on splicing was analyzed using *in vitro* splice assays.

Results: Thirty-four *ABCA4* variants previously identified in 16 STGD1 cases were reliably identified. In 155/411 probands (38%), two causal variants were identified. We identified 11 DI variants present in 62 alleles. Two known and two new noncanonical splice site (NCSS) variants showed splicing defects, and one novel DI variant (c.4539+2065C>G) resulted in a 170-nt mRNA pseudoexon insertion (p.[Arg1514Lysfs*35,=]).

Conclusions: smMIPs-based sequence analysis of coding and selected non-coding regions of *ABCA4* enabled cost-effective mutation detection in STGD1 cases in previously unsolved cases.

Introduction

Although Sanger sequencing is considered the gold standard for the identification of disease-associated variants, it is less suitable for the identification of variants in one or more genes consisting of many exons due to high costs and low throughput (Neveling et al. 2013). The emergence of next generation sequencing (NGS) has revolutionized the genotyping of patients with inherited diseases, including inherited retinal diseases (IRDs) (Levy and Myers 2016; Carss et al. 2017). Whole exome sequencing (WES) is a cost-effective method that reveals the large majority of coding and splice site variants but cannot detect deep-intronic (DI) and regulatory variants. Whole genome sequencing (WGS) in principle detects all variants, including structural variations, but disadvantages are extensive data processing time and storage of a huge amount of data. As an alternative, NGS-based targeted sequencing approaches have been introduced to selectively enrich and sequence the genomic regions of interest. Advantages include a better coverage and faster generation of data (Lin et al. 2012).

One of the target enrichment strategies is based on molecular inversion probes (MIPs). Single-molecule (sm)MIPs contain a unique tag. Due to their dual annealing properties, smMIPs and MIPs are very specific and have been used in multiplex PCR enrichment consisting of 1,312 or 6,200 probes to analyze 33 genes involved in cancer (Hiatt et al. 2013) or 108 genes implicated in IRDs (Weisschuh et al. 2018), respectively. This technology requires an initial investment in synthesizing smMIP oligonucleotides and balancing of their targeting properties, but thereafter it is superior in terms of cost, throughput, scalability, sensitivity and specificity. smMIPs can be employed for simultaneous sequencing of hundreds of patients (Neveling et al. 2017; Weisschuh et al. 2018).

In this study, we aimed to identify mutations in patients with Stargardt disease (STGD1). STGD1 (MIM# 248200) is an autosomal recessive disease caused by mutations in the gene encoding the ATP binding cassette type A4 (*ABCA4*) (MIM# 601691) (Allikmets et al. 1997). With an estimated prevalence of 1 in 10,000 individuals (Blacharski et al. 1988), STGD1 is the most frequent juvenile inherited macular dystrophy and *ABCA4* is the most frequently mutated IRD-associated gene (Tanna et al. 2017; Hussain et al. 2018). In an inventory of all reported variants and cases, 5,962 (allelic) variants identified in 3,928 cases were listed (Cornelis et al. 2017). Until recently, 35% of STGD1 cases carried one or no coding or splice site mutation (Zernant et al. 2014). Copy-number variations, DI variants (Braun et al. 2013; Zernant et al. 2014; Bauwens et al. 2015; Bax et al. 2015) and a low-penetrant frequent coding variant (p.Asn1868Ile) (Zernant et al. 2017; Runhart et al. 2018) explained about 10% of this missing heritability. Recently, upon sequence analysis of the entire *ABCA4* gene, eight

novel and one known DI variant explained approximately 65% of the remaining unsolved cases (Zernant et al. 2018; Bauwens et al. 2019; **Chapter 3**). As the phenotype is specific and no other gene has been described to be mutated in typical STGD cases since *ABCA4* was identified 22 years ago (Allikmets et al. 1997), hundreds of cases have remained genetically unsolved due to the low-sensitive mutation scanning techniques that were used such as single strand conformation polymorphism (Maugeri et al. 1999; Klevering et al. 2004), denaturing high performance liquid chromatography (Maia-Lopes et al. 2009) , and arrayed primer extension (Jaakson et al. 2003).

In this study, we aimed to develop a cost-effective sequencing method for the complete *ABCA4* gene based on smMIPs. As a first step, we developed smMIPs to sequence all 50 exons and flanking splice sites, as well as selected regions carrying 14 known causal DI variants. Using this smMIPs platform, we sequenced 411 genetically unsolved STGD1 cases. Selected novel noncanonical splice site (NCSS) and DI variants were tested by using midigene splice assays.

Materials and Methods

Study cohort

We analyzed the *ABCA4* gene for sequence variants employing STGD1 cases from Lille, France (n=223 cases) and Regensburg, Germany (n=188 cases). Samples were collected according to the tenets of the Declaration of Helsinki and written informed consent was obtained from all participants. STGD1 samples from the French cohort (originating from a cohort of 1,133 probands) were prescreened using the following techniques: high resolution melting (HRM) mutation detection analysis, denaturing high-performance liquid chromatography (dHPLC), or next-generation sequencing (NGS) of four IRD-associated genes, including *ABCA4*. The STGD1 cases from Germany (originating from a cohort of 335 probands) were previously sequenced using a custom-designed GeneChip CustomSeq Resequencing Array (RetChip) (Affymetrix, Santa Clara, CA, USA) (Schulz et al. 2017) or NGS on an ION Torrent semiconductor personal sequencing machine (Thermo Fisher Scientific, Darmstadt, Germany) upon multiplex-PCR amplification of the fragments with the STARGARDT MASTR kit (Multiplicom, Niel, Belgium).

Design of smMIPs

The design of smMIPs was performed in two series. First, smMIPs were designed to sequence the 50 protein-coding exons of *ABCA4* (NM_000350.2), including at least 20 bp upstream and 20 bp downstream sequences encompassing the splice site consensus sequences. Also six previously reported DI variants including

(c.4253+43G>A, c.5196+1137G>A, c.5196+1216C>A, c.5196+1056A>G, c.4539+2001G>A, c.4539+2028C>T) (Braun et al. 2013; Zernant et al. 2018), were captured covering small parts of introns 30 (182 bp) and 36 (226 bp). In the first sequence analysis of 22 test DNA samples, smMIPs targeting exons 7, 10, 13, 31, 33, 36, 37, 38 and 46 resulted in less than 10 reads whereas the average coverage of all smMIPs was >40 reads. To cover these gaps, 23 additional smMIPs were designed and added in the second sequence analysis to improve the coverage of these regions. All regions were covered well except for the last three nucleotides of exon 10 and the first 20 nucleotides of intron 10. Thereafter, the final rebalanced smMIP pool consisted of 309 smMIPs including 299 exonic and 10 intronic smMIPs.

In addition, eight recently discovered causal DI variants (c.769-784C>T, c.859-506G>C, c.859-540C>G, c.1937+435C>G, c.4539+1100A>G, c.4539+1106C>T, c.4539+2064C>T and c.5197-557G>T (Bauwens et al. 2019; **Chapter 3**) as well as 300 bp upstream and 300 bp downstream sequences were captured using 174 smMIPs (**Table S1** and **Figure S1**). Larger segments were sequenced as they may carry novel causal DI variants that may affect the recognition of the same pseudoexons (PEs) by the splicing machinery. Thereafter both pools were combined and consisted of 483 smMIPs.

To design smMIPs, the genomic positions were obtained from the UCSC genome browser (hg19; https://genome.ucsc.edu). smMIPs were designed using MIPgen pipeline (Boyle et al. 2014). For each target region overlapping smMIPs were designed in a double-tiling fashion, i.e., one on the plus strand and one on the minus strand. Details of the smMIPs and their distribution across *ABCA4* are provided in **Table S1** and **Figure S1**.

Each smMIP is 78 nucleotides long and targets 110 nt. The two annealing arms (denoted extension and ligation arm) together are 40 nt and are connected using a common linker sequence of 30 nt. In addition, an 8-nt random tag serves as 'single molecule' identifier (Eijkelenboom et al. 2016). The smMIPs were synthesized by Integrated DNA Technologies (IDT, Leuven, Belgium). All smMIPs were phosphorylated after pooling, as described previously (Neveling et al. 2017).

Automated smMIPs library preparation and sequencing

ABCA4 sequencing libraries were prepared using a fully automated sequencing workflow based on smMIPs enrichment in combination with multiplex-PCR using barcoded PCR primers in two series, followed by sequencing with NextSeq 500 (by using Mid output kits (Illumina) (130 million reads) as described previously (Neveling et al. 2017).

Variant calling and annotation

Data were analyzed using an in-house bioinformatics pipeline starting from the raw sequencing reads (FASTQ). In the first step unique molecular identifiers (UMIs) were trimmed from the sequencing reads and stored within the read identifier for later use. The paired-end sequencing reads were then directly mapped to the reference genome (human genome build GRCh37/hg19) using BWA mem (v0.7.12). The extension and ligation arms as well as overlap between the read-pairs were trimmed after the alignment of each read based on the smMIPs design file, in order to improve the initial alignment. Duplicated reads, located on the same position and having the same UMI, were removed while the remaining unique reads were written to a single bam file per patient based on the barcode. Variants were called using both the UnifiedGenotyper as well as the HaplotypeCaller from GATK (v3.4-46) for each individual sample followed by merging of the two variant call sets using the GATK Combine Variants and joint genotyping functions. The two resulting VCF files were used as input for an in-house annotation pipeline where variants were annotated with effect predictions, gene information as well as frequency information from various population databases (Lelieveld et al. 2016).

Variant prioritization

First, variants were prioritized by using the criteria 'Quality by Depth' >500, which represents the overall coverage of the region, an Allele Frequency (AF) <0.005 in the dbSNP database and in an in-house whole exome data set of 21,559 persons, and an AF <0.01 in control population datasets, such as the Genome Aggregation Database (gnomAD, http://gnomad.broadinstitute.org/), and passing standard quality filters including gene components such as exons and canonical splice sites were prioritized. Previously reported pathogenic variants with high AFs such as c.2588G>C and c.5603A>T (AFs in nFE in gnomAD 0.00784 and 0.06647 respectively) (Cremers et al. 1998; Zernant et al. 2014; Cornelis et al. 2017; Schulz et al. 2017; Zernant et al. 2017), were selected separately. Known DI variants were selected based on prior knowledge from literature (Braun et al. 2013; Zernant et al. 2014; Bauwens et al. 2015; Schulz et al. 2017; Albert et al. 2018; Bauwens et al. 2019; **Chapter 3**).

Coding variants classification

For novel coding variants, along with the AF filters, *in silico* pathogenicity assessment was performed including Sorting Intolerant From Tolerant (SIFT, http://sift.bii.a-star.edu.sg/) (Kumar et al. 2009), Polymorphism Phenotyping v2 (PolyPhen-2) (Adzhubei et al. 2010) and MutationTaster (Schwarz et al. 2014) by using Alamut Visual software version 2.7 (Interactive Biosoftware, Rouen, France; www.interactive-

biosoftware.com) (**Table 1**). Variants were classified according to the guidelines of the American College of Medical Genetics (ACMG) (Richards et al. 2015).

Non-coding variants classification

For the selection of noncanonical splice site (NCSS) and DI variants, *in silico* predictions were performed by using five algorithms (SpliceSiteFinder-like, MaxEntScan, NNSPLICE, GeneSplicer, Human Splicing Finder) via Alamut Visual software version 2.7 (Reese et al. 1997; Pertea et al. 2001; Cartegni et al. 2003; Yeo and Burge 2004; Desmet et al. 2009); Biosoftware, 2014) (Interactive Biosoftware, Rouen, France; www.interactive-biosoftware.com), by comparing splicing scores for wild-type (WT) and variant nucleotides (**Table S2**).

Segregation analysis

Segregation analysis was performed for 45 cases from Lille. PCR products were sequenced in sense and antisense directions using the BigDye® Terminator v3.1 Cycle Sequencing Kit on a 3730 DNA analyzer (Applied Biosystems, Carlsbad, CA).

Midigene-based splice assay and RT-PCR assessment

The effect of seven NCSS and three DI variants was assessed by midigene-based splicing assays employing WT BA constructs described previously (**Chapter 2**) and a newly designed BA31 construct (**Table S3**). Details of mutagenesis primers and sequencing primers are given in **Table S4**. WT and mutant constructs were transfected in Human Embryonic Kidney 293T (HEK293T) cells and the extracted total RNA was subjected to reverse transcription (RT)-PCR, as previously described (**Chapter 2**).

Results

smMIPs performance

Sequence analysis of 16 STGD1 cases carrying known *ABCA4* pathogenic variants and four control samples using 483 smMIPs and a NextSeq 500 mid-output kit was performed to assess the performance of each smMIP and coverage per patient. The average read for these 20 DNA samples ranged from 10 to 37,551 per smMIP, with an overall average coverage of 2687x. A coverage plot of the smMIPs is provided in **Figure S2**. All previously known 34 variants present in the 32 alleles were identified robustly. All the targeted DI variants were found in ≥31% of the total reads ranging from 1521 reads (c.859-506G>C) to 17,538 reads (c.5196+1137G>A) (**Table S5**). Thereafter, 411 STGD1 proband samples were sequenced in two subsequent runs, with 226 samples (average coverage 511x) and 185 samples (average coverage 655x).

Identification of *ABCA4* variants

ABCA4 sequencing was performed using 483 smMIPs in 411 STGD1 persons, identifying 173 unique sequence variants. All variants and the respective cases were uploaded into the *ABCA4* LOVD at www.lovd.nl/ABCA4. Of these, 34 (20%) were new variants, including coding, NCSS and DI variants, as listed in **Table 1**. The 173 unique variants account for 534 alleles in STGD1 patients, details of which are given in **Table S6**. Of these 173 variants, 75% (n=131) are coding variants, comprised of missense (n=96), frameshift (n=23) and nonsense mutations (n=12). The other 25% (n=42) of variants are located in non-coding regions including canonical splice site, NCSS and DI variants.

The most common previously reported alleles were c.4253+43G>A (n=28), c.[5461-10T>C;5603A>T] (n=22), c.5882G>A (n=22), c.[1622T>C;3113C>T] (n=14), c.[2588G>C; 5603A>T] (n=13) and c.5196+1137G>A (n=13) (**Table S7**). The most frequent single variant allele was c.5603A>T (n=144), but its penetrance, when present in *trans* with a severe *ABCA4* variant, is incomplete in STGD1 families and in the Dutch population (see below).

Disease-associated *ABCA4* alleles in 411 STGD1 persons

Two or more pathogenic or likely pathogenic variants in *ABCA4* were found in 155 of 411 STGD1 cases, 97 (24%) of whom were considered solved as they carried two (likely) pathogenic variants. Another 58 (14%) cases were considered possibly solved as they carry c.5603A>T (p.Asn1868Ile), a frequent mild but low-penetrant variant (Zernant et al. 2017; Cremers et al. 2018; Runhart et al. 2018) in *trans* with a moderate or severe coding or DI variant. Details of the identified alleles in each person are given in **Table S7**. Sixty-seven individuals carry complex alleles (**Table S7** and **Table S8**). In 133 (32%) STGD1 cases only one *ABCA4* allele was identified whereas in 123 (30%) of the persons no variants were identified in *ABCA4*.

Novel rare variants identification and classification

Thirty-four novel variants were identified in 534 identified alleles (**Table 1**). Among these, 13 variants were missense mutations, identified in 14 alleles, of which c.4128G>C (p.Gln1376His) and c.4633A>G (p.Ser1545Gly), were also situated in consensus splice site sequences of exons 28 and 31, respectively. Thirteen frameshift mutations were identified in 19 alleles (among which c.4706del, p.Val1569Alafs*12 in 7 alleles), and three canonical splice site variants (c.442+1G>A, c.1760+1G>A and c.5313-2del). One stop mutation, p.Gly1717* was identified in two alleles. In addition, a synonymous variant (c.4539G>A) was located in the splice donor site (SDS) sequence of exon 30 in two alleles, and variant c.3191-19G>A was situated close to exon 22 but

not in consensus splice site sequences. Moreover, three novel DI variants (c.769-605T>C, c.4539+859C>T and c.4539+2065C>G) were identified in six alleles.

Splicing defects due to selected *ABCA4* non-canonical splice site variants

In vitro splice assays were performed in HEK293T cells to assess the impact of NCSS and novel near-exon at the transcript level by using eight WT constructs shown in **Figure 1**. Four of eight tested NCSS or near-exon splice site variants show splicing defects: c.3608G>A, c.4128G>C, c.4539G>A and c.4849G>A. RT-PCR and Sanger sequencing results are provided in **Figure 2**. Variant c.3608G>A, described previously by (Cornelis et al. 2017) was predicted to reduce the strength of the SAS of exon 25 and the creation of a cryptic SAS 2 nt downstream of the canonical SAS (**Table S2**). It was tested in WT construct BA17 and RT-PCR was performed by using *ABCA4* exonic primers located in exons 23 and 26. A fragment of 357 nt corresponding to the WT band and a 151-nt fragment in which exon 25 is skipped were observed (**Figure 2A-C**). The novel missense variants residing in splice site consensus sequences, i.e., c.4128G>C and c.4633A>G, were tested in WT constructs BA19 and BA21 respectively. Variant c.4128G>C, predicted to weaken the SDS of exon 27, resulted in a single band corresponding to a 12-nt elongation of exon 27 due to the presence of a cryptic SDS (**Figure 2D-F**). Variant c.4633A>G, predicted to significantly weaken the SDS of exon 31, showed no splicing defect in HEK293T cells (**Figure S3**).

The novel synonymous variant c.4539G>A, predicted to impair the SDS of exon 30 (**Table S2**), was tested in midigene BA20. RT-PCR was performed by using primers residing in exons 28 and 31, and variant c.4539G>A showed three bands compared to the WT which upon sequencing revealed a band of 441 nt corresponding to WT, a band of 368 nt corresponding to a deletion of 73 nt of the 3' end of exon 30 and a third band of 155 nt corresponding to the deletion of exons 29 and 30 (**Figure 2G-I**).

Variant c.4849G>A, also described previously (Downs et al. 2007) and predicted to reduce the SAS of exon 35, resulted in three fragments when tested in WT BA23. Sanger sequencing revealed a fragment of 578 nt corresponding to the correct transcript, a band of 408 nt corresponding to exon 35 skipping and a third band of 317 nt corresponding to the skipping of exon 35 and a 91-bp reduction of exon 36 due to the recognition of an internal SAS at position c.5110 (**Figure 2K-O**).

The other novel variant c.3191-19G>A, tested in BA16, did not show a splicing defect (**Figure S3**). In addition to this novel variant, the previously reported variant c.6148G>C (Schulz et al. 2017) did not show a splicing defect when tested using BA31 (**Figure S3**).

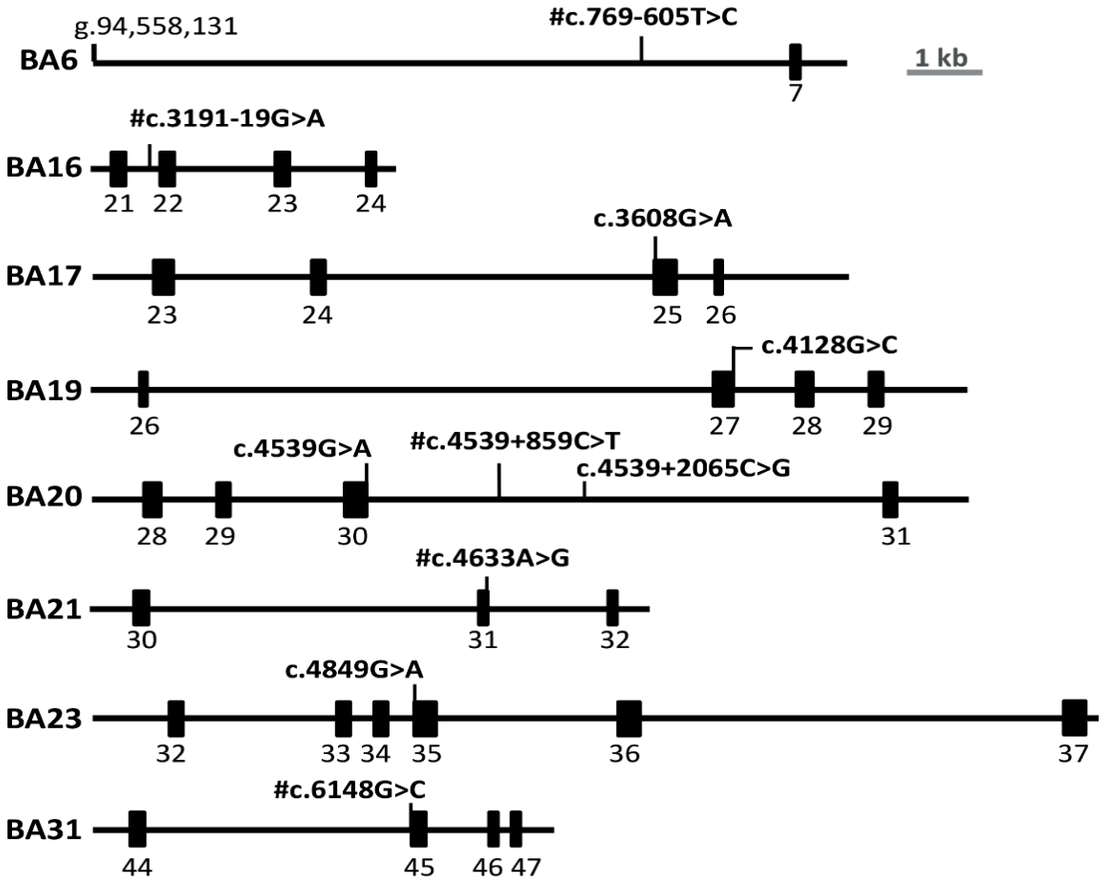


Figure 1. Overview of midigene splice constructs containing novel variants. #Variants showing no splicing defects when tested in HEK293T cells (**Figure S3**). Exons are represented as black rectangles.

Table 1. Novel rare variants found in study cohort and their classification.

Genomic position (hg 19)	c.DNA variant	Protein variant	AF in study	gnomAD_AF (nFE)	*In silico* analysis			ACMG class
					SIFT (scores)	Mutation Taster	PolyPhen-2 (scores)	
94574132	c.442+1G>A	p.(?)	0.00122	-	n.a.	n.a.	n.a.	5
94564408	c.710T>C	p.(Leu237Pro)	0.00122	0.000008952	Deleterious (0)	Disease causing	Prob. dam. (1.000)	3
94549602	c.769-605T>C	p.(=)	0.00365	0.00612925	n.a.	n.a.	n.a.	1
94544258	c.1244A>G	p.(Asn415Ser)	0.00122	-	Deleterious (0)	Disease causing	Pos. dam. (0.870)	4
94544253	c.1249A>G	p.(Thr417Ala)	0.00122	-	Deleterious (0)	Disease causing	Pos. dam. (**0.745**)	4
94543346	c.1454del	p.(Gly485Alafs*83)	0.00122	-	n.a.	n.a.	n.a.	5
94528667	c.1760+1G>A	p.(?)	0.00122	-	n.a.	n.a.	n.a.	5
94520760	c.2494G>T	p.(Asp832Tyr)	0.00122	-	Deleterious (0)	Disease causing	Pos. dam. (0.797)	3
94520684	c.2570T>C	p.(Leu857Pro)	0.00122	-	Deleterious (0.01)	Disease causing	Prob. dam. (**0.969**)	3
94517247_94517248	c.2594_2595del	p.(Tyr865Trpfs*19)	0.00122	-	n.a.	n.a.	n.a.	5
94512608	c.2785dup	p.(Val929Glyfs*11)	0.00122	-	n.a.	n.a.	n.a.	5
94508473	c.3191-19G>A	p.(=)	0.00122	0.0001264	n.a.	n.a.	n.a.	3
94506943	c.3344T>C	p.(Met1115Thr)	0.00122	-	Deleterious (0.01)	Disease causing	Benign (0.062)	5
94506788	c.3499del	p.(Gln1167Argfs*29)	0.00122	-	n.a.	n.a.	n.a.	5
94502847	c.3667dup	p.(Glu1223Glyfs*14)	0.00122	-	n.a.	n.a.	n.a.	5
94497515	c.3947A>G	p.(Asp1316Gly)	0.00122	-	Tolerated (0.41)	Polymorphism	Benign (0.00)	3
94497514	c.3948C>G	p.(Asp1316Glu)	0.00122	-	Tolerated (1)	Polymorphism	Benign (0.00)	3
94497334	c.4128G>C	p.(Gln1376His)	0.00122	-	Deleterious (0)	Disease causing	Prob. dam. (1.000)	5
94495001	c.4539G>A	p.[Cys1490Glufs*12, =]	0.00243	-	n.a.	n.a.	n.a.	5
94494142	c.4539+859C>T	p.(=)	0.03285	-	n.a.	n.a.	n.a.	3
94492936	c.4539+2065C>G	p.[Arg1514Lysfs*35, =]	0.00122	-	n.a.	n.a.	n.a.	5
94490511	c.4633A>G	p.(Ser1545Gly)	0.00122	-	Deleterious (0)	Disease causing	Benign (0.002)	5
94492937	c.4706del	p.(Val1569Alafs*12)	0.00852	-	n.a.	n.a.	n.a.	5
94486887	c.4927del	p.(Leu1643Cysfs*19)	0.00122	-	n.a.	n.a.	n.a.	5
94485185	c.5149G>T	p.(Gly1717*)	0.00243	-	n.a.	n.a.	n.a.	5
94485159	c.5175dup	p.(Thr1726Aspfs*61)	0.00122	-	n.a.	n.a.	n.a.	5
94480248	c.5313-2del	p.(?)	0.00243	-	n.a.	n.a.	n.a.	5
94480230	c.5329A>T	p.(Met1777Leu)	0.00122	0.0001263	Deleterious (0)	Disease causing	Benign (0.008)	3
94480106	c.5453del	p.(Asn1818Ilefs*12)	0.00122	-	n.a.	n.a.	n.a.	5
94476893	c.5509C>T	p.(Pro1837Ser)	0.00243	-	Deleterious (0)	Disease causing	Prob. dam. (1.000)	5
94476366	c.5704dup	p.(Leu1902Profs*10)	0.00122	-	n.a.	n.a.	n.a.	5
94471044	c.6100del	p.(Tyr2034Thrfs*27)	0.00122	-	n.a.	n.a.	n.a.	5
94467468	c.6228del	p.(Lys2076Asnfs*39)	0.00122	-	n.a.	n.a.	n.a.	5
94467442	c.6254T>C	p.(Leu2085Pro)	0.00122	-	Deleterious (0)	Disease causing	Prob. dam. (0.998)	4

Table 1. Novel rare variants found in study cohort and their classification. Data for all the novel coding variants identified in this study is provided with their genomic and cDNA positions (human genome version 19; hg19) and predicted effect at the protein level. SIFT scores range from 0 to 1 and amino acid substitutions are predicted damaging if the score is ≤ 0.05 and tolerated if the score is > 0.05. Mutation Taster had probability-values of 1 for Disease causing variants and those designated to be Polymorphic. PolyPhen-2 scores range between 0.000 to 1.000 are shown (scores close to 1 indicate damaging effect of variants). All the variants were classified according to the ACMG guidelines from Class 1 -5 (1, benign; 2, likely benign; 3, variant of unknown significance; 4, likely pathogenic; 5, pathogenic). AF, allele frequency; nFE, non-Finish European; n.a., not applicable; SIFT, Sorting Intolerant From Tolerant; ACMG, American College of Medical Genetics and Genomics; Prob.dam, probably damaging; Pos. dam, possibly damaging.

Deep-intronic variants identification and *in vitro* assessment

To investigate the occurrence of DI variants among 411 STGD1 persons, 12 selected intronic regions carrying 14 previously identified variants were targeted by smMIPs. Among 534 alleles, 67 alleles carrying different DI variants were identified (**Table S7**). Ten already reported DI variants were identified in 61 (15%) alleles, i.e., c.769-784C>T (4 alleles), c.4253+43G>A (28 alleles), c.4539+1106C>T (1 allele), c.4539+2001G>A (3 alleles), c.4539+2028C>T (1 allele), c.4539+2064C>T (7 alleles), c.5196+1056A>G (1 allele), c.5196+1136C>A (1 allele), c.5196+1137G>A (13 alleles) and c.5196+1159G>A (2 alleles).

In addition, three novel DI variants, i.e., c.769-605T>C (3 alleles), c.4539+859C>T (2 alleles) and c.4539+2065C>G (1 allele), were found in the regions containing known causal variants. Variant c.769-605C>T, showing an allele frequency of 0.00613 in non-Finnish Europeans, was located 12 nt downstream of a PE generated by c.769-784C>T. It was tested in WT construct BA6 but did not show a splicing defect (**Figure S3**). Similarly, variant c.4539+859C>T, showing an allele frequency of 0.01123 in non-Finnish Europeans in gnomAD, was located upstream of variant c.4539+1106C>T and increased the strength of the existing cryptic SAS in the intronic region. It was tested in WT construct BA20 but did not show any splicing defect (**Figure S3**).

Conversely, variant c.4539+2065C>G (not found in 7,713 non-Finnish Europeans), located next to c.4539+2064C>T, creates a new SD and alters the strength of ESEs in the region predicted by Alamut visual. It was also tested in WT construct BA20 and RT-PCR was performed by using *ABCA4* exonic primers located in exons 28 and 31. Gel analysis and Sanger sequencing revealed a band of 441 nt corresponding to the WT and 611 nt corresponding to a 170-nt PE insertion located between (c.4539+1891 and c.4539+2060) that was observed in most of the cDNA product and led to a frameshift p.[Arg1514Lysfs*35,=] (**Figure 2G-J**).

Therefore, a total of 11 causal DI variants were found in 62 STGD1 cases. Most carried DI variants on one allele and a severe or a moderately severe *ABCA4* variant on the other allele. Two patients carried causal DI variants on both alleles, i.e., individual 067332 (c.4539+2064C>T, p.[=,Arg1514Leufs*36] and c.5196+1137G>A, p.[=,Met1733Glufs*78]) and individual 067241 (c.4539+2064C>T and c.4253+43G>A, p.[=,Ile1377Hisfs*3]). Two other individuals, i.e., 066666 and 066688, carried c.4539+859C>T, p.(?) and c.4253+43G>A in a compound heterozygous manner, but the former variant was not shown to result in a splicing defect.

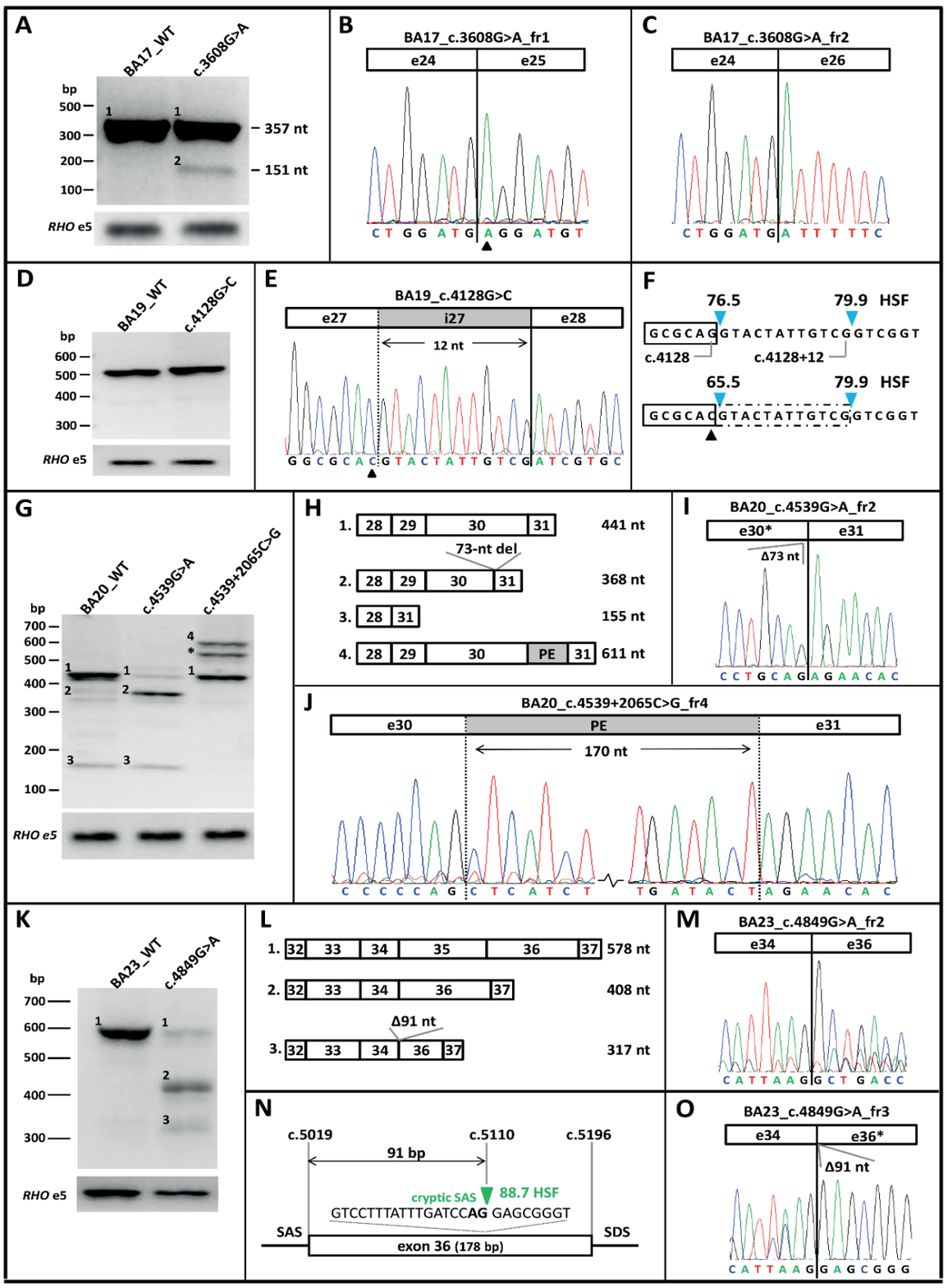


Figure 2. Overview of splicing defects due to three non-canonical splice site variants and one deep-intronic variant in *ABCA4*. All wild-type (WT) and mutant (Mut) midigenes were transfected in HEK293T cells and their RNA subjected to RT-PCR. **A.** RT-PCR for WT and c.3608G>A Mut BA17 midigene showed a splicing defect when using primers in *ABCA4* exons 23 and 26. Fragment 1 of 357 nt corresponds to the correct mRNA and fragment 2 of 151 nt

corresponds to exon 25 skipping. **B, C.** Sanger sequencing of mutant fragment 1 with a missense change (c.3608G>A) highlighted by a triangle, and fragment 2 corresponding to the skipping of exon 25. **D, E, F.** RT-PCR and Sanger sequencing of cDNA corresponding to mutant c.4128G>C resulting in a 12-nt exon 27 elongation. Human Splice Finder (HSF) splice site scores (blue arrowheads) for WT and mutant sites. **G, H, I, J.** RT-PCR for WT and Mut (c.4539G>A and c.4539+2065C>G) midigenes showed a complex splice pattern when analysed using primers in *ABCA4* exons 28 and 31. Variant at the last position of exon 30 resulted in a correct transcript (fragment 1), a deletion of 73 nt of the 3' end of exon 30 (fragment 2) and a deletion of both exons 29 and 30 (fragment 3). Variant c.4539+2065C>G, in addition to the correct fragment 1, resulted in a 170-nt pseudoexon (PE) inclusion between exons 30 and 31 (fragment 4). Asterisks denote to a heteroduplex fragment consisting of fragments 1 and 4. **K, L.** RT-PCR products of WT and Mut BA23 containing a non-canonical splice site variant at the first nucleotide of exon 35. Primers in exon 32 and 37 revealed skipping of exon 35 alone (fragment 2) or skipping of exon 35 in combination with a 91-nt deletion of the 5' end of exon 36 (fragment 3). **M.** Exon 35 skipping confirmed by Sanger sequencing of fragment 2. **N, O.** The 91-nt deletion of exon 36 is the result of the use of a cryptic splice acceptor site (SAS) located at position c.5110, as confirmed by Sanger sequencing.

Discussion

To identify the missing causal variants in STGD1 cases, all 50 exons and 12 selected intronic regions of *ABCA4* were sequenced in 411 previously genotyped cases by employing 483 overlapping smMIPs. In total, 173 unique sequence variants were identified in 534 alleles, solving at least 24% of the cases. Twenty percent of the variants were novel. The novel NCSS variant c.4128G>C showed a 12-nt elongation of exon 27 and variant c.4539G>A resulted in skipping of exon 30.

Ten previously detected DI variants and the novel variant c.4539+2065C>G were identified in 15% (n=62) of the alleles. Variants c.4253+43G>A and c.5196+1137G>A were identified in 28 and 13 alleles, respectively, highlighting their relatively high frequency in these patient cohorts. The novel DI variant c.4539+2065C>G led to the generation of a 170-nt PE in intron 30. This PE was observed as part of one of three splicing products due to c.4539+2064C>T, i.e., a 244-nt PE consisting of the 170-nt PE associated with c.4539+2065C>G, as well as a neighboring 74-nt PE (Bauwens et al. 2019).

A common c.5603A>T (p.Asn1868Ile) variant, strongly associated with late-onset STGD1 (Zernant et al. 2017; Cremers et al. 2018; Runhart et al. 2018), was found in a heterozygous manner as a single variant in 144 STGD1 cases. It was previously shown that p.Asn1868Ile contributes to the pathogenicity of variants c.769-784C>T and c.2588G>C, which were consistently found in *cis* in patients and much less frequent in healthy persons (Zernant et al. 2017; **Chapter 3**).

A pathogenic role was established for one of the three newly identified DI variants, c.4539+2065C>G, when tested in HEK293T cells. However, the disease causing effect of the other two variants c.769-605T>C and c.4539+859C>T cannot be ruled out based on *in vitro* midigene splice assays, as many DI variants have shown to be disease causing only when tested in patient-derived cells due to retina specific splice factors which are missing in HEK293T cells (Albert et al. 2018).

Despite the identification of many causal variants, 62% of STGD1 cases remained unsolved either with one (32%) or no (30%) variant. It is of note that the 411 tested probands originate from a total of 1468 STGD1 and macular dystrophy cases that were screened previously using different genotyping methods, i.e., exon and splice site mutation scanning for 1133 French cases, and exon and splice site sequencing for 335 German cases. This means that the 411 tested probands represent a highly biased patient cohort. Nevertheless, we identified many new variants, including NCSS and DI variants. The absence of *ABCA4* variants in the unsolved cases could first of all be due to genocopies, as several other genes have been implicated in STGD-like phenotypes (Zaneveld et al. 2015; Ma et al. 2019). In view of the high AF of *ABCA4* variants (5%) in the general population (Maugeri et al. 1999; Cornelis et al. 2017), our cohort may contain 40 mono-allelic *ABCA4* cases that in fact are due to variants in these other genes. Second, we also may have missed non-coding variants residing in non-covered sequences as the smMIPs covered ~16,500 bp, leaving ~111,500 bp of intronic sequences unscreened. Third, we may have employed too stringent variant selection criteria when using *in silico* algorithms. Fourth, the midigene *in vitro* splice assay that is performed in human embryonic kidney cells is insensitive to retina-specific splicing defects (Albert et al. 2018). Finally, several types of structural variations such as inversions and insertions may have been missed as they are refractory to smMIPs detection.

By employing smMIPs-based *ABCA4* sequencing, a large number of cases could be simultaneously sequenced for variants in ~11,500 bp, with an average coverage of 583x. In recent studies (Weisschuh et al. 2018; Tayebi et al. 2019), 647,547 bp was sequenced using 6129 MIPs targeting 1,524 coding regions of 108 retinal dystrophy-associated genes, with an average coverage of 213x per MIP. This is comparable to other targeted sequencing methods such as WES or Haloplex-based sequencing. smMIPs-based *ABCA4* sequencing is a cost-effective method compared to Sanger sequencing (≥ €500 for 50 exons) and WES, since it costs ~€20 for reagents only (excluding the smMIPs design and synthesis) to sequence *ABCA4* 50 exons and 12 DI regions which is 25-50 fold less than Sanger sequencing or WES (Schwarze et al. 2018). Other advantages of smMIPs are the low input of DNA per patient. We used 100 ng, but as low as 19 ng of DNA can be sufficient (Eijkelenboom et al. 2016). Moreover, due

to the small target size (140 bp including the annealing primers) many fragmented DNA samples could also be effectively used for sequence analysis. Finally, the ability to detect CNVs with smMIPs data was assessed previously. By including 5 positive control samples in 3 different sequencing runs with a mean coverage per smMIP of 359x, analytical sensitivity and specificity of 100% and 88% was obtained respectively (Neveling et al. 2017). Flexibility, automation and its sensitivity for single nucleotide variants and CNVs render the smMIP technology very attractive for diagnostic requirements. A detailed comparison of smMIPs-based sequencing with other targeted and non-targeted sequencing methods is given in **Table S9**. smMIPs-based sequencing can be the preferred method when the number of samples to be sequenced is (very) high, additional targets should be added in the future, DNA quality and quantity are low, and a next-generation sequencing platform is available for high throughput analysis.

In conclusion, we identified causal DI variants in 15% of our genetically unexplained STGD1 cases, two of which (c.4253+43G>A and c.5196+1137G>A) are frequent. In addition, we identified four exonic NCSS variants that resulted in splicing defects and one novel DI variant resulting in PE formation. Interestingly, many studies have shown the correction of *ABCA4* splicing defects caused by DI variants at the transcript level both *in vitro* and *in vivo*, providing a basis for the development of new therapeutic strategies for individuals with STGD1 carrying DI variants (Albert et al. 2018; **Chapter 3**). In order to find the missing variants in the DI regions, additional smMIPs can be designed to cover the complete *ABCA4* gene.

Acknowledgements

We thank Béatrice Bocquet, Hélène Dollfus, Isabelle Drumare, Emeline Gorecki, Christian P. Hamel, Karsten Hufendiek, Cord Huchzermeyer, Herbert Jägle, Ulrich Kellner, Valérie Pelletier, Yaumara Perdromo, Charlene Piriou, Philipp Rating, Klaus Rüther, Eric Souied, Georg Spital and Xavier Zanlonghi for their cooperation and ascertaining STGD1 cases. SmMIP hybridization, library preparation and NextSeq 500 sequencing was performed at the Genome Technology Center, Department of Human Genetics, Radboud university medical center, Nijmegen.

Supplemental data

Supplemental data are not included in this thesis and they can be found online at the following link: https://onlinelibrary.wiley.com/doi/full/10.1002/humu.23787

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**: 248-249.

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

Allikmets R, Singh N, Sun H, Shroyer NF, 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.

Bauwens M, De Zaeytijd J, Weisschuh N, Kohl S, Meire F, Dahan K, Depasse F, De Jaegere S, De Ravel T, De Rademaeker M et al. 2015. An augmented ABCA4 screen targeting noncoding regions reveals a deep intronic founder variant in Belgian Stargardt patients. *Hum Mutat* **36**: 39-42.

Bauwens M, Garanto A, Sangermano R, Naessens S, Weisschuh N, De Zaeytijd J, Khan M, Sadler F, Balikova I, Van Cauwenbergh C et al. 2019. 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* doi:10.1038/s41436-018-0420-y.

Bax NM, Sangermano R, Roosing S, Thiadens AA, Hoefsloot LH, van den Born LI, Phan M, Klevering BJ, Westeneng-van Haaften C, Braun TA et al. 2015. Heterozygous deep-intronic variants and deletions in ABCA4 in persons with retinal dystrophies and one exonic ABCA4 variant. *Hum Mutat* **36**: 43-47.

Boyle EA, O'Roak BJ, Martin BK, Kumar A, Shendure J. 2014. MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. *Bioinformatics* **30**: 2670-2672.

Braun TA, Mullins RF, Wagner AH, Andorf JL, Johnston RM, Bakall BB, Deluca AP, Fishman GA, Lam BL, Weleber RG et al. 2013. Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. *Hum Mol Genet* **22**: 5136-5145.

Carss KJ, Arno G, Erwood M, Stephens J, Sanchis-Juan A, Hull S, Megy K, Grozeva D, Dewhurst E, Malka S et al. 2017. Comprehensive Rare Variant Analysis via Whole-Genome Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease. *Am J Hum Genet* **100**: 75-90.

Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. 2003. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* **31**: 3568-3571.

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

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 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.

Cremers FPM, Cornelis SS, Runhart EH, Astuti GDN. 2018. Author Response: Penetrance of the ABCA4 p.Asn1868Ile Allele in Stargardt Disease. *Invest Ophthalmol Vis Sci* **59**: 5566-5568.

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**: e67.

Downs K, Zacks DN, Caruso R, Karoukis AJ, Branham K, Yashar BM, Haimann MH, Trzupek K, Meltzer M, Blain D et al. 2007. Molecular testing for hereditary retinal disease as part of clinical care. *Arch Ophthalmol-Chic* **125**: 252-258.

Eijkelenboom A, Kamping EJ, Kastner-van Raaij AW, Hendriks-Cornelissen SJ, Neveling K, Kuiper RP, Hoischen A, Nelen MR, Ligtenberg MJL, Tops BBJ. 2016. Reliable Next-Generation Sequencing of Formalin-Fixed, Paraffin-Embedded Tissue Using Single Molecule Tags. *J Mol Diagn* **18**: 851-863.

Hiatt JB, Pritchard CC, Salipante SJ, O'Roak BJ, Shendure J. 2013. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. *Genome Res* **23**: 843-854.

Hussain RM, Ciulla TA, Berrocal AM, Gregori NZ, Flynn HW, Lam BL. 2018. Stargardt macular dystrophy and evolving therapies. *Expert Opin Biol Th* **18**: 1049-1059.

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

Klevering BJ, Yzer S, Rohrschneider K, Zonneveld M, Allikmets R, van den Born LI, Maugeri A, Hoyng CB, Cremers FP. 2004. Microarray-based mutation analysis of the ABCA4 (ABCR) gene in autosomal recessive cone-rod dystrophy and retinitis pigmentosa. *Eur J Hum Genet* **12**: 1024-1032.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**: 1073-1082.

Lelieveld SH, Reijnders MR, Pfundt R, Yntema HG, Kamsteeg EJ, de Vries P, de Vries BB, Willemsen MH, Kleefstra T, Lohner K et al. 2016. Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat Neurosci* **19**: 1194-1196.

Levy SE, Myers RM. 2016. Advancements in Next-Generation Sequencing. *Annu Rev Genomics Hum Genet* **17**: 95-115.

Lin X, Tang W, Ahmad S, Lu J, Colby CC, Zhu J, Yu Q. 2012. Applications of targeted gene capture and next-generation sequencing technologies in studies of human deafness and other genetic disabilities. *Hear Res* **288**: 67-76.

Ma CJ, Lee W, Stong N, Zernant J, Chang S, Goldstein D, Nagasaki T, Allikmets R. 2019. 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* doi:10.1101/mcs.a003624.

Maia-Lopes S, Aguirre-Lamban J, Castelo-Branco M, Riveiro-Alvarez R, Ayuso C, Silva ED. 2009. ABCA4 mutations in Portuguese Stargardt patients: identification of new mutations and their phenotypic analysis. *Mol Vis* **15**: 584-591.

Maugeri A, van Driel MA, van de Pol DJR, Klevering BJ, van Haren FJJ, Tijmes N, Bergen AAB, Rohrschneider K, Blankenagel A, Pinckers AJLG 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. *American Journal of Human Genetics* **64**: 1024-1035.

Neveling K, den Hollander AI, Cremers FP, Collin RW. 2013. Identification and analysis of inherited retinal disease genes. *Methods Mol Biol* **935**: 3-23.

Neveling K, Mensenkamp AR, Derks R, Kwint M, Ouchene H, Steehouwer M, van Lier B, Bosgoed E, Rikken A, Tychon M et al. 2017. BRCA Testing by Single-Molecule Molecular Inversion Probes. *Clin Chem* **63**: 503-512.

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

Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. *J Comput Biol* **4**: 311-323.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E et al. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**: 405-424.

Runhart EH, Sangermano R, Cornelis SS, Verheij J, Plomp AS, Boon CJF, Lugtenberg D, Roosing S, Bax NM, Blokland EAW et al. 2018. The Common ABCA4 Variant p.Asn1868Ile Shows Nonpenetrance and Variable Expression of Stargardt Disease When Present in trans With Severe Variants. *Invest Ophthalmol Vis Sci* **59**: 3220-3231.

Sangermano R, Garanto A, Khan M, Runhart EH, Bauwens M, Bax NM, van den Born LI, Khan MI, Cornelis SS, Verheij J et al. 2019. Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. *Genet Med* doi:10.1038/s41436-018-0414-9.

Sangermano R, Khan M, Cornelis SS, Richelle V, Albert S, Garanto A, Elmelik D, Qamar R, Lugtenberg D, van den Born LI et al. 2018. ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Res* **28**: 100-110.

Schulz HL, Grassmann F, Kellner U, Spital G, Ruther K, Jagle H, Hufendiek K, Rating P, Huchzermeyer C, Baier MJ et al. 2017. Mutation Spectrum of the ABCA4 Gene in 335 Stargardt Disease Patients From a Multicenter German Cohort-Impact of Selected Deep Intronic Variants and Common SNPs. *Invest Ophthalmol Vis Sci* **58**: 394-403.

Schwarz JM, Cooper DN, Schuelke M, Seelow D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature Methods* **11**: 361-362.

Schwarze K, Buchanan J, Taylor JC, Wordsworth S. 2018. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. *Genet Med* doi:10.1038/gim.2017.247.

Tanna P, Strauss RW, Fujinami K, Michaelides M. 2017. Stargardt disease: clinical features, molecular genetics, animal models and therapeutic options. *Brit J Ophthalmol* **101**: 25-30.

Tayebi N, Akinrinade O, Khan MI, Hejazifar A, Dehghani A, Cremers FPM, Akhlaghi M. 2019. Targeted next generation sequencing reveals genetic defects underlying inherited retinal disease in Iranian families. *Mol Vis* **25**: 106-117.

Weisschuh N, Feldhaus B, Khan MI, Cremers FPM, Kohl S, Wissinger B, Zobor D. 2018. Molecular and clinical analysis of 27 German patients with Leber congenital amaurosis. *PLoS One* **13**: e0205380.

Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* **11**: 377-394.

Zaneveld J, Siddiqui S, Li H, Wang X, Wang H, Wang K, Li H, Ren H, Lopez I, Dorfman A et al. 2015. Comprehensive analysis of patients with Stargardt macular dystrophy reveals new genotype-phenotype correlations and unexpected diagnostic revisions. *Genet Med* **17**: 262-270.

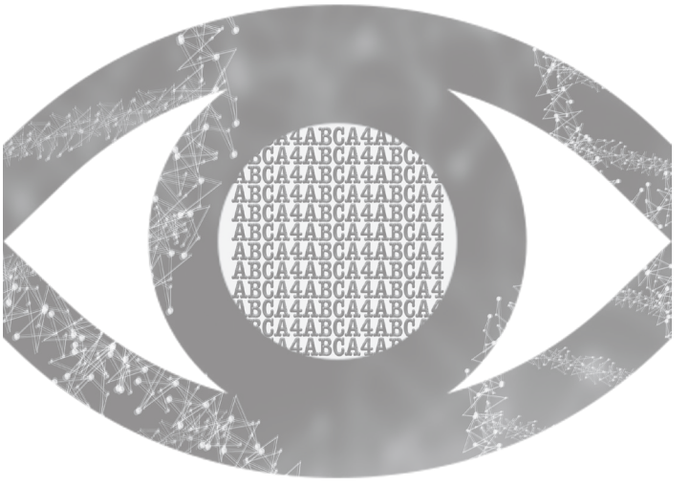
Zernant J, Lee W, Collison FT, Fishman GA, Sergeev YV, Schuerch K, Sparrow JR, Tsang SH, Allikmets R. 2017. Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration. *J Med Genet* **54**: 404-412.

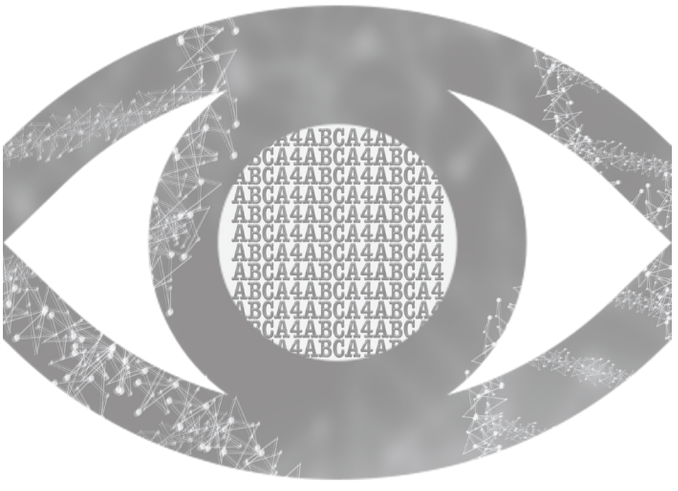
Zernant J, Lee W, Nagasaki T, Collison FT, Fishman GA, Bertelsen M, Rosenberg T, Gouras P, Tsang SH, Allikmets R. 2018. Extremely hypomorphic and severe deep intronic variants in the ABCA4 locus result in varying Stargardt disease phenotypes. *Cold Spring Harb Mol Case Stud* **4**.

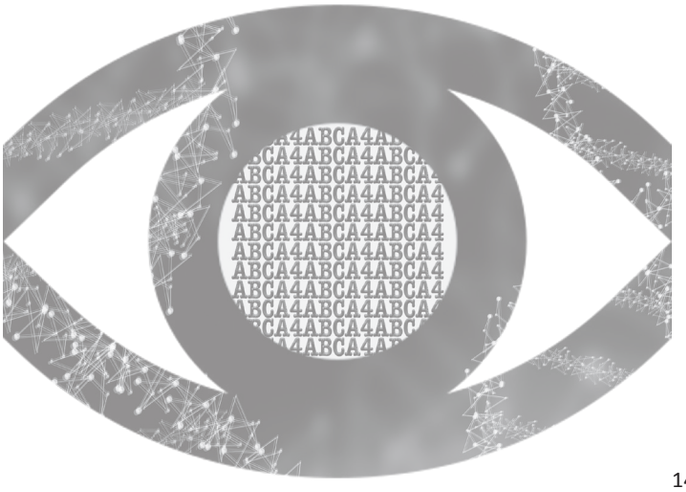
Zernant J, Xie YA, Ayuso C, Riveiro-Alvarez R, Lopez-Martinez MA, Simonelli F, Testa F, Gorin MB, Strom SP, Bertelsen M et al. 2014. Analysis of the ABCA4 genomic locus in Stargardt disease. *Hum Mol Genet* **23**: 6797-6806.

Chapter 5

Detailed phenotyping and therapeutic strategies for intron 36 variants in *ABCA4* underlying *ABCA4*-associated retinal disease







Detailed phenotyping and therapeutic strategies for intronic *ABCA4* variants in Stargardt disease

Mubeen Khan*, Gavin Arno*, Ana Fakin*, David A. Parfitt, Patty P.A. Dhooge, Silvia Albert, Nathalie M. Bax, Lonneke Duijkers, Michael Niblock, Kwan L. Hau, Edward Bloch, Elena R. Schiff, Davide Piccolo, Michael C. Hogden[7], Carel B. Hoyng, Andrew R. Webster, Frans P.M. Cremers, Michael E. Cheetham, Alejandro Garanto[#] and Rob W.J. Collin[#]

Manuscript in preparation

*These authors contributed equally to this work

[#]Shared senior authors

Abstract

Stargardt disease, or *ABCA4*-associated disease, is a retinal disorder characterized by a progressive loss of photoreceptor cells and the underlying retinal pigment epithelium, often leading to severe visual impairment. This disease is caused by bi-allelic mutations in the *ABCA4* gene that encodes the ATP binding cassette type A4 transporter protein. Its tremendous clinical heterogeneity to a large extent can be explained by the allelic heterogeneity, i.e., different *ABCA4* variants result in varying residual activity of the ABCA4 protein. Over the last few years, we and others have identified several pathogenic *ABCA4* variants that reside within the introns of *ABCA4*, including a recurrent variant in intron 36 (c.5196+1137G>A). Genetic and clinical characterization of this variant revealed that it is present on at least three independent haplotypes, and can give rise to various clinical subtypes, to some degree depending on the severity of the second *ABCA4* allele. Moreover, we discovered several additional *ABCA4* variants clustering in intron 36. These variants all result in aberrant splicing of *ABCA4*, i.e., the inclusion of pseudoexons (PEs), albeit to a different extent. The splicing defects caused by the recurrent c.5196+1137G>A variant strongly increased upon differentiation of patient-derived induced pluripotent stem cells into retina-like cells. Finally, all splicing defects could be rescued by the administration of antisense oligonucleotides that were designed to specifically block the PE insertion, including rescue in three-dimensional retinal organoids harboring the c.5196+1137G>A variant. Together, our data further illustrate the importance of intronic variants in *ABCA4* underlying retinal disease and expand the therapeutic possibilities to overcome splicing defects caused by these mutations.

Introduction

Stargardt disease (STGD1), more and more often being referred to as *ABCA4*-associated retinal disease, is a progressive disorder of the retina, initially characterized by a loss of central vision (Tanna et al. 2017). It is the most common form of inherited macular dystrophy and is believed to affect ~1 in 10,000 individuals worldwide. There is a high degree of clinical heterogeneity, in many patients resulting in severe visual impairment or even complete blindness, due to death of the photoreceptors and the retinal pigment epithelium (RPE) cells .

ABCA4-associated retinal disease retinopathy is caused by bi-allelic variants in *ABCA4*, a 50-exon gene that codes for the ATP-Binding Cassette, Subfamily A, Member 4, a transmembrane protein of 2,273 amino acids (*ABCA4*, previously known as ABCR (Allikmets et al. 1997). The protein belongs to the superfamily of ATP-binding cassette transporters, (Allikmets et al. 1997) and within the retinal photoreceptor cells, it is located at the rim of the outer segment disks in rods and lamellae in cones where it acts as a flippase facilitating the active transport of potentially toxic retinoids (Quazi et al. 2012). Dysfunction or absence of ABCA4 leads to an accumulation of these toxic by-products of the visual cycle, commonly known as lipofuscin in the RPE after outer segment phagocytosis. Another potential mechanism of disease is direct toxicity toward the cone photoreceptors due to their different anatomical structure (Conley et al. 2012).

ABCA4 displays a tremendous allelic heterogeneity, with more than 1000 pathogenic mutations identified to date (accessible in https://databases.lovd.nl/shared/variants/ABCA4/) (Cornelis et al. 2017). There is a strong correlation reported between the clinical appearance of and the severity of the *ABCA4* genotype in a given patient (Cremers et al. 1998; Fakin et al. 2016). In fact, some variants lead to complete loss of function of the ABCA4 and result in severe disease (Fakin et al. 2016), whereas other variants are very mild, and in some cases even show reduced penetrance in the absence of other *cis*-acting *ABCA4* variants (Runhart et al. 2018).

Over the last decades, we and others have identified several *ABCA4* variants that are located outside the protein-coding exons of the gene (Bauwens et al. 2015; Bax et al. 2015; Schulz et al. 2017; Albert et al. 2018; Zernant et al. 2018; Bauwens et al. 2019; **Chapter 3**). The majority of these are deep-intronic (DI) pathogenic variants that result in the activation of a splice site within an intron leading to the insertion of an aberrant pseudoexon (PE) into the *ABCA4* pre-mRNA. Following splicing, these PEs often result in disruption of the reading frame and may be expected to result in nonsense mediated decay (NMD) of the aberrant transcript (Zhang et al. 2009). Thus, these mutations can

lead to reduced, or even absence, of functional ABCA4 protein dependent on the extent of PE inclusion. Intriguingly, the recognition of the PEs by the splicing machinery can vary significantly, both in terms of the quantity of PE inclusion (**Chapter 3**) as well as tissue-dependency (Caminsky et al. 2014; Albert et al. 2018). In 2013, Braun and colleagues were the first to report pathogenic DI *ABCA4* variants, including a change in intron 36 (c.5196+1137G>A, reported there as 'V1') that was recurrently present in their primary as well as their validation cohort of patients with *ABCA4*-disease (Braun et al. 2013). This variant, as well as a second closely located variant (c.5196+1216C>A), each strengthened alternative splicing and inclusion of a 73-nt PE into a small fraction of *ABCA4* transcripts in keratinocytes derived from patients compound heterozygous for one of these alleles.

The pathogenic mechanism of PE inclusion consequent upon DI variants has the potential to be highly amenable for splicing modulating therapeutic approaches using antisense oligonucleotides (AONs). These relatively small RNA molecules of 18-25 nucleotides can bind complementarily to their target pre-mRNA, preventing the inclusion of the PE into the final mRNA transcript upon splicing (Collin and Garanto 2017). AONs can be considered safe and well-tolerated when delivered to the retina, as shown in the results of a recent phase I/II clinical trial for a recurrent DI variant in *CEP290* (Cideciyan et al. 2019).

In this study, we performed a detailed phenotypic characterization of a substantial cohort of *ABCA4* patients harboring the c.5196+1137G>A variant, and predicted its severity based on the clinical presentation in patients who harbored a null variant on the other allele. In addition, we expanded the search for DI mutations in intron 36 and assessed their pathogenicity based on their ability to dysregulate *ABCA4* pre-mRNA splicing in both midigene splicing assays as well as patient-derived retina-like cells (Sangermano et al. 2016). Finally, for four variants that cause splicing defects, these defects were rescued by the administration of AONs, thereby opening new avenues for treating selected subgroups of *ABCA4* patients.

Materials and Methods

Subjects

Twenty-five individuals (twenty-four probands) with ABCA4-disease, all carrying a recurrent variant in intron 36 of *ABCA4* (GRCh37 [hg19]:1:94484001C>T, NM_000350.3: c.5196+1137G>A) were recruited for this study. Twenty patients were ascertained from the genetic records of the Inherited Eye Disease clinics at Moorfields Eye Hospital London, UK, while four (three probands and one sibling) were ascertained

from the Radboud University Medical Center in Nijmegen, The Netherlands. One individual was identified and recruited from a cohort at the Princess Alexandra Hospital in Brisbane, Australia. For all British and Australian individuals, *ABCA4* mutation screening was performed by next generation sequencing. Specifically, five individuals underwent whole genome sequencing (WGS) analysis, either as part of the Genomics England 100,000 genomes project (PIDs #1-3) or the National Institute of Health Research Bioresource Rare disease study (PIDs #13-14) (Carss et al. 2017). Individuals #4-12 and #15-21 were sequenced as part of the retinal panel at the Molecular Vision Lab (http://www.molecularvisionlab.com), as described elsewhere (Fakin et al. 2016). For Dutch individuals 22 and 23, haloplex-based *ABCA4* targeted sequencing was performed (**Chapter 3**), while for individual 24, array-based *ABCA4* sequencing was done (Jaakson et al. 2003). Genotypic data are listed in **Table 1**. Written informed consent adhering to the Declaration of Helsinki was obtained from all individuals prior to participation in the study. This study was approved by the Institutional Review Board (IRB)/Ethics Committee approval (12/LO/0141) in the UK, and the Medical Ethical Committee 2010-359 (Protocol nr. 2009-32; NL nr. 34152.078.10) and the Commissie Mensgebonden Onderzoek Arnhem-Nijmegen (Dossier no. 2015-1543; dossier code sRP4h) in The Netherlands.

Clinical analysis

All individuals underwent complete clinical examination by an experienced ophthalmologist and were diagnosed with STGD1 or cone-rod dystrophy (CRD). Collected phenotype data included age at disease onset, visual acuity, fundus autofluorescence (FAF), optical coherence tomography (OCT), full-field electroretinography (ffERG) and pattern electroretinography (PERG). Disease onset was defined as the age when patients first noticed visual problems. Electroretinography (ERG) incorporated the recommendations of the International Society for Clinical Electrophysiology of Vision (McCulloch et al. 2015). PERG was used to study the function of the macula and ffERG to study the function of the peripheral retinal. Fundus autofluorescence and optical coherence tomography (OCT, Spectralis, Heidelberg, Germany) were used to determine the structural integrity of photoreceptors and RPE. Each of the persons exhibited at least one of the following clinical features of STGD1: characteristic yellow-white pisciform flecks in the retinal pigment epithelium of the posterior pole which were hyperautofluorescent on fundus autofluorescence imaging or progressive atrophy of the macular retinal pigment epithelium (**Chapter 3**). For each individual, clinical details are presented in **Table 2**.

Patients were classified into three groups according to the ffERG results, as described previously (Lois et al. 2001). Since there was a high similarity of ffERG

responses from both eyes, the right eye was chosen for the analysis of amplitudes in correlation with age and genotype. To determine the severity of the c.5196+1137G>A allele, a subgroup of patients harboring this variant *in trans* with an allele expected to produce no or minimal ABCA4 function was analyzed (n = 8, marked with § in **Table 1**). This enabled us to isolate the effect of the c.5196+1137G>A allele on the severity of the phenotype. Comparative groups consisted of 102 patients harboring different combinations of *ABCA4* alleles with previously determined severity (Fakin et al. 2016). The patient that was homozygous for the c.5196+1137G>A variant was compared to one carrying this allele *in trans* with a null allele.

Table 1. Gender and genotypic details of STGD1 cases with c.5196+1137G>A

PID	Sex	Allele 1		Allele 2		Allele 1 haplotype	Moorfields family number (GCID)
		cDNA	Protein	cDNA	Protein		
1	F	c.5196+1137G>A	p.[=,M1733Efs*78]	c.5714+5G>A	p.[=,E1863Lfs33]	A	21567
2	F	c.5196+1137G>A	p.[=,M1733Efs*78]	c.293A>G	p.(N98S)	B	4646
3	**M**	**c.5196+1137G>A**	**p.[=,M1733Efs*78]**	**c.4469G>A**	**p.(C1490Y)**	**C**	**4988**
4	**F**	**c.5196+1137G>A**	**p.[=,M1733Efs*78]**	**c.2160+2T>G**	**p.(?)**	**B**	**20160**
5	F	c.5196+1137G>A	p.[=,M1733Efs*78]	c.2063A>T	p.(N688I)	B	20122
6	**M**	**c.5196+1137G>A**	**p.[=,M1733Efs*78]**	**c.2564G>A**	**p.(W855*)**	/	**25804**
7	M	c.5196+1137G>A	p.[=,M1733Efs*78]	c.[4216C>T;6148G>C]	p.[(H1406Y);(V2050L)]	C	21670
8	F	c.5196+1137G>A	p.[=,M1733Efs*78]	c.4468T>C	p.(C1490R)	U	20567
9	F	c.5196+1137G>A	p.[=,M1733Efs*78]	c.4139C>T	p.(P1380L)	A	19004
10	**F**	**c.5196+1137G>A**	**p.[=,M1733Efs*78]**	**c.3364G>A**	**p.(G1122K)**	**B**	**20522**
11	M	c.5196+1137G>A	p.[=,M1733Efs*78]	c.4139C>T	p.(P1380L)	A	4198
12	**F**	**c.5196+1137G>A**	**p.[=,M1733Efs*78]**	**c.2239del**	**p.(L747Cfs*40)**	**A**	**4617**
13	M	c.5196+1137G>A	p.[=,M1733Efs*78]	c.4319T>C	p.(F1440S)	A	16063
14	F	c.5196+1137G>A	p.[=,M1733Efs*78]	c.1804C>T	p.(R602W)	A	20898
15	F	c.5196+1137G>A	p.[=,M1733Efs*78]	c.[1715G>A;2588G>C]	p.[(R572Q);G863A,G863del]	U	20546
16	**F**	**c.5196+1137G>A**	**p.[=,M1733Efs*78]**	**c.3329-1G>A**	**p.(?)**	**U**	**17478**
17	F	c.5196+1137G>A	p.[=,M1733Efs*78]	c.161G>A	p.[C54Sfs*14,C54Y]	A	20655
18	F	c.5196+1137G>A	p.[=,M1733Efs*78]	c.6079C>T	p.(L2027F)	U	21317
19	F	c.5196+1137G>A	p.[=,M1733Efs*78]	*unknown*	*unknown*	A	21659
20	M	c.5196+1137G>A	p.[=,M1733Efs*78]	c.[4126C>T; 6148G>C]	p.[(H1406Y);(V2050L)]	/	21257
21	F	c.5196+1137G>A	p.[=,M1733Efs*78]	c.5196+1137G>A	p.[=,M1733Efs*78]	/	n.a.
22a	F	c.5196+1137G>A	p.[=,M1733Efs*78]	c.859-506G>C	p.[F287Tfs*32,=]	A	n.a.
22b	M	c.5196+1137G>A	p.[=,M1733Efs*78]	c.859-506G>C	p.[F287Tfs*32,=]	(A)	n.a.
23	**F**	**c.5196+1137G>A**	**p.[=,M1733Efs*78]**	**c.[2918+775_3328+640del; 4462T>C]**	**p.[(S974Qfs*64);(C1488R)]**	**A**	**n.a.**
24	**M**	**c.5196+1137G>A**	**p.[=,M1733Efs*78]**	**c.3874C>T**	**p.(Q1292*)**	**A**	**n.a.**

PID: patient ID; F: female; M: male. Patients 22a and 22b are siblings. For individual 22b, haplotype A was assumed given its relation to individual 22a, hence the parentheses. U: haplotype unknown. /: haplotyping not performed. Individuals that carry c.5196+1137G>A *in trans* with a predicted null allele are depicted in bold.

Haplotype analysis

Where additional family member WGS data were available (individuals #1-3), phasing of informative variants to identify those *in cis* with c.5196+1137G>A was possible, thereby establishing the haplotypes on which this mutation was present in those individuals. Single nucleotide variants with minor allele frequency (MAF) <0.1 in the gnomAD dataset were identified in the *ABCA4* gene region. Variants present in the carrier parent, and absent in the other parent, were assumed to be *in cis* with the c.5196+1137G>A variant. Once these haplotypes and informative variants were established, direct variant interrogation in the singleton (unphased) WGS data (NIHR-RD, individuals #13-14) was performed to identify which (if any) haplotype was most plausible in these individuals. Direct Sanger sequencing of selected informative variants was performed in the remaining individuals. Primer sequences are provided in **Table S1**.

Selection of *ABCA4* intron 36 variants

To select additional DI variants in intron 36 of *ABCA4*, a literature search was conducted. All variants in intron 36 of *ABCA4* that were reported in affected individuals were selected based upon at least two of the three following selection criteria: i) variants have an allele frequency (AF) of ≤0.005 in general population databases such as in non-Finnish European (nFE) gnomAD (http://gnomad.broadinstitute.org/) or Genome of the Netherlands (GoNL; http://www.nlgenome.nl); ii) variants that cause an alteration in the splice acceptor site (SAS) or splice donor site (SDS) scores, or iii), a change in predicted splicing enhancer or splicing silencer elements when compared to the reference sequence. Reference and mutant sequences were analyzed using five algorithms (SpliceSiteFinder-like, MaxEntScan, NNSPLICE, GeneSplicer, Human Splicing Finder) (Shapiro and Senapathy 1987; Reese et al. 1997; Pertea et al. 2001; Yeo and Burge 2004; Desmet et al. 2009) via Alamut Visual software version 2.7. By applying these criteria, 12 variants in intron 36 of *ABCA4* were selected. Details of the selected variants and the *in silico* analyses are shown in **Table 3** and **Table S2.**

Table 2. Clinical characteristics of ABCA4-disease patients harboring c.5196+1137G>A

PID	Sex	Age at onset (yrs)	Age at last examination (yrs)	BCVA		Fishmann classifi-cation*	Foveal photo-receptors	Extent of FAF abn. with regards to vascular arcades	Age at ERG	ffERG group	PERG
				OD	OS						
1	F	55	60	6/9	6/6	2	Spared	Beyond	/	/	/
2	F	22	43	3/60	2/60	4	Atrophy	Beyond	41	3	A
3	M	22	39	3/60	3/60	4	Atrophy	Beyond	36	1	A
4	F	23	30	6/24	6/24	3	Atrophy	Beyond	/	/	/
5	F	41	64	6/18	CF	4	Atrophy	Beyond	61	3	A
6	M	12	12	/	/	0	Early changes	Within (minimal)	/	/	/
7	M	33	44	6/60	6/92	4	Atrophy	Beyond	42	1	A
8	F	10	26	6/36	6/36	2	Early changes	Within (minimal)	21	1	A
9	F	24	43	6/60	6/60	4	Atrophy	Beyond	36	3	A
10	F	15	21	6/36	6/36	2	Atrophy	Beyond	15	1	A
11	M	46	61	6/60	6/60	3	Atrophy	Beyond	/	/	/
12	F	22	62	HM	6/60	4	Atrophy	Beyond	52	3	A
13	M	46	56	6/18	6/12	4	Spared	Beyond	47	1	A
14	F	25	30	6/6	6/19	2	Atrophy	Beyond	26	1	N
15	F	39	51	6/9	6/9	2	Early changes	Beyond	40	1	A
16	F	24	50	2/60	6/6	3	Atrophy	Beyond	39	1	A
17	F	33	43	3/60	6/12	3	Atrophy	Beyond	39	3	A
18	F	55	62	6/6	6/36	3	Spared	Beyond	57	1	N
19	F	10	20	6/48	6/30	2	Atrophy	Beyond	17	1	A
20	M	40	72	3/60	3/60	4	Atrophy	Beyond	68	1	A
21	F	35	35	6/9	6/9	3	Spared	Beyond	35	1	N
22a	F	13	29	6/60	3/60	3	Atrophy	Beyond	25	1	A
22b	M	27	35	20/60	24/60	2	Atrophy	Beyond	27	1s	/
23	F	4	41	1/60	1/60	4	Atrophy	Beyond	34	1s	/
24	M	15	43	1/60	1/60	4	Atrophy	Beyond	18	1	/

PID: patient ID; F: female; M: male; BCVA: Best corrected visual acuity; OD: right eye; OS: left eye; PERG: pattern electroretinography; ffERG: full-field electroretinography; OCT: optical coherence tomography; 1s: normal photopic, low scotopic; / = not available; *Fishman classification: I – flecks limited to within the vascular arcades, II – fleck-like lesions anterior to the vascular arcades and/or nasal to the optic disc, III – most diffuse flecks resorbed leaving diffuse RPE atrophy, and IV – not only diffusely resorbed fundus flecks and atrophy of the RPE but also diffuse choriocapillaris atrophy. Individuals that carry c.5196+1137G>A *in trans* with a predicted null allele are depicted in bold.

Table 3. *ABCA4* intron 36 variants and allele frequencies in population databases

Intron 36 variant	cDNA variant	Genomic position (hg 19)	AF_gnomAD_nFE (accessed_18-10-2019)	AF_gnomAD (accessed_18-10-2019)	Reference
M1	c.5196+235G>A	94,484,903	-	-	Zernant et al, 2014
M2	c.5196+771G>A	94,484,367	-	-	Zernant et al, 2014
M3	c.5196+899C>T	94,484,239	0.03126	0.02097	Schulz et al, 2017
M4	c.5196+1013A>G	94,484,125	-	-	Schulz et al, 2017
M5	c.5196+1015A>G	94,484,123	0.05542	0.05228	Schulz et al, 2017
M6	c.5196+1056A>G	94,484,082	-	-	Zernant et al, 2014
M7	c.5196+1078delA	94,484,060	0.03418	0.05990	Schulz et al, 2017
M8	c.5196+1136C>A	94,484,002	0.0001296	0.01073	Bauwens et al, 2015
M9	c.5196+1137G>A	94,484,001	0.0001297	0.00009558	Braun et al, 2013
M10	c.5196+1159G>A	94,483,979	0.0003888	0.003694	Bauwens et al, 2015
M11	c.5196+1216C>A	94,483,922	-	-	Braun et al, 2014
M12	c.5196+1614G>A	94,483,524	-	-	Zernant et al, 2014

hg 19, human genome version 19; AF, allele frequency; nFE, non-Finnish European

Cell lines and culture conditions

Human Embryonic Kidney 293 (HEK293T) cells were purchased from ECACC (Catalogue no. 12022001; Salisbury, UK), genomic integrity was tested by STR PCR. HEK293T cells were cultured and passaged twice a week as described previously (**Chapter 3**).

Midigene-based *in vitro* splicing assay

The individual effect of the c.5196+1137G>A and 11 additional 12 selected DI variants was assessed by midigene-based splicing assays employing a newly designed wild type construct (coined BA32) harboring a region of *ABCA4* spanning from intron 34 to intron 38. The reference sequence was derived from the bacterial artificial chromosome clone CH17-325O16 (insert g.94,434,639-94,670,492), as described previously (**Chapter 2**). Details of primers used for Gateway cloning, PCR and/or sequencing are provided in **Table S1**. The BA32 construct was used as a template to generate mutant midigene constructs by site-directed mutagenesis, which were subsequently validated by Sanger sequencing. Details of mutagenesis and sequencing primers are given in **Table S1**. Wild type and mutant BA32 constructs were transfected in HEK293T cells and the extracted total RNA was subjected to reverse transcription (RT)-PCR, as previously described (**Chapter 2**). Rhodopsin (exon 5) amplification was used as transfection and loading control. All experiments were performed in two independent replicates. Primer sequences used for RT-PCR analysis are presented in **Table S1**.

Antisense Oligonucleotides

AON molecules were designed as described previously (Aartsma-Rus 2012; Garanto and Collin 2018). Briefly, the PE region including 50 bp up- and downstream were subjected to RNA structure and splicing enhancer factors analyses employing several freely available software programs. AONs were ordered from Eurogentec with a 2'*O*-methyl sugar modification and a phosphorothioate backbone (commonly called 2OMe/PS AONs). A sense oligonucleotide (SON) with the same chemical modifications

was used as a negative control. All AON sequences and characteristics are provided in **Table S3**.

***In vitro* rescue studies in HEK293T cells using midigenes and AONs**

Wild-type (WT) and mutant midigene constructs were transfected in HEK293T cells as described above. Subsequently, cells were transfected with AONs targeting the PE of which the insertion was induced by the different mutations, respectively. AONs were delivered as described previously (**Chapter 3**). Forty-eight hours post AON treatment, cells were harvested and total RNA was extracted, which was then subjected to RT-PCR. Rhodopsin amplification was used as transfection and loading control. All experiments were performed in two independent replicates. Primer sequences can be found in the **Table S1**.

Rescue studies in patient-derived photoreceptor precursor cells (PPCs)

Patient-derived fibroblast cells were reprogrammed to induced pluripotent stem cells (iPSCs) using the Yamanaka factors (Takahashi and Yamanaka 2006), together with a control line derived from a healthy individual. The pluripotency of these cells was validated by qPCR and immunocytochemistry for pluripotency markers, as described previously (Albert et al. 2018). Subsequently, cells were subjected to a two-dimensional differentiation protocol, to allow the formation of retinal precursor cells, as described elsewhere (Flamier et al. 2016). For this, iPSCs were dissociated and seeded in matrigel-coated 12 well plates, and differentiated for 30 days. On day 28, AONs were added to the cells at a final concentration of 1 µM. The day after, cells were subjected to cycloheximide (CHX, final concentration 100 µg/ml) treatment to block nonsense mediated decay. On day 30, cells were harvested and subjected to RNA analysis. RNA was isolated using the Nucleospin RNA kit (Macherey-Nagel, Düren, Germany) following the manufacturer's instructions. One microgram of RNA was converted into cDNA using SuperScript™ VILO™ Master Mix (Invitrogen, Carlsbad, USA). Fifty nanograms of cDNA were used for PCR analysis, as described in previous sections. Primers used for validation of the differentiation into PPCs are provided in **Table S1**.

Quantification of RT-PCR products

To assess the quantity of the aberrantly and correctly spliced RT-PCR products, densitometric analysis of the gel electrophoresis images was performed with Image J software (Schindelin et al. 2012), after which the ratio of aberrantly spliced products was calculated.

Rescue studies in patient-derived retinal organoids

Patient fibroblasts were reprogrammed as previously described (Okita 2011). iPSCs were maintained in Essential 8 Flex media (ThermoFisherScientific), and retinal organoids were produced as described previously, with modifications (Nakano et al. 2012; Parfitt et al. 2016b). Briefly, iPSCs were dissociated into a single cell suspension and resuspended at 9000 cells per well of V-bottomed 96-well plates, in E8Flex media supplemented with 20μM Y-27632 (Millipore) from day 0 and from day 2 EB2 media supplemented with 20μM Y-27632, 3mM IWR1e (Calbiochem) and 1% matrigel. Media was changed every two days. On day six only, media was additionally supplemented with 55 ng/ml BMP4 (Preprotech). From day 12 onwards, EB2 media were supplemented with 10% FBS and 100 nM SAG (Tocris). On day 20, embryoid bodies were transferred to U-bottomed 96-well plate and media was changed to NR media, replenished every two days (DMEM/F12 with 10% FBS, 1x N2 supplement (ThermoFisherScientific) and 0.5 μM RA). Around this timepoint, retinal organoid formation could be visualized. On day 40, embryoid bodies with clearly defined neuroepithelial structure suggestive of retinal organoids (~30%) were transferred to 25 well square petri dishes and media were changed twice a week. At day 100, RA was withdrawn to encourage formation of outer segments. At D180 (when 'brush border' was visible at apical edge) retinal organoids were treated gymnotically with 1 μM, 5 μM or 10 μM AON6, or 10 μM SON for 72 hours, before processing for analysis.

For RNA analysis, organoids were washed gently with PBS twice before RNA extraction via RNeasy Plus Micro kit, as per manufacturer's instructions (Qiagen). cDNA was transcribed using oligodT primer via Tetro cDNA synthesis kit (Bioline). PCR was performed using GoTaq Green with standard cycling conditions (Promega). Primers are listed in **Table S1**. Gel band densitometry analysis was performed in ImageJ by measuring average pixel density. For immunofluorescence, organoids were washed gently with PBS once before fixation in 4% PFA/5% sucrose for 45 minutes at 4°C. Organoids were cryoprotected using increasing concentrations of sucrose (6.25% and 12.5%) for 30 minutes each at 4°C before overnight cryoprotection in 25% sucrose at 4°C. Organoids were frozen in OCT matrix using cooled acetone before cryosectioning at 10 μm. Sections were processed for immunofluorescence as previously described (Parfitt et al.). Primary antibodies were anti-mouse rhodopsin 4D2 (1:1000; Merck Millipore), anti-mouse cone arrestin clone 7G6 (1:100; gift from Peter MacLeish), anti-rabbit L/M opsin (1:500; Merck Millipore) and anti-rabbit recoverin (1:500; Merck Millipore). Secondary antibodies were donkey anti-mouse Alexa Fluor 488 and donkey anti-rabbit Alexa Fluor 555 (1:1000; ThermoFisherScientific).

Statistical analysis

Statistical analysis was performed using SPSS software v.22 (IBM SPSS Statistics, IBM Corporation, Chicago, IL, USA). The Mann-Whitney U test was used to test for significant differences in the median age of onset between patients with different genotypes. The 95% confidence interval of the double null regression line was used to determine whether the ERG amplitudes of patients harboring c.5196+1137G>A differed significantly from those harboring two null alleles.

Results

Variant c.5196+1137G>A is an allele of intermediate severity

Phenotypic data of 25 patients harboring the c.5196+1137G>A mutation are shown in **Table 2**. The median age at onset was 24 years (range: 4-55 years), and the last exam was performed at the median age of 43 years (range: 12-72 years). The majority (92%; 22/24) exhibited FAF abnormalities (hyper- or hypoautofluorescence) within and beyond the vascular arcades, while in 8% (2/24) the abnormalities confined were confined within the vascular arcades (**Figure 1**). Foveal photoreceptors were affected in the majority (21/24) of patients and were preserved (foveal sparing) in three. The median age at ERG recording was 36 years (range: 15-68 years; N = 21). At that time, according to the previously established ERG classification) (Lois et al. 2001), 67% (14/21) of the patients were classified into group 1 (normal full field ERG; ages 15-68 years), and 24% (5/21) into group 3 (abnormal cone and rod function; ages 36-61 years). Two patients had normal photopic responses while their scotopic responses were subnormal, which is a pattern not included in the previous classification (**Table 2**). For a number of patients, cross-sectional ERG results were plotted against their age (**Figure 2**).

To determine the severity of the c.5196+1137G>A mutation, a subgroup of eight patients harboring this allele *in trans* with a null or null-like allele (Table 1) was analyzed separately. This allowed the assessment of the c.5196+1137G>A allele in isolation, an approach that has been used previously to determine the severity of *ABCA4* alleles (Fakin et al. 2016). The distribution of ages at onset for the different genotypes is shown in **Figure S1**. The median age at onset of the patients c.5196+1137G>A patients was significantly higher than that of the patients harboring two null alleles (19 *vs*. 6 yrs, Mann-Whitney U test, $p < 0.01$). It was most comparable to previously defined alleles of intermediate severity. e.g., p.(R24H), c.5714+5G>A and p.(L2027F).

The FAF images of the eight selected patients were also arranged by age and compared with age-matched patients harboring two null alleles. In the early stages, the

macular area was notably more preserved in c.5196+1137G>A patients. In later stages, when the atrophy expanded outside the imaged area, the differences were less obvious. None of these patients exhibited foveal sparing (**Figure 1** and **Figure S2**). The phenotype of the patient that was homozygous for c.5196+1137G>A was compared to the patients above to determine whether there was any evidence for a reduced effect of this genotype on the clinical appearance. In comparison to the patients harboring c.5196+1137G>A *in trans* with a loss-of-function allele, the c.5196+1137G>A homozygous patient had significantly delayed disease onset (35 yrs *vs*. 19 yrs (range: 4-24 years). Although FAF imaging showed widespread abnormalities at the posterior pole and beyond the vascular arcades, the OCT revealed mainly preserved photoreceptors across the whole macula, including those in the fovea (**Figure S3**). The full-field ERG and pattern ERGs in this patient were normal. Finally, the c.5196+1137G>A homozygous patient also had well preserved visual acuity (6/9 on both eyes at age 35, **Table 2**) while none of the patients harboring V1 *in trans* with a loss-of-function allele had visual acuity above 6/24 (ages 12-62).

Variant c.5196+1137G>A occurs on different haplotypes

Haplotype analysis was performed using variants located within the *ABCA4* locus and is summarized in **Table 1**. For some of the cases carrying the c.5196+1137G>A mutation, WGS data of probands, as well as unaffected parents were available, enabling phasing of variants in the genomic region harboring *ABCA4*. In total, three probands affected by *ABCA4*-retinopathy in our cohort of cases recruited to the Genomics England 100,000 genomes project included this allele.

For individual #2, c.5196+1137G>A was identified on the paternal allele *in trans* with the reported missense variant (p.N98S). Filtering of *ABCA4* variants with an allele frequency of <0.1 in the gnomAD dataset and an in-house cohort of 7766 alleles from the WGS cohort revealed four variants informative for the haplotype (defined as haplotype B, **Table 1** and **Table S4**).

For individual #3, c.5196+1137G>A was identified on the paternal allele *in trans* with the previously reported missense variant (p.C1490Y). Filtering of *ABCA4* variants applying the same criteria revealed a different haplotype (haplotype termed C) which contained the variant in intron 36 (**Table 1** and **Table S4**).

Individual #1 was recruited for WGS with an affected sibling, it was not possible to accurately phase variants located in or near *ABCA4* in the absence of parental samples, yet we were able to identify the presence of the upstream variant chr1:94531618T>C shared with individual 2 (haplotype B) and found in only 2 out of 15,430 European alleles in gnomAD in the absence of the downstream variants found on haplotype B. This may suggest that the haplotype observed in individuals #1 and #2

have a common ancestral gene that underwent recombination in the *ABCA4* gene region, giving rise to the additional haplotype observed in subject #1 (defined here as haplotype A, **Table 1** and **Table S4**). However, it is also possible that the chr1:94531618T>C variant was present on the opposite allele, in which case the haplotype would be unknown. Besides individuals #1-3, subjects #13 and #14 also underwent WGS as singletons. In both cases, the presence of the four rare variants previously found to define haplotype B were identified. Subsequently, these and other rare variants close to c.5196+1137G>A that marked the different haplotypes (including chr1:94531618T>C to determine haplotype A) were genotyped in all available remaining cases via Sanger sequencing (**Table S4**). An additional 5/17 cases were suggestive of haplotype A, 3/17 were consistent with haplotype B and 1/17 being consistent with haplotype C. Of the remaining four cases available for testing, three did not carry any of the haplotype associated variants, whereas one individual carried all variant alleles tested (**Table 1** and **Table S4**). This may be explained by the relatively high frequency of chr1:94513853TAA>T in gnomAD (MAF 0.07479) and therefore it may also be present on the *trans* allele.

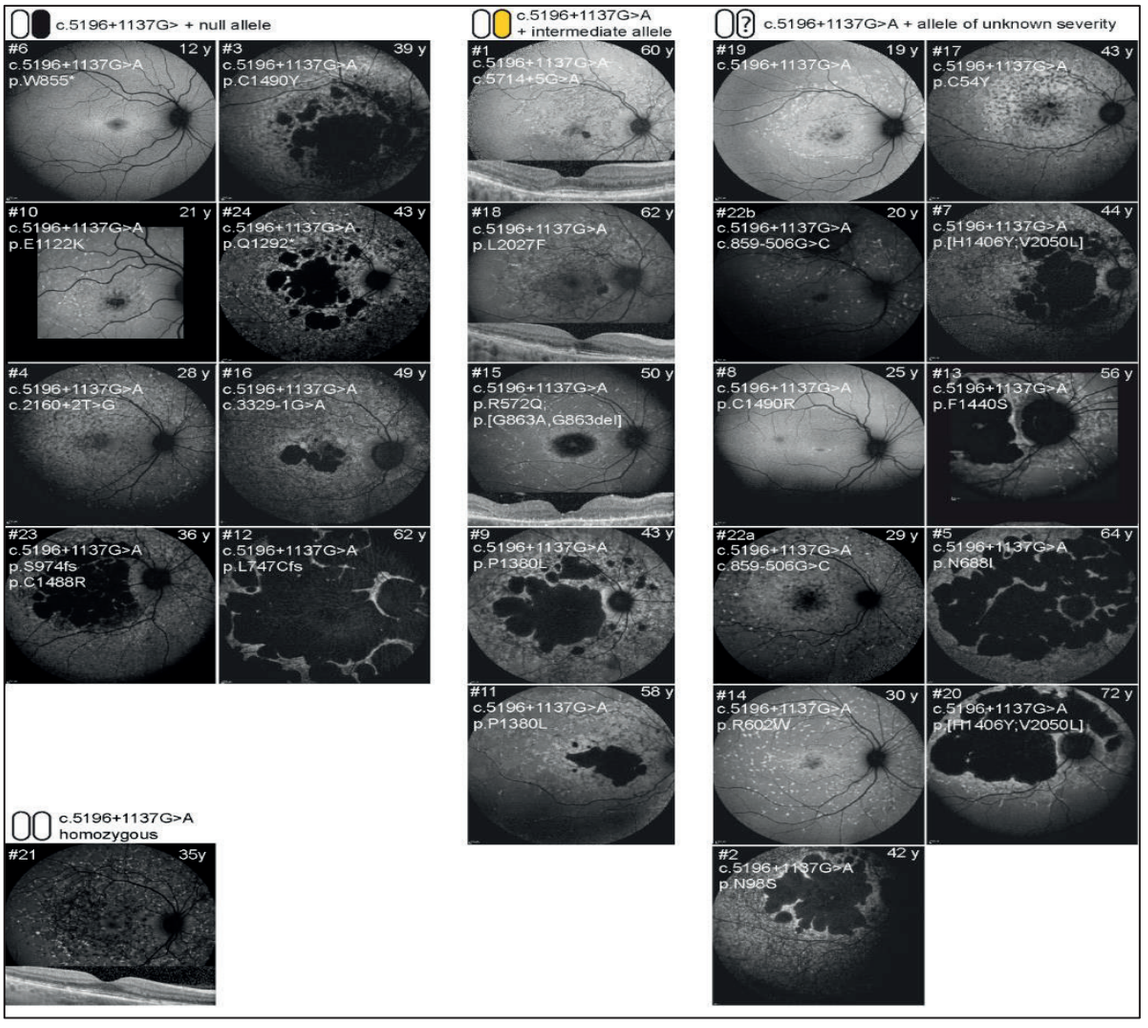


Figure 1. Fundus autofluorescence images of *ABCA4*-disease cases harboring c.5196+1137G>A. FAF images of patients harboring c.5196+1137G>A *in trans* with different groups of alleles. Classification of *ABCA4* alleles was described previously (Fakin et al. 2016). Note the increasing retinal degeneration (no flecks, fleck outside the arcades, absorbed flecks, chorioretinal-atrophy) in patients harboring c.5196+1137G>A *in trans* with null, when arranged by age (first two columns). Note the relatively milder phenotypes in patients harboring c.5196+1137G>A *in trans* with intermediate alleles c.5714+5G>A, p.L2027F and p.[G863A,G863del] and in homozygous state.

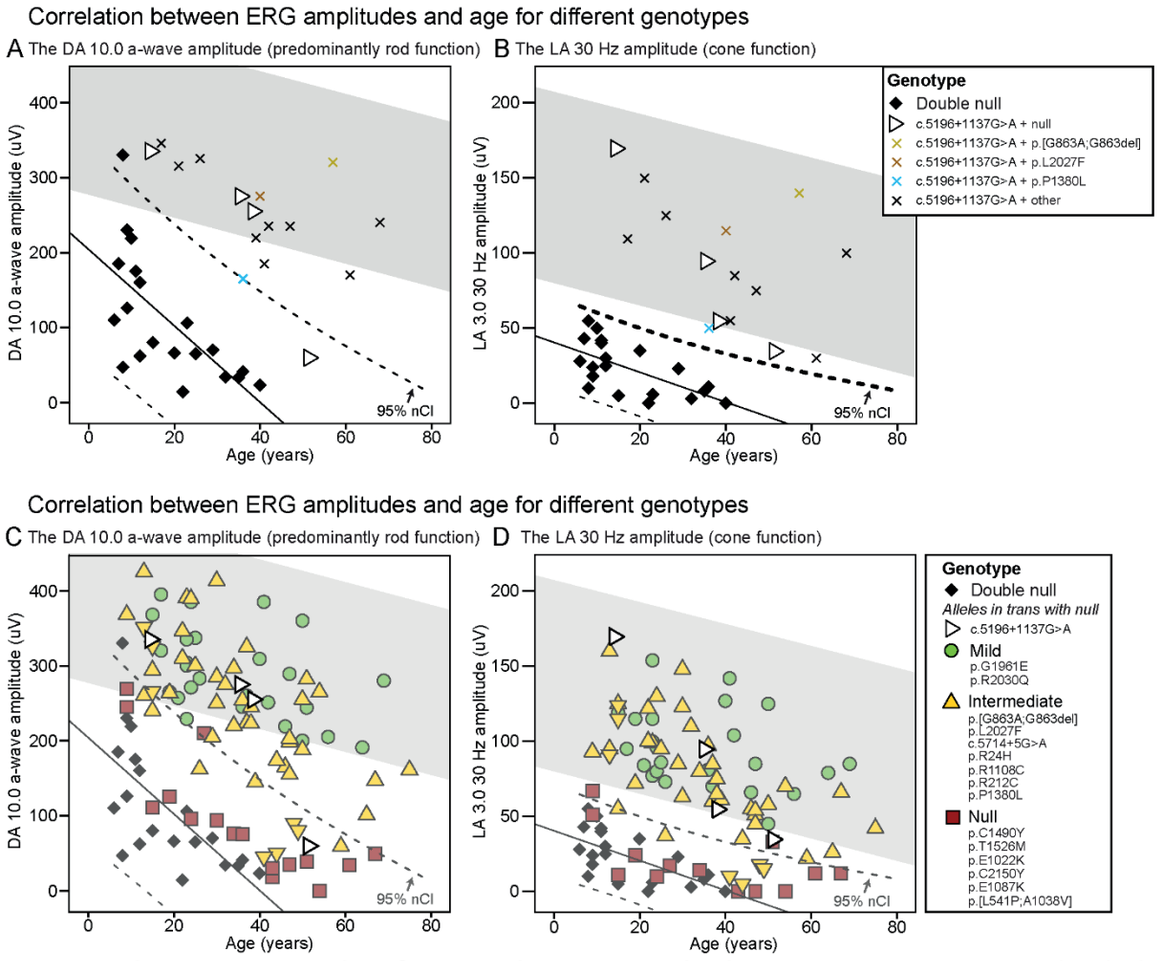


Figure 2. Electroretinography of ABCA4-disease cases harboring c.5196+1137G>A. The dark-adapted (DA) 10.0 a-wave amplitudes **A.** and the light-adapted (LA) 3.0 30-Hz amplitudes. **B.** of 'double null' patients and patients harboring c.5196+1137G>A with different mutations *in trans*, plotted by age. **C and D.** The DA 10.0 a-wave amplitudes **C** and the LA 3.0 30-Hz amplitudes **D** of patients harboring c.5196+1137G>A in *trans* with null alleles and patients harboring 15 other alleles of previously determined severity (Fakin et al. 2016), plotted by age. The grey area represents the 95% confidence interval of the healthy volunteers. On each chart, the 'double null' patients are shown as a baseline reference with their regression line and their confidence intervals (CI). Patients harboring c.5196+1137G>A had significantly better retinal function compared to 'double null' patients.

Identification of *ABCA4* variants in intron 36

Following the clinical characterization of patients harboring the c.5196+1137G>A variant, an in-depth molecular analysis of this, and other variants in close proximity (i.e., intron 36 of the *ABCA4* gene) was performed. Previously, a few variants in intron 36 potentially causing ABCA4-disease have been described, including an *in silico* analysis of the splicing defects they exert (Braun et al. 2013). In addition, other reports have described the clustering of DI variants in certain introns, suggesting that some introns are more prone to harbor disease-causing mutations than others (Albert et al.

2018; **Chapter 3**). To investigate which *ABCA4* variants in intron 36 could potentially lead to splicing defects (and can be considered pathogenic), all previously described variants were collected and analyzed for frequency statistics and splicing predictions. This led to the identification of eleven additional variants, besides c,5196+1137G>A, that could potentially influence splicing, e.g., by the generation of new splice acceptor or donor sites, or changing splicing enhancer or silencer motifs (**Table 3** and **Table S2**). Although some of the variants had a relatively high AF, these were still tested, as previous studies demonstrated a pathogenic role of frequent DI variants (c.769-784C>T, c.4253+43G>A) (**Chapter 3**).

To study the causative nature of all these variants, a new construct was generated that could be used in a midigene splicing assay (**Chapter 2**). This construct, coined BA32, harbors the region of *ABCA4* spanning from intron 34 to intron 38, thus allowing to study the splicing of intron 36 in a genomic context more extended than a minigene. Following confirmation that transfection of this midigene indeed leads to the expected splicing pattern (merge of exons 35, 36, 37 and 38), site-directed mutagenesis was performed to generate mutant BA32 constructs each harboring one of the 12 selected DI variants. Transfection of the WT and the various mutant BA32 constructs into HEK293T cells followed by RT-PCR analysis revealed that four out of the twelve mutant constructs led to aberrant splicing, i.e., the insertion of additional sequences in between exons 36 and 37 (**Figure 3**). For two variants (c.5196+1013A>G (M4) and c.5196+1056A>G (M6)), the predominant transcript contained a PE (of 129 or 177 nt, respectively) that overlapped, defined by an already existing cryptic splice acceptor site in the reference sequence, and the strengthening of different cryptic splice donor sites by either of the variants. Two other variants, the recurrent c.5196+1137G>A (M9) and c.5196+1216C>T (M11), led to the insertion of a different, 73-nt PE, but as a smaller fraction of transcripts. Both variants led to the insertion of the exact same PE, although the c.5196+1137G>A variant slightly strengthens a cryptic splice acceptor site, whereas c.5196+1216C>A clearly strengthens a cryptic splice donor site of the PE (**Figure 3A** and **B**, **Figure S4**). Seven out of the other eight variants did not alter the splicing pattern (**Figure S5**), suggesting that these variants do not have a drastic impact on the processing of *ABCA4* pre-mRNA, at least not in HEK293T cells employed here. Variant c.5196+1078delA was not assessed since site-directed mutagenesis to introduce this change into the BA32 construct was repeatedly unsuccessful.

Aberrant splicing induced by c.5196+1137G>A variant in photoreceptor precursor cells

As can be observed in **Figure 3B**, the splicing defect evoked by the c.5196+1137G>A variant was only present in a small proportion of transcripts following transfection of the mutant midigene into HEK293T cells. Previous reports demonstrated that for some DI variants, the degree of PE inclusion can be tissue- or even cell type-dependent, e.g., the splicing defect is more prominent in retina-like tissue (Parfitt et al. 2016b; Albert et al. 2018). Given the association of distinct c.5196+1137G>A haplotypes with patients with ABCA4-disease, we explored whether this was also true for this variant. Fibroblast cells of affected individual #23 harboring this allele *in trans* with a protein-truncating mutation (p.[S974Qfs*64;C1488R]), together with that of a healthy control individual, were reprogrammed into iPSCs as described previously (Sangermano et al. 2016). Subsequently, iPSCs were differentiated into photoreceptor precursor cells (PPCs) following a thirty-day protocol (Flamier et al. 2016), giving rise to an increased expression of several retina-specific genes including *ABCA4*. The degree of differentiation was assessed via qPCR analysis (**Figure S6**). Prior to harvesting, cells were treated with cycloheximide (CHX), an agent that blocks nonsense-mediated decay (NMD) of aberrant transcripts, as both alleles are predicted to give rise to premature termination codons that can lead to this phenomenon. As shown in **Figure 3C**, especially following CHX treatment, the degree of PE insertion caused by c.5196+1137G>A was notably higher compared to that in the midigene splicing assay, in particular when taking into account that only half of the *ABCA4* transcripts in these PPCs are derived from the allele harboring this mutation. Of note, a small amount of transcripts showed PE inclusion in the control PPCs, following CHX treatment, demonstrating that in the absence of the DI variant, this PE can be recognized by the splicing machinery in retina-like cells.

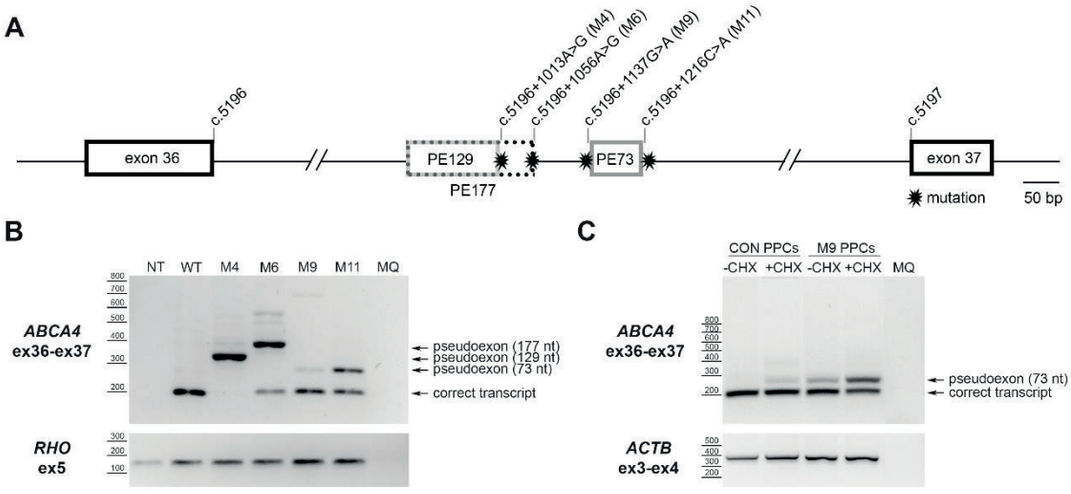


Figure 3. Pseudoexon insertion caused by deep-intronic variants in intron 36. A. Schematic representation of the intron 36 with the location of all the mutations and the pseudoexon (PE) insertion caused by each variant (M). The number of the different PEs indicates the size in nucleotides. **B.** Identification of PEs using a midigene-based approach in HEK293T cells. *RHO* amplification was used as transfection and loading control. **C**. Identification of a 73-nt PE in patient-derived photoreceptor precursor cells (PPCs). This PE is subjected to nonsense-mediated decay as its detection is clearly increased upon cycloheximide (CHX) treatment. *ACTB* was used as a loading control. MQ is the negative control of the PCR.

Antisense oligonucleotides rescue splicing defects due to intron 36 variants in *ABCA4*

Following the identification of the PE insertions caused by the four variants in intron 36, antisense oligonucleotides were designed to block splicing of these PEs into the *ABCA4* mRNA, with the aim to restore normal splicing. For the partially overlapping PEs (PE129 and PE177) induced by c.5196+1013A>G and c.5196+1056A>G, respectively, four AONs were designed. AON1 and 2 target a region present in both PEs, whereas AON3 and 4 overlap with PE177 but not, or at least not entirely, with PE129 (**Figure 4A**). Of note, AON3 is targeting a region in PE177 that encompasses the reference nucleotide at position c.5196+1013, and therefore contains a mismatch towards the c.5196+1013A>G change. As shown in **Figure 4B**, co-transfection of the corresponding midigene with AON1 or 2 resulted in a full correction of the aberrant splicing events induced by c.5196+1013A>G, AON3 showed only a partial correction and AON4 had no effect. In contrast, for the c.5196+1056A>G variant, all four AONs showed restoration of correct splicing, AONs 1 and -3 being the most potent (**Figure 4C**). For the 73-nt PE inclusion arising from either c.5196+1137G>A or c.5196+1216C>A, four AONs were designed, none of which were allele-specific. To assess the efficacy of these AONs to abolish PE inclusion by the c.5196+1216C>A variant, co-transfection of the midigene with the AONs resulted in a complete correction of splicing for AONs -5 and -6, whereas AON7 and -8 showed a partial effect (**Figure 4D**).

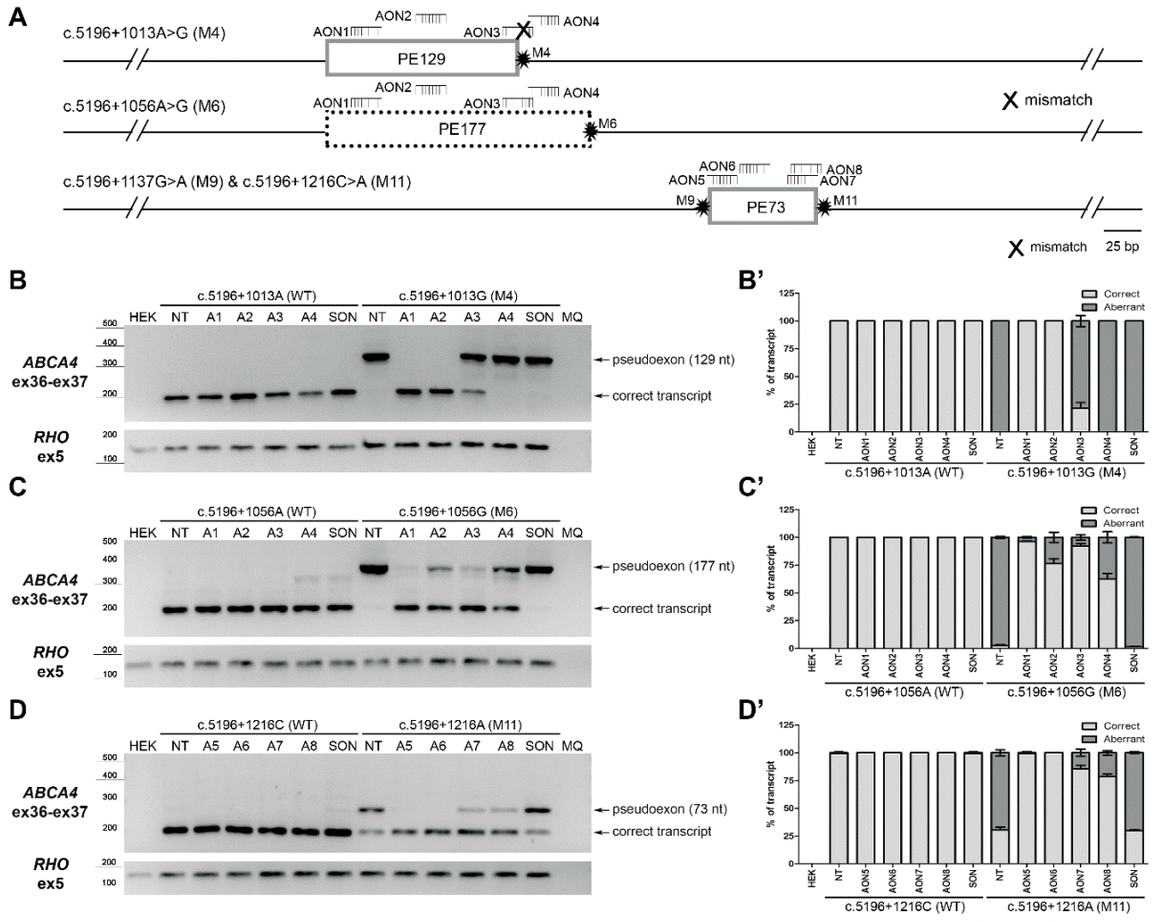


Figure 4. Assessment of splicing redirection by antisense oligonucleotides (AON) using a midigene-based system in HEK293T cells. A. Schematic representation of the pseudoexons (PEs) introduced by each variant (M) and the relative position of the different AONs within the region. **B** and **B'**. Assessment by RT-PCR of AON1 to AON4 (A1 to A4) in HEK293T transfected with either the wild-type (WT) midigene or the midigene containing the c.5196+1013A>G (M4) variant (B) and its corresponding semi-quantification expressed in ratio of correct *vs*. aberrant transcript (B'). **C and C'**. Splicing redirection by AON1 to AON4 for the c.5196+1056A>G (M6) variant in HEK293T cells (C) accompanied by its semi-quantification expressed in ratio of correct *vs*. aberrant transcript (C'). **D** and **D'**. Effect of AON5 to AON8 (A5 to A8) to redirect the splicing defect introduced by c.5196+1216C>A (M11) in HEK293T cells (D) as well as its semi-quantification expressed in correct *vs*. aberrant transcript ratio. In all cases, a scrambled oligonucleotide (SON) was used as negative control. MQ is the negative control of the PCR. In all graphs results are presented as average

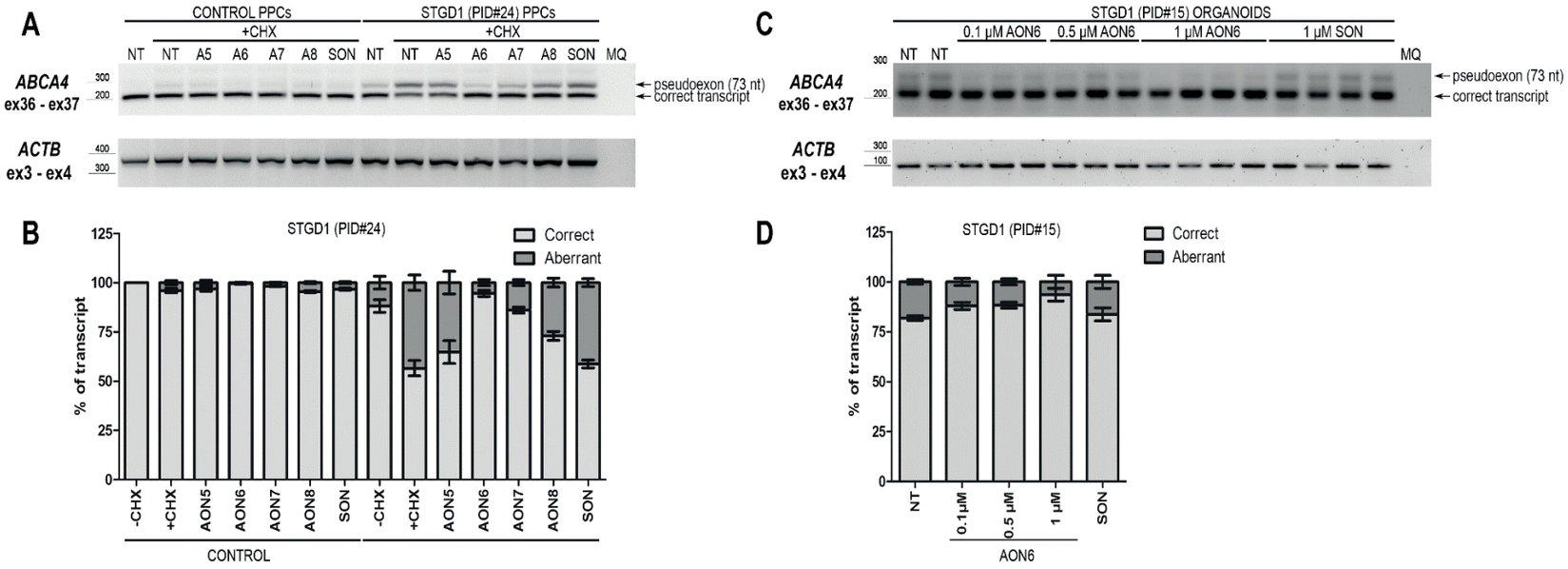


Figure 5. AON-based splicing correction for the c.5196+1137G>A (M9) variant using patient derived retina-like cells. A. RT-PCR analysis of control and STGD1 individual PPCs treated with AON5 to AON8 (A5 to A8) or scrambled oligonucleotide (SON). Pseudoexon detection was enhanced by cycloheximide (+CHX) treatment. *ACTB* amplification was used as loading control. **B.** Representation of the ratio of correct and aberrant transcript upon semi-quantification. **C**. RT-PCR analysis of STGD1 individual-derived 185 day old retinal organoids treated with AON6 at the indicated concentrations or scrambled oligonucleotide (SON). *ACTB* amplification was used as loading control (of note, primers were different to those used in B). **D**. Graphical representation of the ratio of correct and aberrant transcript upon semi-quantification.

To study splicing defects caused by c.5196+1137G>A, iPSC-derived PPCs were again employed, as these show a more robust insertion of PE73 compared to the HEK293T cells transfected with the corresponding mutant midigene. As shown in **Figure 5A** and -**B**, AON6 and -7 were most potent in converting PE73-containing transcripts to the correct transcripts, whilst AON5 and -8 had little effect. Of note, the PE-containing transcripts again were more visible following CHX-treatment. Finally, to study PE73 inclusion upon a longer differentiation period, iPSCs were differentiated for 180 days, i.e., to form 3D retinal organoids. Previously, we have shown that antisense oligonucleotides can be used to restore splicing of aberrantly spliced *CEP290* in retinal organoids (Parfitt et al. 2016), and used them as part of the preclinical development of an AON-based therapy (Dulla et al. 2018; Cideciyan et al.2019). Retinal organoids were generated from iPSCs derived from individual #15 (**Table 1**) that carries c.5196+1137G>A *in trans* with a complex allele (p.[R572Q;G863A,G863del;N1868I]), as previously described (Nakano et al.2012; Parfitt et al.2016). Retinal organoids developed as shown previously, with maturation of a neuroblastic cell layer into a defined outer nuclear layer-like structure, with expression of photoreceptor markers evident by day 120 (data not shown). As retinal organoids matured further to day 180, a 'brush border' of outer segment-like structures appeared at the apical edge of the organoids, which were positive for rhodopsin expression (**Figure S7A**). Organoids were also positive for photoreceptor markers such as recoverin and cone arrestin, and L/M opsins, suggesting the correct formation of retina-like structure (**Figure S7B**). At this stage, the retinal organoids were treated gymnotically with AON6 or SON days for 72 hours, in the absence of any NMD inhibition, to mimic the conditions that might be applied *in vivo*. Treatment with AON6 led to a dose-dependent decrease in PE-containing transcripts (**Figure 5C** and **D**). In all experiments, transfection of a negative control sense oligonucleotide (SON) had no effect on splicing, showing the exact same results as the untreated condition. In addition, all AONs did not seem to have an effect when added to control cells or to those transfected with a WT midigene.

Discussion

In this study, we assessed variants in intron 36 of *ABCA4*, and their potential involvement with ABCA4 retinopathy. Detailed clinical characterization and genomic analysis of the most frequent variant, c.5196+1137G>A, showed that this variant is likely to have arisen on different haplotype backgrounds and acts as an allele of intermediate severity. This is in line with the molecular defects that were found in PPCs and retinal organoids harboring this mutation, which showed a residual amount of a

normally spliced product. For eleven other *ABCA4* variants in intron 36, three were also found to result in clear splicing defects, all by inserting PEs into the final *ABCA4* transcript. Furthermore, by using cultured HEK293T cells or patient-derived PPCs and 3D retinal organoids, all splicing defects could be rescued by the administration of AONs targeting these PE inclusions.

The c.5196+1137G>A variant was one of the first DI mutations in *ABCA4* described (Braun et al. 2013). In that study, it was shown that this mutation could lead to splicing that results in inclusion of a PE in a small proportion of *ABCA4* transcripts in patient-derived keratinocytes. These data are similar to the observations made here, where the PE insertion was at a low to level in our midigene splicing assay that was performed in HEK293T cells. Only when differentiating patient-derived cells to a more retinal lineage, either as PPCs or retinal organoids, did the proportion of PE-containing *ACBA4* transcripts increase, as has been described for other splicing defects due to DI changes (Parfitt et al. 2016; Albert et al. 2018). Despite a relatively high frequency of this allele in patients, there has remained some controversy about the pathogenic nature of the c.5196+1137G>A variant. Based on the allele frequency of this variant in African American control subjects in a single study, but mainly on the notion that in macaques, c.5196+1137A (the mutant allele in humans) was found to be the most frequent major allele at this position. In macaque retina harboring this A allele, no PE-insertion was detected, thus the pathogenicity of this variant was questioned (Zernant et al. 2014). However, when directly comparing the sequence of the entire region encompassing the PE in humans, several additional nucleotide differences were identified between the human and macaque genomic sequence either in the 73-nt PE or the cryptic splice donor site (**Figure S8**). Whereas Human Splicing Finder software HSF attributes equal scores to the human mutant and macaque splice acceptor and donor site, another splicing prediction software, MaxEntScan predicts a significantly lower strength of the splice donor site in macaque compared to human. In addition, a number of the differences between human and macaque within the PE are predicted to generate splice silencer sequences in macaque but not in human. Since pre-mRNA splicing is well-known to be a complex process involving not only the splice acceptor and donor site sequences, but also binding sites for exonic splicing enhancers and silencers (Lee and Rio 2015), this divergence may explain why c.5196+1137G>A induces PE insertion in humans, but not in macaques.

Patients carrying the c.5196+1137G>A allele *in trans* with known disease alleles had clinical presentations within the spectrum of *ABCA4*-retinopathy, supporting the pathogenicity of the allele. In addition, using WGS no more plausible potentially pathogenic variant was identified in the patients that could exclude pathogenicity of c.5196+1137G>A. Although the majority of cases exhibited FAF abnormalities beyond

the vascular arcades and fovea was rarely spared, the retina-wide photoreceptor dysfunction was relatively mild, as demonstrated by preserved peripheral cone and rod function on the full-field ERG in the majority of subjects. Categorization of the c.5196+1137G>A allele was performed by selecting a subgroup of patients harboring this allele *in trans* with a (predicted) null allele, and comparing their ages at onset, ERG amplitudes and FAF patterns with patients harboring two null alleles or one of the 15 *ABCA4* alleles of known severity (Fakin et al. 2016) in trans with null alleles. The c.5196+1137G>A was similar to alleles of intermediate severity. These had been defined previously as alleles that *in trans* with null produce a significantly milder phenotype than two null alleles, however, often progress to cone-rod dystrophy at older ages. The moderately severe nature of this variant is in line with other findings, i.e., none of the cases harboring c.5196+1137G>A carried a mild allele *in trans* (e.g., p.G1961E and p.R2030Q, which *in trans* with null produce normal full-field ERG even at older ages,) (Fakin et al. 2016), as it is likely that such a combination would result in ABCA4 functional protein reduction surpassing the disease threshold. In addition, a recent study on DI *ABCA4* variants in a French cohort also concluded that the c.5196+1137G>A variant was mostly associated with a milder phenotype, based on six subjects harboring this allele (Nassisi et al. 2019). It should also be noted however that evidence is emerging that some variants that were long considered pathogenic or benign by themselves, such as c.2588G>C;p.[G863A,G863del] and c.5603A>T;p.(N1868I), only in combination on the same allele are fully penetrant. The same is true for the *cis* configuration of p.(N1868I) with a recently identified intronic mutation, c.769-784C>T (Zernant et al. 2017; Runhart et al. 2018; Runhart et al. 2019). In addition, modifier variants, either within or outside *ABCA4*, as well as environmental factors will likely play a role in determining disease outcome (Runhart et al. 2018). Thus, care is warranted when attributing a severity score to individual alleles. The existence of residual ABCA4 function produced by the c.5196+1137G>A variant is also supported by the clinical presentation of the homozygous patient, namely a significantly delayed disease onset and better preserved photoreceptors in comparison to the patients with only a single c.5196+1137G>A allele in combination with a null allele. Finally, the molecular studies revealed that the PE insertion induced by this variant was only present in a proportion of the *ABCA4* transcripts, thus always giving rise to residual WT transcript and translation of a functional protein.

The fact that the c.5196+1137G>A variant was identified on different haplotypes suggests that it may have arisen independently, which would further support a case for pathogenicity. However, alternatively, one could argue that the C haplotype is ancestral and that haplotype A occurred by a centromeric recombination, followed by a telomeric recombination giving rise to haplotype B. Of note, no clear correlation

between the severity of the clinical phenotype and the nature of the haplotypes (A, B or C) was observed. This suggests that the surrounding variant landscape is unlikely to play an important role in the degree of pathogenicity, although the numbers of cases were small.

In this study, we selected twelve *ABCA4* variants in intron 36 that were predicted to affect pre-mRNA splicing. Upon testing in our midigene assay, however, only three of them showed clear splicing aberration in HEK293T cells. Given the high prevalence of the c.5196+1137G>A change, this variant was further analyzed in retina-like cells to identify potential 'retina-specific' splicing defects as previously reported for other variants in *ABCA4* and *CEP290* (Parfitt et al. 2016; Albert et al. 2018). For the other eight, a similar tissue-specific mechanism may be operational, such that it is not possible to exclude pathogenicity solely based on the midigene assay in HEK293T. Thus, other *in vitro* systems alternative to the costly and time-consuming retinal progenitor cell/organoid assays are needed to assess the pathogenic nature of such variants in a high-throughput manner.

For each of the two overlapping PEs that we identified, we designed four AONs to redirect the aberrant splicing processes. Similar to our previous findings (Albert et al. 2018; Garanto et al. 2019; **Chapter 3**), not all AONs were effective. In general, AONs targeting certain predicted splicing enhancer motifs (SC35), a melting temperature above 48°C and having a GC-content between 40 and 60% are good predictors of efficacy (Aartsma-Rus et al. 2009; Garanto et al. 2019), yet for every PE this needs to be empirically determined. Given the low number of AON molecules tested, no detailed correlation between the parameters of the selected AONs and their efficacy was possible. As we observed previously (Albert et al. 2018; Garanto et al. 2019), a single mismatch between an AON and its target sequence was sufficient to completely abrogate its splicing modulation capacity. AON3, that was designed to block the 129-nt and 177-nt PE inclusion caused by the c.5196+1013A>G and c.5196+1056A>G variant, respectively, harbored the c.5196+1013G allele in its sequence, and was indeed only effective for the c.5196+1013A>G variant. For the two variants that create the same PE, c.5196+1137G>A and c.5196+1216C>A, the efficacy of the AONs were tested in two different systems. Interestingly, the results were similar but not entirely matching. When AON delivery was combined with the midigene assay in HEK293T cells for the c.5196+1216C>A variant, AONs 5 and -6 were the two most efficacious ones, followed by a partial effect of AONs 7 and -8. In contrast, whilst AON6 also remained the most efficacious in the iPSC-derived PPCs harboring the c.5196+1137G>A variant, the ability of AON5 to redirect splicing was hardly detectable anymore. This discrepancy is most likely due to the limitations in the test systems, e.g., the delivery to HEK293T cells is easier than that to iPSC-derived PPCs. In addition, the molecular

context is different: iPSC-derived PPCs have the entire 128-kb *ABCA4* pre-mRNA expressed, while in the midigene assay only an artificial fragment of 7.5 kb is expressed.

RNA therapies are gaining momentum for the treatment of a variety of diseases, both inherited as well as acquired ones (Zhou et al. 2019). For inherited retinal diseases, the first clinical trial based on AONs to redirect a splicing defect in *CEP290* has recently shown promising results (Cideciyan et al. 2019), whereas a second one, for *USH2A*, is currently ongoing (NCT03780257). We have used the same retinal organoid technology that was used in the proof of concept studies for the preclinical development of the CEP290 therapy (Dulla et al. 2018), to show our AON approach for c.5196+1137G>A is effective, which might facilitate a faster translation to the clinic. Both of the ongoing clinical trials consist of unilateral intravitreal delivery of AONs to patients harboring the corresponding mutation in either homozygous or compound heterozygous state. AONs are administered every six months, since these oligos have been chemically modified to enhance cellular uptake and stability. However, further improvement of the stability and/or delivery of these molecules would reduce the burden for the patients - as well as health care costs - for this type of therapy even more. Alternatives include a viral delivery (e.g., AONs packaged into adeno-associated viruses) (Goyenvalle et al. 2004; Garanto et al. 2016), although these strategies also come with downsides.

In conclusion, in this study we demonstrate a cluster of *ABCA4* variants in intron 36 which, in their own way, result in aberrant processing of *ABCA4* pre-mRNA. Whilst some variants appear to act in a cell-independent manner, a recurrent DI variant in *ABCA4* (c.5196+1137G>A) shows a more retina-specific defect. Detailed clinical characterization of patients harboring this variant show that the severity of the phenotype largely depends on the nature of the second *ABCA4* allele *in trans*, reasoning that the c.5196+1137G>A allele on itself can be considered as a moderate hypomorphic allele. Finally, we show in patient derived PPCs and 3D retinal organoids that AONs appear to be a promising tool to correct splicing defects associated with the pathogenic variants identified in this study, warranting further development of these molecules towards clinical trials in order to halt the progression of this disease.

Acknowledgements

First, we wish to thank the patients and families who kindly participated in this study. This work was supported by the Algemene Nederlandse Vereniging ter Voorkoming van Blindheid, Stichting Blinden-Penning, Landelijke Stichting voor Blinden en Slechtzienden, Stichting Oogfonds Nederland, Stichting Macula Degeneratie Fonds, and Stichting Retina Nederland Fonds (who contributed through UitZicht 2015-31 and 2018-21), together with the Rotterdamse Stichting Blindenbelangen, Stichting

Blindenhulp, Stichting tot Verbetering van het Lot der Blinden, Stichting voor Ooglijders, and Stichting Dowilvo (to A.G. and R.W.J.C.). This work was also supported by the Foundation Fighting Blindness USA, grant no. PPA-0517-0717-RAD (to S.A., F.P.M.C., A.G. and R.W.J.C.), the RetinaUK, grant no. GR596 (to F.P.M.C. and R.W.J.C), and the European Union, Marie Curie Sklodowska Action Initial Training Network StarT, grant agreement 813490 (to F.P.M.C, M.E.C and R.W.J.C). In addition, this work was supported by the National Institute Fight for Health Research (NIHR) Biomedical Research Centre at Moorfields Eye Hospital and UCL Institute of Ophthalmology, NIHR Bioresource – Rare Disease study, GA is supported by a Fight For Sight (UK) Early Career Investigator Award, NIHR Biomedical Research Centre at Great Ormond Street Hospital Institute of Child Health and Fight for Sight to, The Wellcome Trust (to MEC) and Moorfields Eye Charity. Finally, this work was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre at Moorfields Eye Hospital and UCL Institute of Ophthalmology, and the NIHR Bioresource Translational Research Consortium for Rare Disease (grant to A.W. and A.F.).

This research was made possible through access to the data and findings generated by the 100,000 Genomes Project. The 100,000 Genomes Project is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The 100,000 Genomes Project is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure. The 100,000 Genomes Project uses data provided by patients and collected by the National Health Service as part of their care and support. None of the funding organizations had a role in the design nor conduct of this research.

Supplemental data

Supplemental tables S1, S2, S3 and S4 are not included in this thesis and are available upon request.

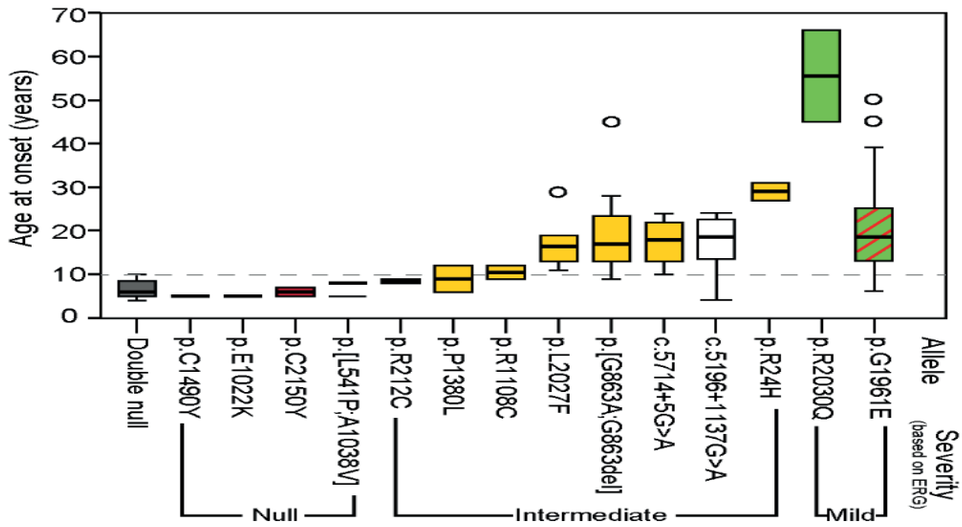


Figure S1. Distribution of the ages at onset associated with different *ABCA4* alleles in *trans* with null alleles. The figure is modified from Fakin et al. 2016, with the addition of c.5196+1137G>A. Different categories of alleles were determined based on full-field ERG. Allele p.G1961E had a specific preponderance to foveal damage, reflected by a relatively early onset of visual symptoms. Dashed line marks the 95% confidence interval of the patients harboring two null alleles.

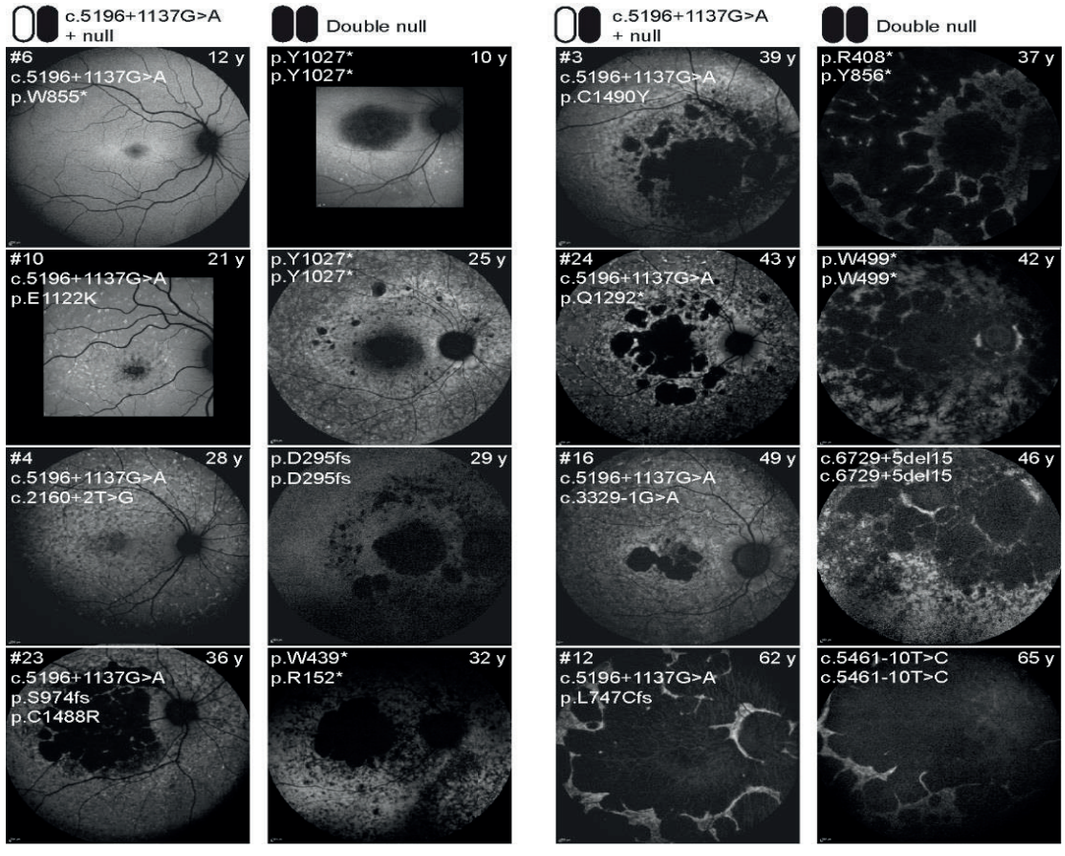


Figure S2. Fundus autofluorescence images of patients harboring c.5196+1137G>A in *trans* with null alleles and age-matched patients harboring two null alleles. Patient ID and genotypes are noted in the top left corner; age is noted in the top right corner of each image.

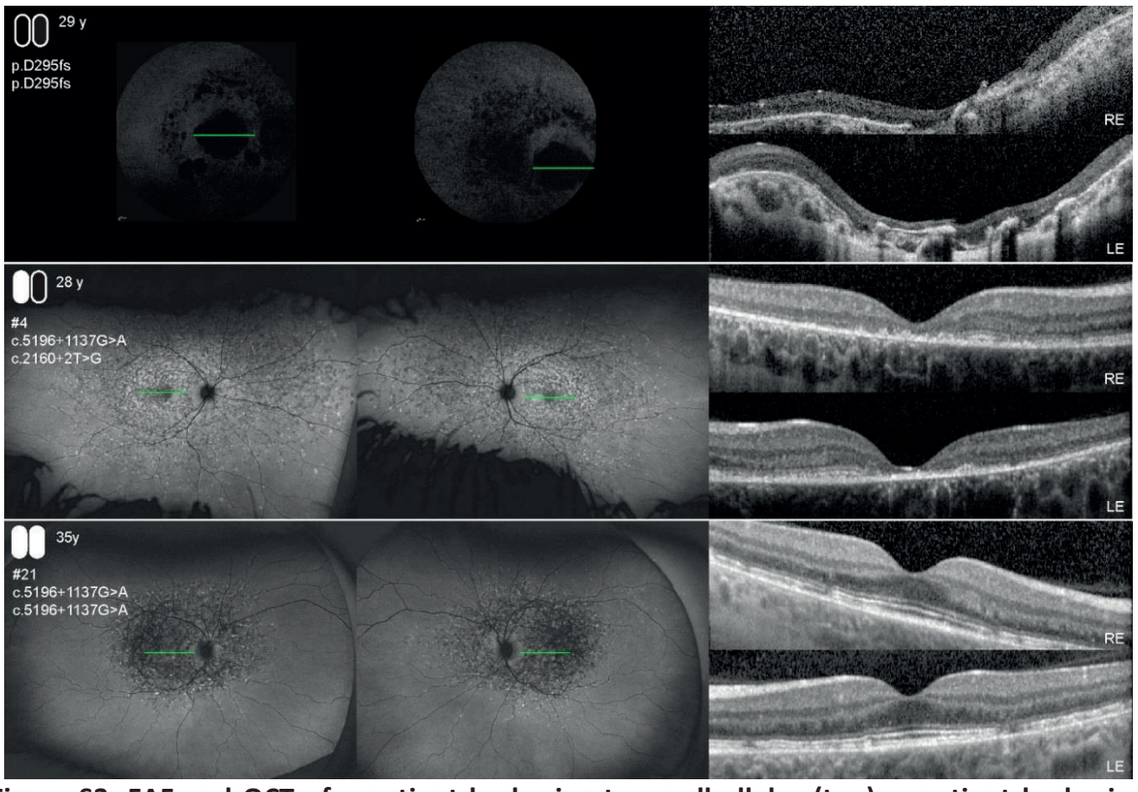


Figure S3. FAF and OCT of a patient harboring two null alleles (top), a patient harboring c.5196+1137G>A in *trans* with a null allele (middle) and a patient homozygous for c.5196+1137G>A patient of similar ages. Note the increasingly milder phenotype in the presence of c.5196+1137G>A alleles

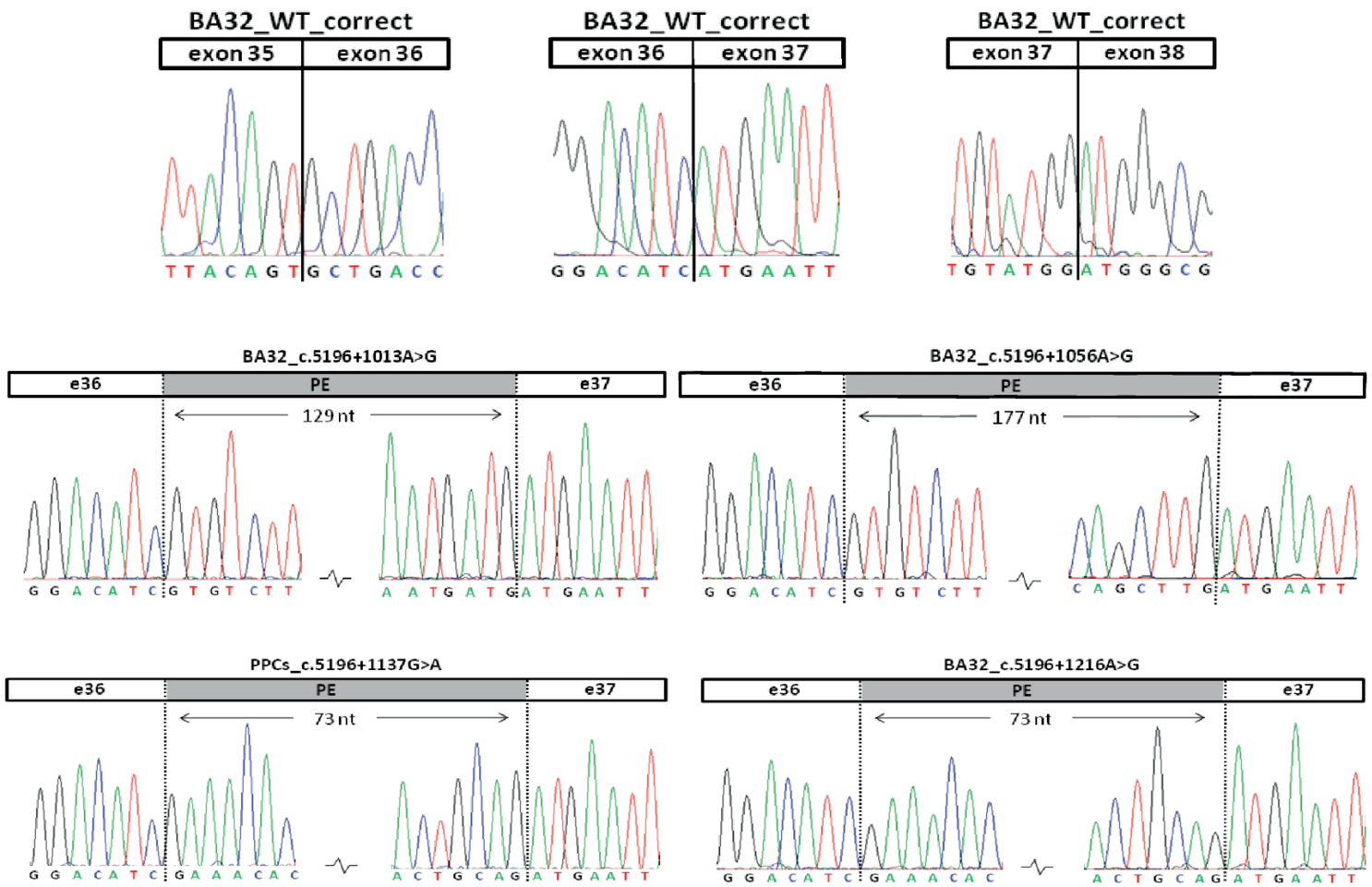


Figure S4. Sanger sequencing electropherograms for four causal *ABCA4* variants in intron 36 and correct transcript are given. Sequence traces for variants tested in *in vitro* splice assays are labeled as BA32_respective variant, whereas results obtained from photoreceptor progenitor cells (PPCs) are labelled as PPCs variants. PE, pseudoexon; nt, nucleotide.

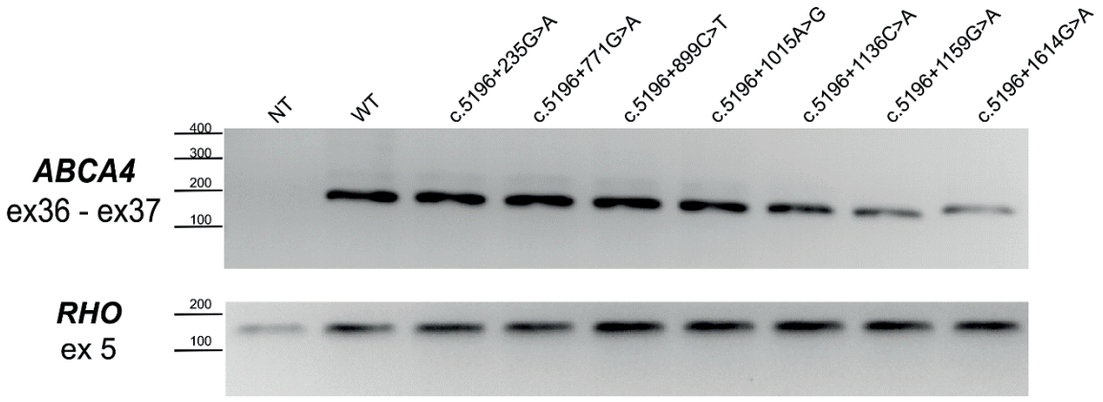


Figure S5. Seven *ABCA4* variants in intron 36 were tested in wild-type (WT) construct BA32 (*ABCA4* exons 35 - exon 38). RT-PCR was performed by using exonic (ex) primers in exon 36 and 37 of *ABCA4* for WT and all the variants. Rhodopsin (*RHO*) exon 5 amplification was used as a transfection and loading control.

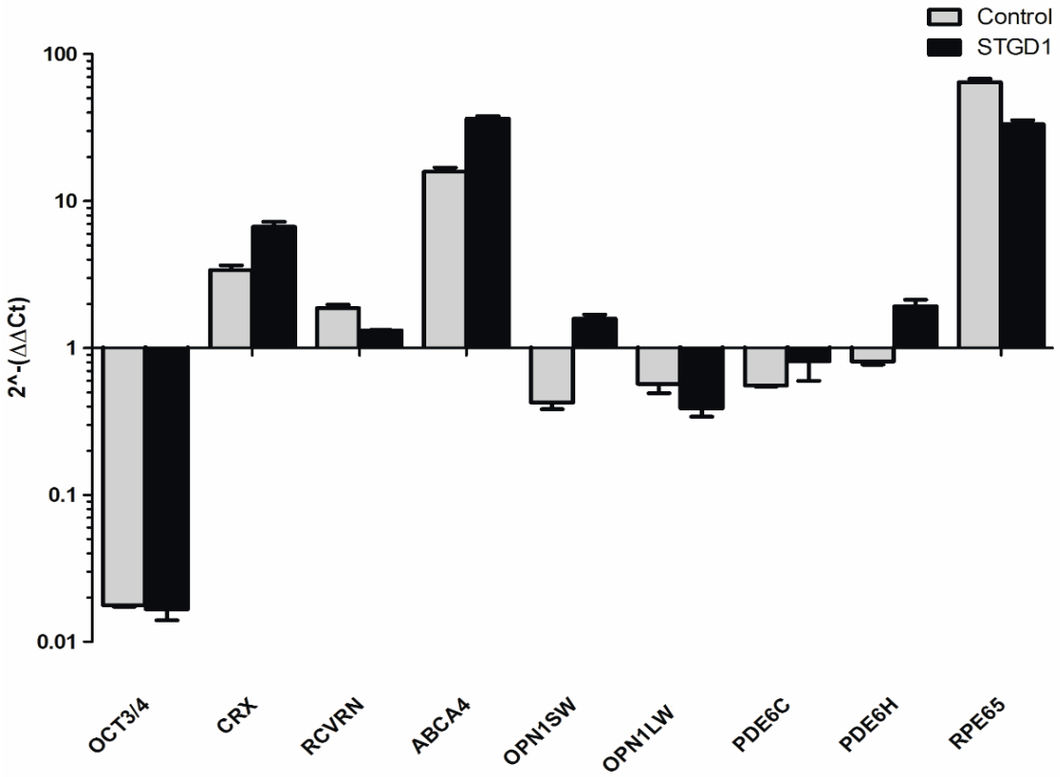


Figure S6. Gene expression profile of photoreceptor precursor cells (PPCs) derived from control (grey bars) and c.5196+1137G>A STGD1 (black bars) iPSCs. Expression levels were assessed by qPCR, normalized to GUSB expression and compared to the expression profile of day 0 iPSCs (undifferentiated). Pluripotency marker expression (*OCT3/4*) was reduced, while the expression of the photoreceptor precursor marker (*CRX*) was increased. The differentiation into RPE-like cells is shown by the increased expression of *RPE65*, while PPCs (especially STGD1 cells) showed some expression of early (*RCVRN, OPN1SW*) and late (*OPN1LW, PDE6C or PDE6H*). The expression of *ABCA4* was highly increased in all cell lines.

The results are shown as the mean ± SD of two experimental replicates with three technical replicates each.

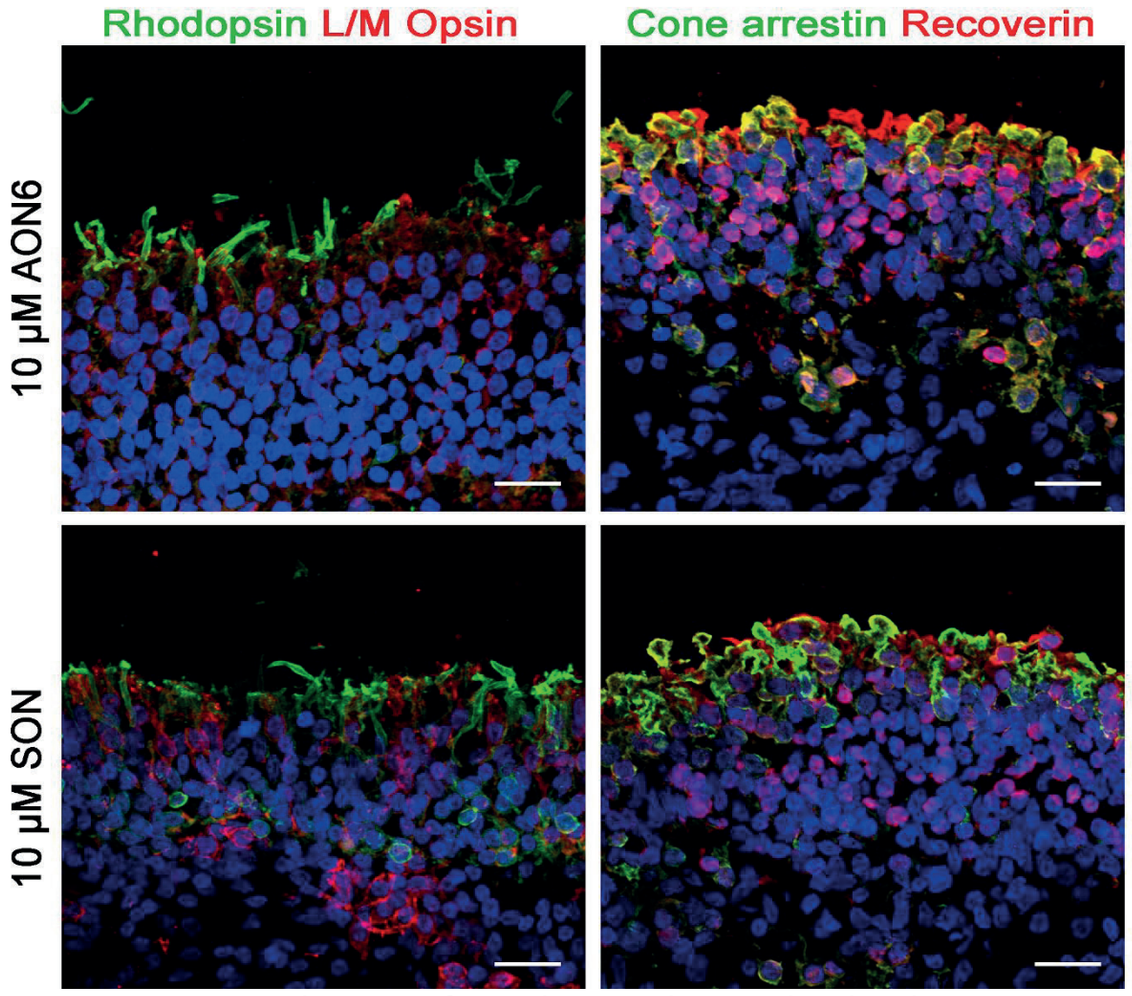


Figure S7. Immunocytochemistry of retinal organoids treated with either AON or SON. Sections were stained with antibodies for rhodopsin (green, left panels), L/M opsin (red, left panels), cone arrestin (green, right panels) and recoverin (red, right panels). Nuclei were stained with DAPI in blue. Scale bar: 20 µM.

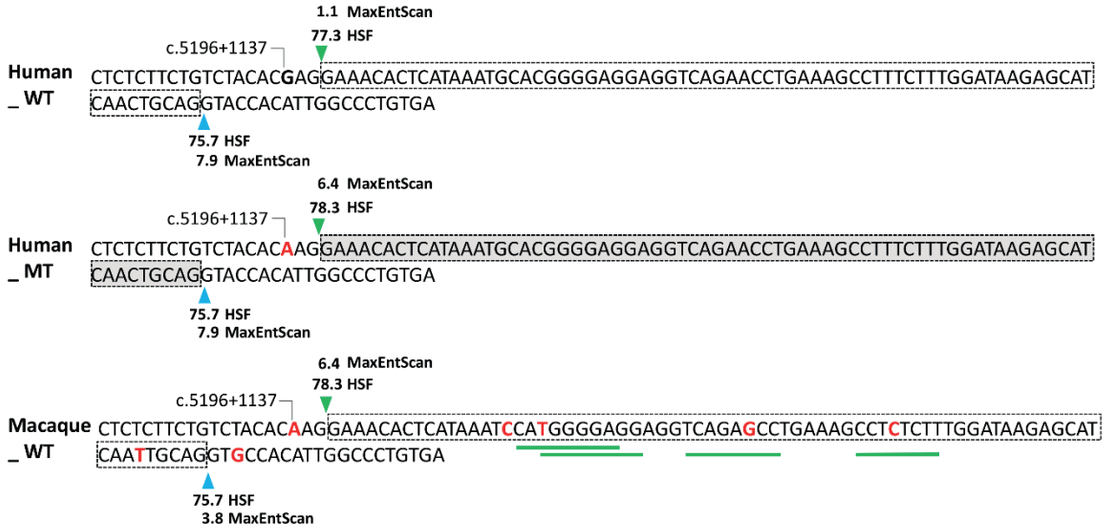


Figure S8. Splice-site predictions are given for human wild type (WT) and mutant (MT) sequence as well as for the macaque WT. Strength of the WT and MT splice acceptor site (SAS) and splice donor site (SDS) were calculated using Human splicing Finder (HSF) and MaxEnt splicing prediction tools, by taking 20 nt upstream and downstream of the 73-nt pseudoexon (PE). The PE found in the human mutant sequence is highlighted in grey. Nucleotides that are different among human WT/MT and macaque WT are shown in red. Green bars indicate the newly created splicing silencer elements due to the difference in human and macaque sequences. Green and blue rectangles indicate the SAS and SDS respectively.

References

Aartsma-Rus A. 2012. Overview on AON design. *Methods Mol Biol* **867**: 117-129.

Aartsma-Rus A, van Vliet L, Hirschi M, Janson AA, Heemskerk H, de Winter CL, de Kimpe S, van Deutekom JC, t Hoen PA, van Ommen GJ. 2009. Guidelines for antisense oligonucleotide design and insight into splice-modulating mechanisms. *Mol Ther* **17**: 548-553.

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

Allikmets R, Singh N, Sun H, Shroyer NF, 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.

Bauwens M, De Zaeytijd J, Weisschuh N, Kohl S, Meire F, Dahan K, Depasse F, De Jaegere S, De Ravel T, De Rademaeker M et al. 2015. An augmented ABCA4 screen targeting noncoding regions reveals a deep intronic founder variant in Belgian Stargardt patients. *Hum Mutat* **36**: 39-42.

Bauwens M, Garanto A, Sangermano R, Naessens S, Weisschuh N, De Zaeytijd J, Khan M, Sadler F, Balikova I, Van Cauwenbergh C et al. 2019. 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* doi:10.1038/s41436-018-0420-y.

Bax NM, Sangermano R, Roosing S, Thiadens AA, Hoefsloot LH, van den Born LI, Phan M, Klevering BJ, Westeneng-van Haaften C, Braun TA et al. 2015. Heterozygous deep-intronic variants and deletions in ABCA4 in persons with retinal dystrophies and one exonic ABCA4 variant. *Hum Mutat* **36**: 43-47.

Braun TA, Mullins RF, Wagner AH, Andorf JL, Johnston RM, Bakall BB, Deluca AP, Fishman GA, Lam BL, Weleber RG et al. 2013. Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. *Hum Mol Genet* **22**: 5136-5145.

Caminsky N, Mucaki EJ, Rogan PK. 2014. Interpretation of mRNA splicing mutations in genetic disease: review of the literature and guidelines for information-theoretical analysis. *F1000Res* **3**: 282.

Carss KJ, Arno G, Erwood M, Stephens J, Sanchis-Juan A, Hull S, Megy K, Grozeva D, Dewhurst E, Malka S et al. 2017. Comprehensive Rare Variant Analysis via Whole-Genome Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease. *Am J Hum Genet* **100**: 75-90.

Cideciyan AV, Jacobson SG, Drack AV, Ho AC, Charng J, Garafalo AV, Roman AJ, Sumaroka A, Han IC, Hochstedler MD et al. 2019. Effect of an intravitreal antisense oligonucleotide on vision in Leber congenital amaurosis due to a photoreceptor cilium defect. *Nat Med* **25**: 225-228.

Collin RW, Garanto A. 2017. Applications of antisense oligonucleotides for the treatment of inherited retinal diseases. *Curr Opin Ophthalmol* **28**: 260-266.

Conley SM, Cai X, Makkia R, Wu Y, Sparrow JR, Naash MI. 2012. Increased cone sensitivity to ABCA4 deficiency provides insight into macular vision loss in Stargardt's dystrophy. *Biochim Biophys Acta* **1822**: 1169-1179.

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

Cremers FPM, van de Pol DJ, van Driel MA, den Hollander AI, van Haren FJ, Knoers NV, Tijmes N, Bergen AA, 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.

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**: e67.

Dulla K, Aguila M, Lane A, Jovanovic K, Parfitt DA, Schulkens I, Chan HL, Schmidt I, Beumer W, Vorthoren L et al. 2018. Splice-Modulating Oligonucleotide QR-110 Restores CEP290 mRNA and Function in Human c.2991+1655A>G LCA10 Models. *Mol Ther Nucleic Acids* **12**: 730-740.

Fakin A, Robson AG, Chiang JP, Fujinami K, Moore AT, Michaelides M, Holder GE, Webster AR. 2016a. The Effect on Retinal Structure and Function of 15 Specific ABCA4 Mutations: A Detailed Examination of 82 Hemizygous Patients. *Invest Ophthalmol Vis Sci* **57**: 5963-5973.

Fakin A, Robson AG, Fujinami K, Moore AT, Michaelides M, Pei-Wen Chiang J, G EH, Webster AR. 2016b. Phenotype and Progression of Retinal Degeneration Associated With Nullizigosity of ABCA4. *Invest Ophthalmol Vis Sci* **57**: 4668-4678.

Flamier A, Barabino A, Gilbert B. 2016. Differentiation of Human Embryonic Stem Cells into Cone Photoreceptors. *Bio-protocol* **6**: e1870.

Garanto A, Chung DC, Duijkers L, Corral-Serrano JC, Messchaert M, Xiao R, Bennett J, Vweandenberghe LH, Collin RW. 2016. In vitro and in vivo rescue of aberrant splicing in CEP290-associated LCA by antisense oligonucleotide delivery. *Hum Mol Genet* **25**: 2552-2563.

Garanto A, Collin RWJ. 2018. Design and In Vitro Use of Antisense Oligonucleotides to Correct Pre-mRNA Splicing Defects in Inherited Retinal Dystrophies. *Methods Mol Biol* **1715**: 61-78.

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

Goyenvalle A, Vulin A, Fougerousse F, Leturcq F, Kaplan JC, Garcia L, Danos O. 2004. Rescue of dystrophic muscle through U7 snRNA-mediated exon skipping. *Science* **306**: 1796-1799.

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

Lee Y, Rio DC. 2015. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem* **84**: 291-323.

Lois N, Holder GE, Bunce C, Fitzke FW, Bird AC. 2001. Phenotypic subtypes of Stargardt macular dystrophy-fundus flavimaculatus. *Arch Ophthalmol* **119**: 359-369.

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

Nakano T, Ando S, Takata N, Kawada M, Muguruma K, Sekiguchi K, Saito K, Yonemura S, Eiraku M, Sasai Y. 2012. Self-formation of optic cups and storable stratified neural retina from human ESCs. *Cell Stem Cell* **10**: 771-785.

Nassisi M, Mohand-Said S, Andrieu C, Antonio A, Condroyer C, Mejecase C, Varin J, Wohlschlegel J, Dhaenens CM, Sahel JA et al. 2019. Prevalence of ABCA4 Deep-Intronic Variants and Related Phenotype in An Unsolved "One-Hit" Cohort with Stargardt Disease. *Int J Mol Sci* **20**.

Okita K. 2011. iPS cells for transplantation. *Curr Opin Organ Transplant* **16**: 96-100.

Parfitt DA, Lane A, Ramsden C, Jovanovic K, Coffey PJ, Hardcastle AJ, Cheetham ME. 2016a. Using induced pluripotent stem cells to understand retinal ciliopathy disease mechanisms and develop therapies. *Biochem Soc Trans* **44**: 1245-1251.

Parfitt DA, Lane A, Ramsden CM, Carr AJ, Munro PM, Jovanovic K, Schwarz N, Kanuga N, Muthiah MN, Hull S et al. 2016b. Identification and Correction of Mechanisms Underlying Inherited Blindness in Human iPSC-Derived Optic Cups. *Cell Stem Cell* **18**: 769-781.

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

Quazi F, Lenevich S, Molday RS. 2012. ABCA4 is an N-retinylidene-phosphatidylethanolamine and phosphatidylethanolamine importer. *Nat Commun* **3**: 925.

Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. *J Comput Biol* **4**: 311-323.

Runhart EH, Sangermano R, Cornelis SS, Verheij J, Plomp AS, Boon CJF, Lugtenberg D, Roosing S, Bax NM, Blokland EAW et al. 2018. The Common ABCA4 Variant p.Asn1868Ile Shows Nonpenetrance and Variable Expression of Stargardt Disease When Present in trans With Severe Variants. *Invest Ophthalmol Vis Sci* **59**: 3220-3231.

Runhart EH, Valkenburg D, Cornelis SS, Khan M, Sangermano R, Albert S, Bax NM, Astuti GDN, Gilissen C, Pott JR et al. 2019. Late-Onset Stargardt Disease Due to Mild, Deep-Intronic ABCA4 Alleles. *Invest Ophthalmol Vis Sci* **60**: 4249-4256.

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

Sangermano R, Garanto A, Khan M, Runhart EH, Bauwens M, Bax NM, van den Born LI, Khan MI, Cornelis SS, Verheij J et al. 2019a. Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. *Genet Med* **21**: 1751-1760.

Sangermano R, Garanto A, Khan M, Runhart EH, Bauwens M, Bax NM, van den Born LI, Khan MI, Cornelis SS, Verheij J et al. 2019b. Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. *Genet Med* doi:10.1038/s41436-018-0414-9.

Sangermano R, Khan M, Cornelis SS, Richelle V, Albert S, Garanto A, Elmelik D, Qamar R, Lugtenberg D, van den Born LI et al. 2018. ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Res* **28**: 100-110.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B et al. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**: 676-682.

Schulz HL, Grassmann F, Kellner U, Spital G, Ruther K, Jagle H, Hufendiek K, Rating P, Huchzermeyer C, Baier MJ et al. 2017. Mutation Spectrum of the ABCA4 Gene in 335 Stargardt Disease Patients From a Multicenter German Cohort-Impact of Selected Deep Intronic Variants and Common SNPs. *Invest Ophthalmol Vis Sci* **58**: 394-403.

Shapiro MB, Senapathy P. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* **15**: 7155-7174.

Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**: 663-676.

Tanna P, Strauss RW, Fujinami K, Michaelides M. 2017. Stargardt disease: clinical features, molecular genetics, animal models and therapeutic options. *Br J Ophthalmol* **101**: 25-30.

Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* **11**: 377-394.

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**: 404-412.

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

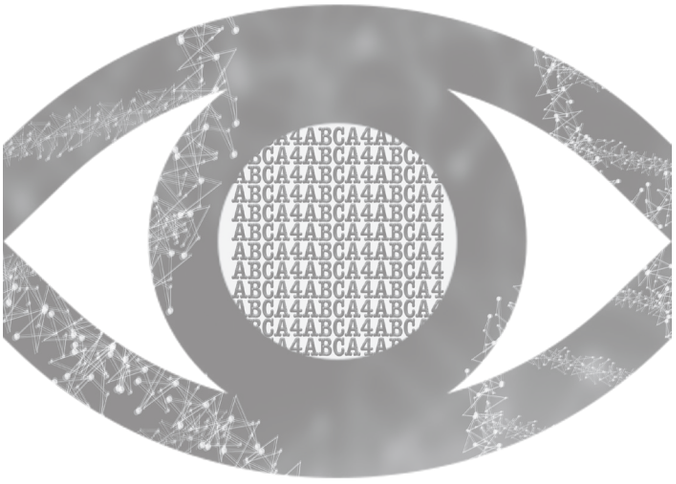
Zernant J, Xie YA, Ayuso C, Riveiro-Alvarez R, Lopez-Martinez MA, Simonelli F, Testa F, Gorin MB, Strom SP, Bertelsen M et al. 2014. Analysis of the ABCA4 genomic locus in Stargardt disease. *Hum Mol Genet*

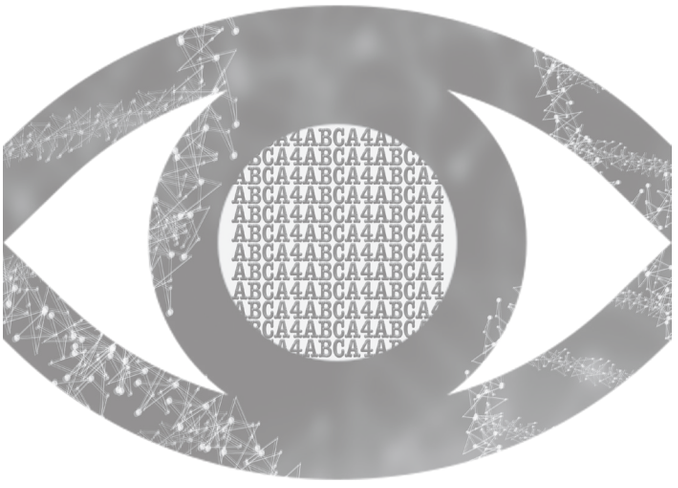
Zhang Z, Xin D, Wang P, Zhou L, Hu L, Kong X, Hurst LD. 2009. Noisy splicing, more than expression regulation, explains why some exons are subject to nonsense-mediated mRNA decay. *BMC Biol* **7**: 23.

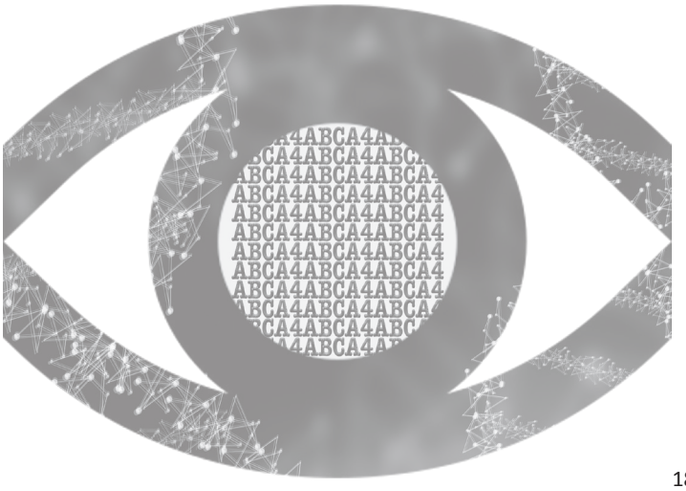
Zhou LY, Qin Z, Zhu YH, He ZY, Xu T. 2019. Current RNA-based Therapeutics in Clinical Trials. *Curr Gene Ther* **19**: 172-196.

Chapter 6

Identification of splice defects due to noncanonical splice site or deep-intronic variants in *ABCA4*







Identification of splice defects due to noncanonical splice site or deep-intronic variants in *ABCA4*

Zeinab Fadaie, **Mubeen Khan**, Marta Del Pozo-Valero, Stéphanie S. Cornelis, Carmen Ayuso, [*ABCA4* Studygroup, Rando Allikmets, Miriam Bauwens, Mohammad Ghofrani, Michael B. Gorin, Mohammad Keramatipur, Francesca Simonelli, Naeimeh Tayebi, Andrea Vincent, Nicole Weisschuh] Frans P.M. Cremers, Susanne Roosing

***Hum Mutat*.** 2019; **40**: 2365-2376.

Abstract

Pathogenic variants in the ATP-binding cassette transporter A4 (*ABCA4*) gene cause a continuum of retinal disease phenotypes, including Stargardt disease (STGD1). Noncanonical splice site (NCSS) and deep-intronic (DI) variants constitute a large fraction of disease-causing alleles, defining the functional consequences of which remains a challenge. We aimed to determine the effect on splicing of nine previously reported or unpublished NCSS variants, one near exon splice variant and nine DI variants in *ABCA4*, using *in vitro* splice assays in Human Embryonic Kidney (HEK293T) cells. Reverse transcription PCR and Sanger sequence analysis revealed splicing defects for 12 out of 19 variants. Four DI variants create pseudoexons or elongate the upstream exon. Furthermore, eight NCSS variants cause a partial deletion or skipping of one or more exons in mRNAs. Among the 12 variants, nine lead to premature stop codons and predicted truncated ABCA4 proteins. At least two DI variants are located in splicing enhancers and silencers and, therefore, these conserved sequences should be carefully evaluated when predicting the outcome of NCSS and DI variants.

Introduction

Inherited retinal diseases are clinically and genetically heterogeneous conditions (Sullivan and Daiger 1996; Berger et al. 2010), which makes it a great challenge for clinicians to come to a genetic diagnosis in affected individuals. However, by defining the genetic cause, the genetic risk of the disease for other family members can be assessed and it provides essential prognostic information for affected family members and possible therapeutic approaches. More knowledge on genetic variability in a gene will also provide a better insight and understanding of the disease mechanism (Ellingford et al. 2015; Carss et al. 2017).

The rise of next generation sequencing technology has drastically changed the opportunities in obtaining a genetic diagnosis in affected individuals (Neveling et al. 2012; Vaz-Drago et al. 2017). Using these techniques, hundreds of thousands of single-nucleotide variants are detected in each individual. However, in many instances the functional significance of variants remains unclear. Several *in silico* tools exist which predict the putative effect of missense and splice variants (SpliceSiteFinder-Like (SSFL), MaxEntScan, NNSPLICE, GeneSplicer, and Human Splicing Finder (HSF) (Shapiro and Senapathy 1987; Reese et al. 1997; Pertea et al. 2001; Yeo and Burge 2004; Desmet et al. 2009). The latter variants can be experimentally assessed for their pathogenicity by performing *in vitro* midigene splice assays (Runhart et al. 2018).

Stargardt disease (STGD1) (STGD1; MIM# 600110) is the most common inherited macular disease. It is characterized by bilateral progressive loss of central vision, color vision defect, photophobia and importantly a delayed dark adaptation and fundus imaging shows accumulation of lipofuscin (Stargardt 1909; Fishman et al. 1991). STGD1 is an autosomal recessive disease caused by pathogenic variants in the ATP-binding cassette subfamily A member 4 (*ABCA4*) gene (ABCA4; MIM# 601691; NM_000350.2) (Allikmets et al. 1997). The protein is comprised of two tandem halves, each of which consists of a nucleotide-binding domain, a cytoplasmic domain and a transmembrane domain, followed by a large extracellular segment (Bungert et al. 2001). *ABCA4* is a 128-kb gene containing 50 exons that encodes a polypeptide of 2,273 amino acids and is located in the rod and cone photoreceptor cells, as well as the retinal pigment epithelium (RPE) (Sun et al. 1999; Ahn et al. 2000; Lenis et al. 2018). One or both copies of this gene were found to be mutated in the majority of patients with STGD1 (Zernant et al. 2014), in 30% of patients with cone-rod dystrophy (Cremers et al. 1998; Maugeri et al. 2000), and approximately 5% of individuals with pan-retinal dystrophy or a phenotype resembling retinitis pigmentosa (Cremers et al. 1998).

The *ABCA4* gene carries many pathogenic non-canonical splice site (NCSS) variants. A comprehensive study on all *ABCA4* variants published up to 2016 showed a total of 5,962 likely causal variants or alleles in *ABCA4* of which 13.6% (809/5,962) are located in NCSS (Cornelis et al. 2017). These NCSS variants are located at the first and last three nucleotides of an exon as well as the -3 to -14 nucleotides from the acceptor site, and +3 to +6 from the donor site. Besides NCSS variants, many pathogenic variants were observed in canonical sequences at the AG-acceptor (-1 and -2) and GT-donor (+1 and +2) nucleotides. The functional consequences of many NCSS variants in *ABCA4* were revealed using *in vitro* mini- and midigene splice assays in HEK293T cells (Schulz et al. 2017; Runhart et al. 2018).

Until recently, a single pathogenic variant was identified in ~25% of STGD1 cases worldwide (Zernant et al. 2014; Zernant et al. 2017). About 40% of these cases (~10% of all STGD1 cases), many of whom showing late onset STGD1, were explained by a frequent coding variant c.5603A>T (p.Asn1868Ile) with an allele frequency of 0.07 in most control populations. This variant generally was found in *trans* with severe *ABCA4* variants (Zernant et al. 2017; Runhart et al. 2018). Another 40% of mono-allelic STGD1 probands (10% of all STGD1 cases) carried deep-intronic (DI) variants in their second alleles (Braun et al. 2013; Zernant et al. 2014; Bauwens et al. 2019; **Chapter 3**). The functional consequences of many of these variants only came to light when performing *in vitro* splice assays (Bauwens et al. 2019; **Chapter 3**). or by RT-PCR analysis of photoreceptor progenitor cells differentiated from patient-derived induced pluripotent stem cells (Sangermano et al. 2016; Albert et al. 2018).

To expand the knowledge on the consequences of NCSS and DI variants in *ABCA4*, we studied reported and unpublished NCSS, as well as reported DI variants in *ABCA4* employing midigene-based splice assays. This study contributes to a deeper understanding of alternative splicing through the activation of cryptic splice sites and the presence of exonic splicing enhancers (ESEs) or exonic splicing silencers (ESSs) and provides evidence for the pathogenicity of 12 splicing variants, thereby significantly expanding our knowledge on the effects of putative splicing defects.

Materials and Methods

Editorial policies and ethical considerations

The study adhered to the tenets of the Declaration of Helsinki and was approved by the local ethics committees of each participating center. Written informed consent was obtained from patients prior to inclusion in the study.

Clinical studies

The 12 probands carried pathogenic *ABCA4* variants (**Table S1**) and were diagnosed with macular dystrophies or cone-rod dystrophy (**Tables S2** and **S3**). Medical records of each patient were reviewed for clinical examination including age of onset, visual acuity, fundus photography and electroretinogram, where available.

Selection of non-canonical splice site and deep-intronic *ABCA4* variants

Variants were selected for *in vitro* analysis when they adhered to the following criteria: 1) NCSS and 'Near Exon Aberrant RNA' (NEAR) splice variants were predicted to result in a reduction of at least 2% of the relative strength in at least two of five different splice site prediction algorithms (SpliceSiteFinder-Like (SSFL), MaxEntScan, NNSPLICE, GeneSplicer, and Human Splicing Finder (HSF)) (Shapiro and Senapathy 1987; Reese et al. 1997; Pertea et al. 2001; Yeo and Burge 2004; Desmet et al. 2009). Moreover, variants were selected when nearby (up to 300 bps) a cryptic splice site was strengthened or created.

DI variants were included if a mutant cryptic splice site was predicted by at least two out of five algorithms with a ≥75% score in the presence of an already existing other splice site within 300-nt, which together could result in the formation of a pseudoexon (PE) (**Chapter 3**). As it has been shown that ESEs and ESSs have a significant effect on the splicing process in human cells we assessed ESEs through ESEfinder, an *in silico* prediction tool integrated in *Alamut*® version 2.10 (Fairbrother and Chasin 2000; Cartegni et al. 2002). ESEfinder determines the presence of five different ESE elements, i.e., SF2/ASF, SF2/ASF (IgM-BRCA1), SC35, SRp40, and SRp55 (Cartegni et al. 2003; Smith et al. 2006). Moreover, suggested pathogenic variants adhered to the selection criteria when differences could be observed in predicted ESEs or ESSs between wild-type (WT) and mutant sequences. ESSs were assessed through algorithms incorporated in Human Splice Finder (HSF) as introduced by Wang et al. 2004 and Sironi et al. 2004 (Sironi et al. 2004; Wang et al. 2004).

Nineteen variants were selected to be assessed by a midigene splice assays. All *in silico* splice site prediction scores of the variants investigated in this study are provided in **Table S4**. Of these, 16 variants were previously reported (Zhang et al. 2009; Fujinami et al. 2013; Zernant et al. 2014; Cornelis et al. 2017; Zernant et al. 2017; Bauwens et al. 2019; Tayebi et al. 2019), we also assessed in-house data based on molecular inversion probes-based sequence analysis of 108 inherited retina disease-associated genes in ~5,000 probands (Roosing S. and Cremers FPM., unpublished data). While the latter analysis led to the inclusion of c.1937+5G>A, c.5715-5T>G and c.6147G>A, as these fulfilled the criteria described above.

Generation of *ABCA4* wild-type and mutant midigenes

Previously, we generated a library of 31 overlapping WT midigenes (BA1-BA31) (Runhart et al. 2018; **Chapter 4**). Through Gateway cloning and subsequent site-directed mutagenesis, mutant constructs were generated for all 19 variants investigated in this study (**Figure 1**). Subsequently, WT and mutant constructs were independently transfected in HEK293T cells, assessed through RT-PCR, gel analysis, and followed by Sanger sequencing of the observed fragments. When a multitude of products were observed after gel electrophoresis they were quantified using Fiji software as previously described (Runhart et al. 2018). In addition, for the c.5715-5T>G variant, all observed fragments were cloned via the pGEM-T Easy Vector system I (Promega, Madison, WI, USA) according to the manufacturer's protocol and analyzed by Sanger sequencing. All mutagenesis, exonic primers and quantification measurements are available in **Tables S5**, **S6 and S7**.

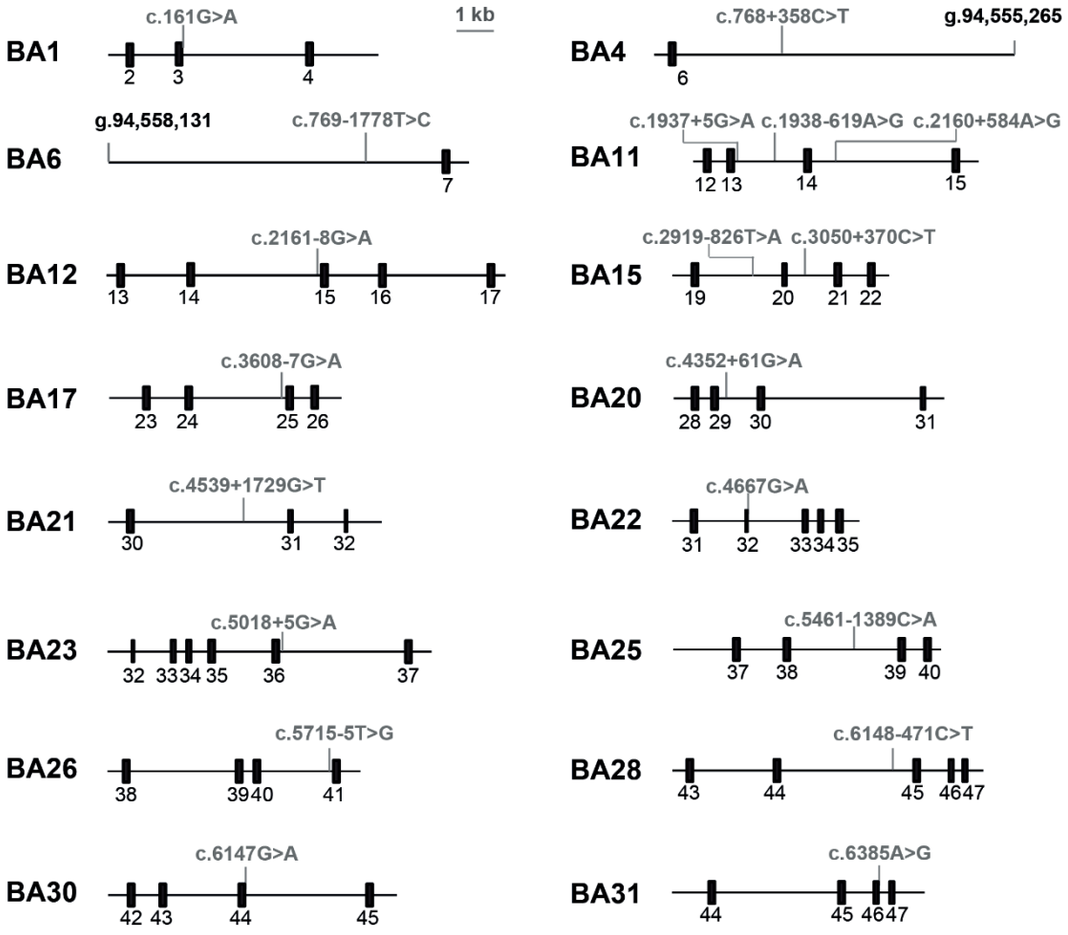


Figure 1. Schematic representation of mutant midigene splicing constructs of *ABCA4* and corresponding locations of the NCSS and deep-intronic variants tested. The exons are represented as black rectangles. BA depicts the BAC-clones used that were previously

described in **Chapter 2**. The BA11 and BA15 variants each were introduced separately into the wild-type constructs.

Results

Splicing effect of non-canonical splice site variants in *ABCA4*

The nine NCSS variants experimentally assessed with their observed effect on splicing are listed in **Table 1**. Of all variants only two were observed in the 'control' population database gnomAD (c.161G>A, allele frequency: 0.00003608; c.3608-7G>A, allele frequency: 0.000004087). Four of the selected variants (c.161G>A, c.4667G>A, c.6147G>A, and c.6385A>G) were in the coding regions of *ABCA4*, whereas the remaining NCSS variants were located in introns. Eight of the nine NCSS variants lead to partial or entire exon skipping, while only c.3608-7G>A did not show an effect on splicing. The c.1937+5G>A, c.2161-8G>A, and c.4667G>A variants led to partial in-frame deletions of *ABCA4* exons. The c.6385A>G variant caused an open reading frame-disrupting by 47-nt deletion of exon 46 due to the use of a cryptic splice donor site (SDS) at position c.6340 in the exon (**Figure 2**). The other five NCSS variants caused frameshift mutations and led to predicted truncated proteins (**Table 1**).

Variant c.5715-5T>G, located 5 nucleotides upstream of exon 41, showed a complex defect in splicing compared to the WT product. Among the different fragments, we observed exon 39 skipping as well as exon 39/40 skipping. The most prominent mutant mRNA contained a 4-nt insertion of the exon 41 splice acceptor site (SAS) along with exon 39/40 skipping. We did not observe any mRNA with only exon 40 skipping.

Splicing effect of deep-intronic variants

The ten NEAR and DI variants with their observed effect on splicing from this study are presented in **Table 2**. None of these variants was observed in gnomAD, indicating that they are very rare. Four showed a splicing defect, while the other six variants did not show an effect on splicing (**Figure S1**). Of the four variants that did show a splicing defect, c.1938-619A>G, c.2919-826T>A, and c.3050+370C>T created a PE in the mature mRNAs, leading to a frameshift and a subsequent predicted truncated protein. The NEAR splicing variant c.4352+61G>A elongated the *ABCA4* mRNA downstream of exon 29 by 57 nucleotides, resulting in a premature stop codon, likely due to an increased cryptic SDS score according to 5 prediction algorithms, even though there is an inactivation of several ESE protein binding motifs such as SC35 and SF2/ASF (**Figure 3**).

Of the six variants that showed no effect on splicing, the c.768+358C>T, c.769-1778T>C, and c.5461-1389C>A variants showed alternative ESEs being recognized in

the corresponding mutant and the c.2160+584A>G, c.4539+1729G>T, and c.6148-471C>T variants significantly increased the splicing scores in different algorithms (**Table S3**). However, none of these six variants showed splicing defects in HEK293T cells in the *in vitro* assay (**Figure S1**).

Table 1. *In vitro* assessed non-canonical splice site variants and the observed RNA and predicted protein effects.

DNA variant	RNA effect	Protein effect	Splicing defect	Variant effect
c.161G>A	r.161_302del	p.[Cys54Serfs*14,Cys54Tyr]	Exon 3 skipping	Moderate[#]
c.1937+5G>A	r.1806_1937del	p.(Tyr603_Ser646del)	132-nt exon 13 deletion	Severe
c.2161-8G>A	r.2161_2382del	p.[His721_Val794del,=][*]	Exon 15 skipping	Severe
c.3608-7G>A	r.=	p.(=)	None	n.a.
c.4667G>A	r.4635_4667del	p.(Ser1545_Gln1555del)	Exon 32 skipping	Severe
c.5018+5G>A	r.4849_5018del	p.(Val1617Alafs*113)	Exon 35 skipping	Severe
c.5715-5T>G	r.5461_5714;5460_5715del_ins(5715-4_5715-1)	p.[Thr1821Serfs*34,=][*]	Exon 39/40 skipping	Severe
c.6147G>A	r.6006_6147del	p.(Ser2002Argfs*11)	Exon 44 skipping	Severe
c.6385A>G	r.6340_6386del	p.(Val2114Hisfs*4)	47-nt exon 46 deletion	Severe

n.a., not applicable. [#] The splicing defect is classified as moderate. However, the missense variant Cys54Tyr was proposed earlier to be considered a severe variant in Stargardt cases. [*]The equal sign depicts the presence of >15% wild-type RNAs. The RNA and protein effect annotations show the most abundant product followed by the less abundant RNA product observed.

Table 2. ***In vitro*** **assessed near exon and deep-intronic variants and the observed RNA and predicted protein effects.**

DNA variant	RNA effect	Protein effect	Splicing defect	Variant effect
c.768+358C>T	r.=	p.(=)	None	n.a.
c.769-1778T>C	r.=	p.(=)	None	n.a.
c.1938-619A>G	r.1937_1938ins[1938-797_1938-624, 1937+396_1937+529;1938-797_1938-624]	p.[Phe647Alafs*22,Phe647Serfs*22]	134-nt and 174-nt pseudoexon(s)	Severe
c.2160+584A>G	r.=	p.(=)	None	n.a.
c.2919-826T>A	r.2918_2919ins2919-957_2919-825	p.[Leu973Phefs*1,=]*	133-nt pseudoexon	Severe
c.3050+370C>T	r.3050_3051ins3050+164_3050+368	p.(Leu1018Glufs*4)	205-nt pseudoexon	Severe
c.4352+61G>A	r.4352_4353ins4352+1_4352+57	p.[Glu1452*,=]*	57-nt exon 29 elongation	Severe
c.4539+1729G>T	r.=	p.(=)	None	n.a.
c.5461-1389C>A	r.=	p.(=)	None	n.a.

n.a., not applicable. *The equal sign depicts the presence of >15% wild-type RNAs. The RNA and protein effect annotations show the most abundant product followed by the less abundant RNA product observed

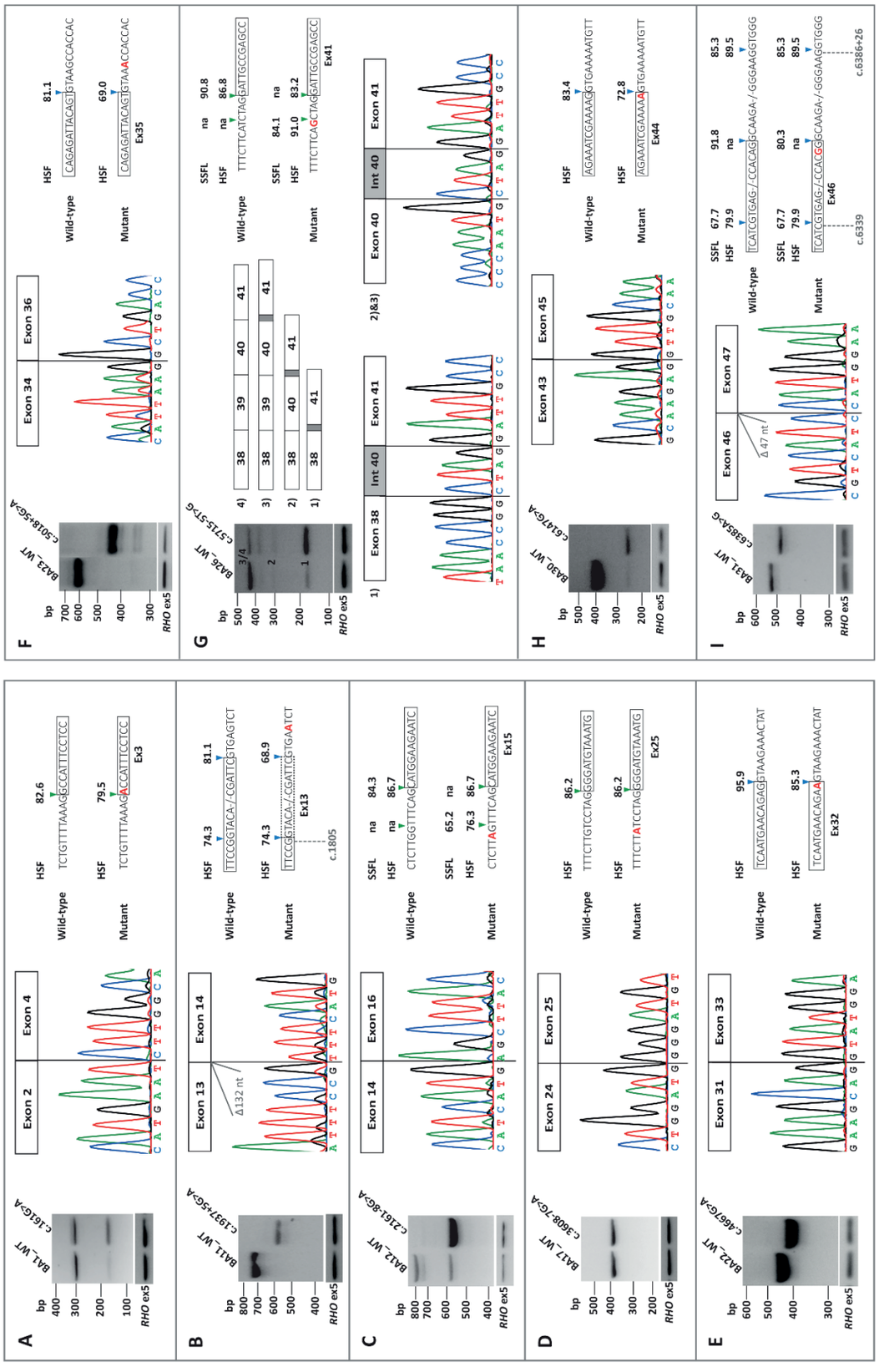


Figure 2. Overview of splicing defects for nine NCSS variants. Exon 5 *RHO* RT-PCR was used as a control for transfection efficiency. The chromatogram presents the nucleotides identified in the mutant midigene construct. **A.** Exon 3 showed weak natural exon skipping in the wild-type (WT) construct, which is significantly increased for the c.161G>A mutant. **B.** The recruitment of a cryptic SDS in exon 13 at position c.1806 resulted in a 132-nt deletion. **C.** The c.2161-8G>A mutant construct showed full exon 15 skipping. Note that exon 15 also shows natural exon skipping. **D.** c.3608-7G>A did not result in a splicing defect. **E, F.** Variants c.4667G>A and c.5018+5G>A led to complete exon 32 and 35 skipping, respectively. **G.** RT-PCR for the c.5715-5T>G mutant construct showed a complex splicing pattern. Exons 39 and 40 are partially skipped in the WT mRNA; c.5715-5T>G induced exon 39/40 skipping. This variant also created a new SAS upstream of exon 41 which led to the insertion of 4-nt into the mature mRNA. **H.** c.6147G>A caused complete exon 44 skipping. **I.** The use of a cryptic SDS site in exon 46 caused the 47-nt exon deletion. WT, wild-type; int, intron; nt, nucleotide; bp, base pairs.

Clinical characteristics of STGD1 cases carrying causal NCSS or deep-intronic variants

Clinical data were collected from the corresponding patients carrying the variants assessed in this study. The overview of the *ABCA4* variants observed in *trans* in these patients as well as phenotypic details are provided in **Tables S1** and **S2**, respectively.

For all variants assessed in this study we quantified visible fragments from gel electrophoresis analysis to determine the severity of variants. The quantifications observed ≤15% of WT mRNA in four of five variants with visible WT fragments suggesting that they represent severe variants (**Table S7**). For the analysis of c.161G>A we noted naturally occurring exon skipping for exon 3 in 14% of the WT construct. Therefore, we normalized the full-length fragment in mutant construct (44%) to the full-length RNA including exon 3 in the WT (86%). The splicing defect of c.161G>A was classified as moderately severe due to the resulting 51% (86/44*100) remaining product. However, already earlier the missense variant was proposed to be severe in STGD1 cases, and therefore we argue that both the exon skipping and missense mutation are contributing to the severity of this variant. Six other variants did not show any WT fragments besides the mutant fragment and are therefore deemed severe causal variants. Variant c.3608-7G>A did not show a splicing defect and is therefore likely not causative, until proven otherwise with another experimental setup.

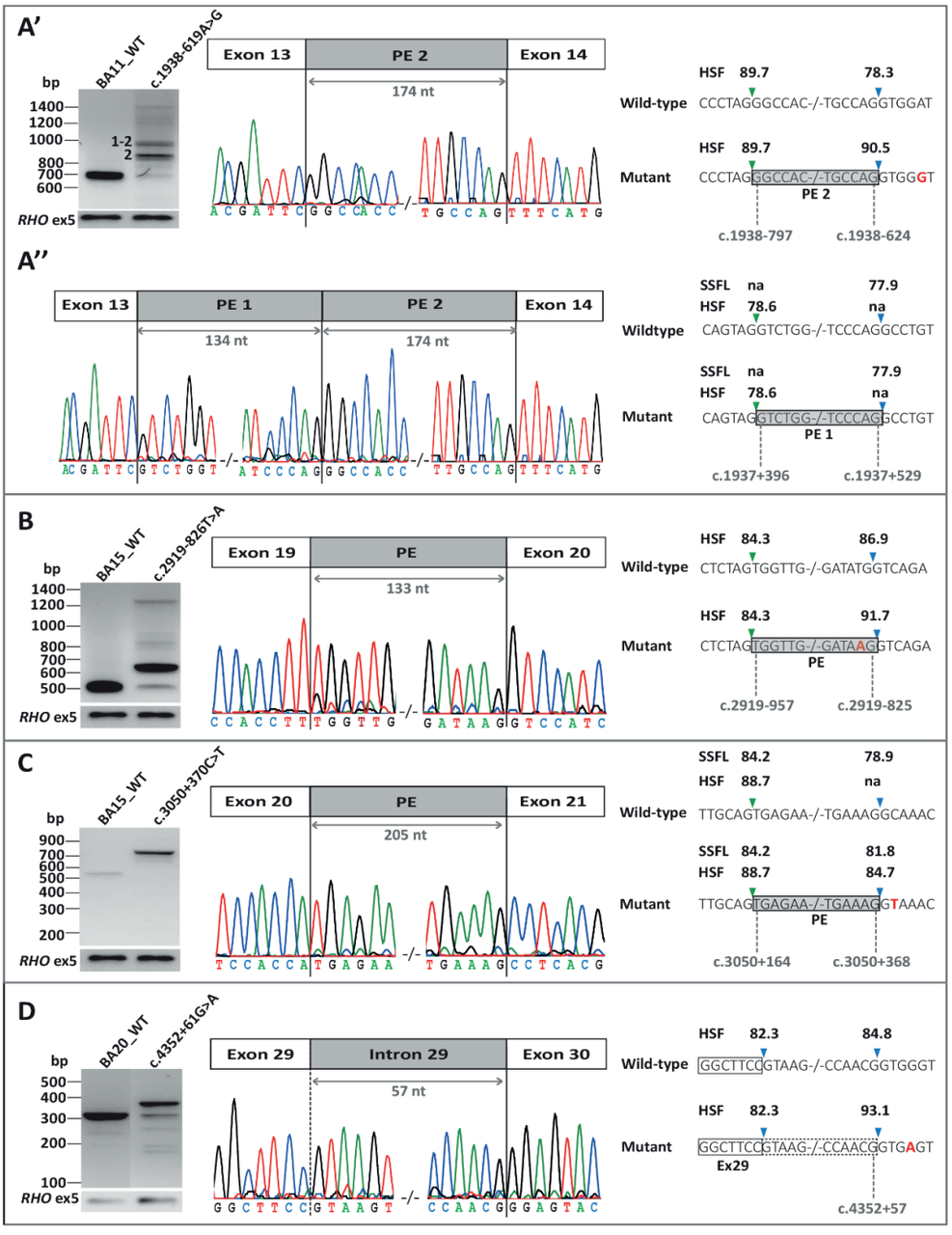


Figure 3. Overview of splicing defects for three deep-intronic variants and one near exon variant. All wild-type and mutant midigenes were transfected in HEK293T cells and their splicing effects were identified by RT-PCR. The exon 5 *RHO* RT-PCR was used as a control for transfection efficiency. The chromatograms present the nucleotides identified in the mutant midigene construct. **A.** Variant c.1938-619A>G strengthened a cryptic SDS and surprisingly resulted in two independent pseudoexon (PE) insertions. Fragment 1-2 consists of a fused fragment of two PEs encompassing 134-nt and 174-nt (PE1-2). PE1 (134-nt) starts from c.1937+396 to c.1937+529 and the boundaries of PE2 (174-nt) are c.1938-797 and c.1938-624. Fragment 2 contains the second PE of 174-nt (PE2) only. **B.** Variant c.2919-826T>A strengthened a cryptic SDS and led to a 133-nt PE. **C.** Variant c.3050+370C>T created a new canonical SDS which resulted in a 105-nt PE insertion. **D.** The NEAR splicing variant c.4352+61G>A enhanced a cryptic SDS's HSF score and elongated exon 29 with 57 nt in 84% percent of the mRNA. bp,

base pairs; HSF, Human Splicing Finder; nt, nucleotide; PE, pseudoexon; SSFL, SpliceSiteFinder-Like.

Through this study we established the splicing defects for 12 *ABCA4* variants found in 12 macular dystrophy probands. The c.1938-619A>G variant is located at position +5 of the PE and strengthened a cryptic SDS which likely led to the recognition of ESE SC35 and SRp55 motifs. These ESEs are located in the exon-intron boundaries and are shown to promote the recognition of exons with weak 5' and 3' splice sites and are be involved in exon definition by assisting in the recruitment of splicing factors prior to the removal of the adjacent introns. Therefore, the recognition of these ESE motifs will facilitate the splicing machinery to detect the nearby sequence as an exon (Lam and Hertel 2002; Wu et al. 2005; Buvoli et al. 2007). The individual carrying this variant was identified with c.5882G>A (p.Gly1961Glu) as a second variant and was shown to segregate with the disease in the family with a typical STGD1 phenotype (Zernant et al. 2014). The c.2919-826T>A variant is located at the penultimate nucleotide of the newly recognized PE, increasing the SDS scores at the c.2919-824 position. The corresponding patient carries the mild c.5882G>A (p.Gly1961Glu) variant on the other allele and has a characteristic for the p.Gly1961Glu variant, bull's eye maculopathy phenotype.

Discussion

We found splicing defects for 12 of the 19 (63%) assessed NCSS or DI *ABCA4* variants (**Figure S2**). Based on the midigene splice assays, eleven could be classified as severe variants and one variant (c.161G>A) is considered to have a moderately severe splicing effect. The splicing effects of all tested variants and their RNA and protein annotations were uploaded into the *ABCA4* Leiden Open (source) Variant Database (*ABCA4*-LOVD). In addition, we updated the protein outcome for those variants that were already present in the database (www.lovd.nl/ABCA4).

Among nine NCSS variants, six variants were found to cause skipping of one or two exons. The c.5715-5T>G variant revealed multiple fragments due to the splicing defect. All erroneous fragments showed a 5' elongation of exon 41, however the variant also leads to skip of both exon 39 and 40 and of exon 39 only. Whereas most NCSS variants lead to complete exon skipping, the midigene splice assays for both c.1937+5G>A and c.6385A>G variants resulted in partial

exon deletions due to the use of cryptic splice sites within exons 13 and 46, respectively. As observed for c.1937+5G>A, an upstream cryptic SDS at c.1807 was utilized which led to an in-frame 132-nt deletion of exon 13, thereby removing amino acid residues 603 to 646. This affects the first extracellular domain of the ABCA4 protein, but residual function of the protein cannot be excluded.

The prediction for the coding variant c.6385A>G was challenging as the canonical SDS of exon 46 contains GC instead of GT, which is recognized only by the SSFL algorithm. The SSFL values were reduced from 91.8 to 80.3% for the mutant variant. While we anticipated a 26-nt exon elongation due to high predicted scores for a cryptic splice site, or intron retention due to the small intron 46 size (73nt), we observed the use of an upstream cryptic SDS at position c.6335. This resulted in a 47-nt deletion of exon 46, leading to a frameshift resulting in a predicted truncated protein. We hypothesize that the strong cryptic SDS in c.6386+28 position (SSFL score: 85.3) may not be used by the splicing machinery due to the high abundance of silencers preventing the binding of splicing factors (**Figure S3**). The preferred cryptic SDS within the exon has a relatively low SSFL score of 67.7, but the region has few predicted silencer motifs and therefore likely is preferred over other SDSs. The corresponding patient with a retinitis pigmentosa-like phenotype has the c.5461-10T>C (p.[Thr1821Aspfs*6,Thr1821Valfs*13]) pathogenic variant as the other allele (Braun et al. 2013), which is the most frequent severe variant in *ABCA4* (Sangermano et al. 2016).

The prediction for the coding variant c.6385A>G was challenging as the canonical SDS of exon 46 contains GC instead of GT, which is recognized only by the SSFL algorithm. The SSFL values were reduced from 91.8 to 80.3% for the mutant variant. While we anticipated a 26-nt exon elongation due to high predicted scores for a cryptic splice site, or intron retention due to the small intron 46 size (73nt), we observed the use of an upstream cryptic SDS at position c.6335. This resulted in a 47-nt deletion of exon 46, leading to a frameshift resulting in a predicted truncated protein. We hypothesize that the strong cryptic SDS in c.6386+28 position (SSFL score: 85.3) may not be used by the splicing machinery due to the high abundance of silencers preventing the binding of splicing factors (**Figure S3**). The preferred cryptic SDS within the exon has a relatively low SSFL score of 67.7, but the region has few predicted silencer motifs and therefore likely is preferred over other SDSs. The corresponding patient with a retinitis pigmentosa-like phenotype has the c.5461-10T>C (p.[Thr1821Aspfs*6,Thr1821Valfs*13]) pathogenic variant as the other allele

(Braun et al. 2013), which is the most frequent severe variant in *ABCA4* (Sangermano et al. 2016).

Among the 10 NEAR splice and DI variants evaluated, c.1938-619A>G, c.2919-826T>A, and c.3050+370C>T generated PEs that contain stop codons and thus result in predicted truncated ABCA4 proteins. Vaz-Drago and colleagues recently showed that the majority of DI variants generate a canonical SAS or SDS, while the minority creates or disrupts an ESE or ESS element (Vaz-Drago et al. 2017). In our study only one out of three variants introduced a new splice site leading to a PE. Analysis of the c.1938-619A>G variant revealed a complex effect and introduced two PEs. The variant created a 134-nt PE as well as a second PE of 174-nt which resided 491-nt downstream. To our knowledge, the phenomenon of two PEs generated by a single variant has not been described before and the underlying mechanism remains to be elucidated.

The WT c.3050+370C residue is part of a cryptic 'GC-type' splice site as predicted by SSFL (SSFL score: 78.9%), but apparently is not employed by the splicing machinery. The c.3050+370C>T variant however creates a canonical GT (SSFL score: 81.8%) which was recognized by all five splice site algorithms, and led as expected to a PE insertion that contains a premature stop codon after four amino acids p.Leu1018Glufs*4.

Six DIvariants did not show a splicing effect, while there were strong predictions for cryptic SDSs for the c.2160+584A>G and c.4539-1729G>T variants. An explanation for the absence of PEs may be a paucity of retina-specific splicing motifs and/or an abundance of silencer motifs (Murphy et al. 2016). Variants showing no effect on splicing in this study may still be proven to be pathogenic when assessed in iPSC-derived photoreceptor precursors. For example, pathogenicity was proven when studying *ABCA4* variants c.4539+2001G>A and c.4539+2028C>T in photoreceptor precursor cells derived from patient fibroblasts while no effect on splicing was detected in fibroblasts of the same patients (Albert et al. 2018). Moreover, a retina-specific increase of a 128-nt PE insertion was also observed for the most frequent Leber congenital amaurosis-associated *CEP290* variant, c.2991+1655A>G (den Hollander et al. 2006; Dulla et al. 2018).

As current estimates indicate that NCSS, NEAR splice and DI variants represent 15-20% of the causes of recessive human diseases, (Matlin et al. 2005; Carss et al. 2017) we sought to assess pathogenicity of 19 *ABCA4* variants that were previously published or identified in our cohort. We clearly determined the effect on splicing and, consequently, the highly likely pathogenicity for six NCSS variants as well as three DI variants and thereby contribute to a growing

list of NCSS variants and the 16 pathogenic DI variants published previously (**Figure 4**) (Braun et al. 2013; Albert et al. 2018; Bauwens et al. 2019; **Chapter 3**). Moreover, our study revealed a fourth NEAR splicing variant to be pathogenic in addition to the three previously described NEAR splice variants (Runhart et al. 2018; Bauwens et al. 2019). An overview of the currently known pathogenic DI and NEAR splice variants is presented in **Figure 4**.

Determining the precise effects of splice-site variants will open new opportunities for therapeutic approaches for patients carrying these variants. As previously shown, modulation of *ABCA4* pre-mRNA splicing can be executed through antisense oligonucleotides, which can bind complementarily to mRNA and manipulate the splicing process by skipping PEs (Albert et al. 2018; Bauwens et al. 2019; **Chapter 3**). Moreover, antisense oligonucleotides have proven their effect in *in vitro* and *in vivo* studies for the most frequent DI variant in CEP290 that causes Leber congenital amaurosis (Collin et al. 2012; Garanto et al. 2016). The latter is currently in a Phase II clinical trial: https://clinicaltrials.gov/ct2/show/NCT03140969.

By evaluating the pathogenic effects of putative splicing variants, we gained crucial knowledge for evaluation of yet to be identified NCSS, NEAR or DI variants. While predictions by splice-site algorithms are crucial, we also observed an important, and possibly essential, role of ESE and ESS motifs. Future studies regarding retina-specific splicing motifs and proteins will improve predictions for the effect of novel NCSS and deep-intronic variants in *ABCA4* as well as in other IRD-associated genes.

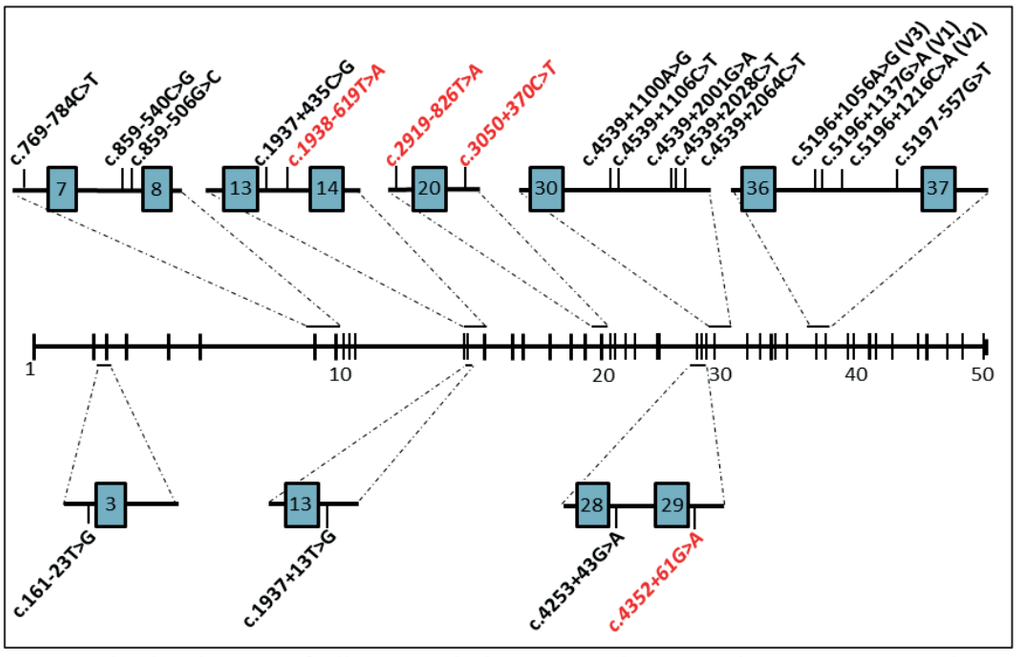


Figure 4. The landscape of all currently known pathogenic deep-intronic and NEAR splicing variants in *ABCA4*. The three DI variants and one NEAR splicing variant deemed as pathogenic in this study are depicted in red. Other variants were previously described elsewhere (Albert et al., 2018; Bauwens et al., 2019; Braun et al., 2013; **Chapter 3**).

Supplemental data

Supplemental data are not included in this thesis and they can be found online at:

https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fhumu.23890&file=humu23890-sup-0001-Fadaie_et_al__data_No_track-changes_rebuttal.pdf.

References

Ahn, J., Wong, J. T., & Molday, R. S. (2000). The effect of lipid environment and retinoids on the ATPase activity of ABCR, the photoreceptor ABC transporter responsible for Stargardt macular dystrophy. *Journal of Biological Chemistry, 275*(27), 20399-20405.

Albert, S., Garanto, A., Sangermano, R., Khan, M., Bax, N. M., Hoyng, C. B., . . . Collin, R. W. (2018). Identification and rescue of splicing defects caused by two neighboring deep-intronic ABCA4 mutations underlying Stargardt disease. *The American Journal of Human Genetics, 102*(4), 517-527.

Allikmets, R., Shroyer, N. F., Singh, N., Seddon, J. M., Lewis, R. A., Bernstein, P. S., . . . Hutchinson, A. (1997). Mutation of the Stargardt disease gene (ABCR) in age-related macular degeneration. *Science, 277*(5333), 1805-1807.

Bauwens, M., Garanto, A., Sangermano, R., Naessens, S., Weisschuh, N., De Zaeytijd, J., . . . De Baere, E. (2019). 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*. doi:10.1038/s41436-018-0420-y

Berger, W., Kloeckener-Gruissem, B., & Neidhardt, J. (2010). The molecular basis of human retinal and vitreoretinal diseases. *Progress in retinal and eye research, 29*(5), 335-375.

Braun, T. A., Mullins, R. F., Wagner, A. H., Andorf, J. L., Johnston, R. M., Bakall, B. B., . . . Weleber, R. G. (2013). Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. *Human molecular genetics, 22*(25), 5136-5145.

Bungert, S., Molday, L. L., & Molday, R. S. (2001). Membrane Topology of the ATP Binding Cassette Transporter ABCR and Its Relationship to ABC1 and Related ABCA Transporters Identification of N-linked Glycosylation Sites. *Journal of Biological Chemistry, 276*(26), 23539-23546.

Buvoli, M., Buvoli, A., & Leinwand, L. A. (2007). Interplay between exonic splicing enhancers, mRNA processing, and mRNA surveillance in the dystrophic Mdx mouse. *PloS one, 2*(5), e427. doi:10.1371/journal.pone.0000427

Carss, K. J., Arno, G., Erwood, M., Stephens, J., Sanchis-Juan, A., Hull, S., . . . Malka, S. (2017). Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. *The American Journal of Human Genetics, 100*(1), 75-90.

Cartegni, L., Chew, S. L., & Krainer, A. R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature reviews genetics, 3*(4), 285-298.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q., & Krainer, A. R. (2003). ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic acids research, 31*(13), 3568-3571.

Collin, R. W., Den Hollander, A. I., Van Der Velde-visser, S. D., Bennicelli, J., Bennett, J., & Cremers, F. P. (2012). Antisense oligonucleotide (AON)-based therapy for Leber congenital amaurosis caused by a frequent mutation in CEP290. *Molecular Therapy-Nucleic Acids, 1*, e14.

Cornelis, S. S., Bax, N. M., Zernant, J., Allikmets, R., Fritsche, L. G., den Dunnen, J. T., . . . Cremers, F. P. (2017). In Silico Functional Meta-Analysis of 5,962 ABCA4 Variants in 3,928 Retinal Dystrophy Cases. *Human mutation, 38*(4), 400-408.

Cremers, F. P., van de Pol, D. J., van Driel, M., den Hollander, A. I., van Haren, F. J., Knoers, N. V., . . . Hoyng, C. B. (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.

den Hollander, A. I., Koenekoop, R. K., Yzer, S., Lopez, I., Arends, M. L., Voesenek, K. E., . . . Brunner, H. G. (2006). Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. *The American Journal of Human Genetics, 79*(3), 556-561.

Desmet, F.-O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M., & Béroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic acids research, 37*(9), e67-e67.

Dulla, K., Aguila, M., Lane, A., Jovanovic, K., Parfitt, D. A., Schulkens, I., . . . Cheetham, M. E. (2018). Splice-Modulating Oligonucleotide QR-110 Restores CEP290 mRNA and Function in Human c.2991+1655A>G LCA10 Models. *Mol Ther Nucleic Acids, 12*, 730-740. doi:10.1016/j.omtn.2018.07.010

Ellingford, J. M., Sergouniotis, P. I., Lennon, R., Bhaskar, S., Williams, S. G., Hillman, K. A., . . . Lloyd, I. C. (2015). Pinpointing clinical diagnosis through whole exome sequencing to direct patient care: a case of Senior-Loken syndrome. *The Lancet, 385*(9980), 1916.

Fairbrother, W. G., & Chasin, L. A. (2000). Human genomic sequences that inhibit splicing. *Molecular and cellular biology, 20*(18), 6816-6825.

Fairbrother, W. G., Yeo, G. W., Yeh, R., Goldstein, P., Mawson, M., Sharp, P. A., & Burge, C. B. (2004). RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. *Nucleic acids research, 32*, W187-W190.

Fishman, G. A., Farbman, J. S., & Alexander, K. R. (1991). Delayed rod dark adaptation in patients with Stargardt's disease. *Ophthalmology, 98*(6), 957-962.

Fujinami, K., Zernant, J., Chana, R. K., Wright, G. A., Tsunoda, K., Ozawa, Y., . . . Allikmets, R. (2013). ABCA4 gene screening by next-generation sequencing in a British cohort. *Investigative ophthalmology & visual science, 54*(10), 6662-6674.

Garanto, A., Chung, D. C., Duijkers, L., Corral-Serrano, J. C., Messchaert, M., Xiao, R., . . . Collin, R. W. (2016). In vitro and in vivo rescue of aberrant splicing in CEP290-associated LCA by antisense oligonucleotide delivery. *Human molecular genetics, 25*(12), 2552-2563.

Lam, B. J., & Hertel, K. J. (2002). A general role for splicing enhancers in exon definition. *Rna, 8*(10), 1233-1241.

Lenis, T. L., Hu, J., Ng, S. Y., Jiang, Z., Sarfare, S., Lloyd, M. B., . . . Radu, R. A. (2018). Expression of ABCA4 in the retinal pigment epithelium and its implications for Stargardt macular degeneration. *Proc Natl Acad Sci U S A, 115*(47), E11120-e11127. doi:10.1073/pnas.1802519115

Matlin, A. J., Clark, F., & Smith, C. W. (2005). Understanding alternative splicing: towards a cellular code. *Nature reviews Molecular cell biology, 6*(5), 386.

Maugeri, A., Klevering, B. J., Rohrschneider, K., Blankenagel, A., Brunner, H. G., Deutman, A. F., . . . Cremers, F. P. (2000). Mutations in the ABCA4 (ABCR) gene

are the major cause of autosomal recessive cone-rod dystrophy. *The American Journal of Human Genetics, 67*(4), 960-966.

Murphy, D., Cieply, B., Carstens, R., Ramamurthy, V., & Stoilov, P. (2016). The Musashi 1 Controls the Splicing of Photoreceptor-Specific Exons in the Vertebrate Retina. *PLoS Genet, 12*(8), e1006256. doi:10.1371/journal.pgen.1006256

Neveling, K., Collin, R. W., Gilissen, C., van Huet, R. A., Visser, L., Kwint, M. P., . . . de Ligt, J. (2012). Next-generation genetic testing for retinitis pigmentosa. *Human mutation, 33*(6), 963-972.

Pertea, M., Lin, X., & Salzberg, S. L. (2001). GeneSplicer: a new computational method for splice site prediction. *Nucleic acids research, 29*(5), 1185-1190.

Reese, M. G., Eeckman, F. H., Kulp, D., & Haussler, D. (1997). Improved splice site detection in Genie. *Journal of computational biology, 4*(3), 311-323.

Runhart, E. H., Sangermano, R., Cornelis, S. S., Verheij, J. B., Plomp, A. S., Boon, C. J., . . . Blokland, E. A. (2018). The common ABCA4 variant p. Asn1868Ile shows nonpenetrance and variable expression of Stargardt disease when present in trans with severe variants. *Investigative ophthalmology & visual science, 59*(8), 3220-3231.

Sangermano, R., Bax, N. M., Bauwens, M., Van den Born, L. I., De Baere, E., Garanto, A., . . . Rohrschneider, K. (2016). Photoreceptor progenitor mRNA analysis reveals exon skipping resulting from the ABCA4 c. 5461-10T→ C mutation in Stargardt disease. *Ophthalmology, 123*(6), 1375-1385.

Sangermano, R., Garanto, A., Khan, M., Runhart, E. H., Bauwens, M., Bax, N. M., . . . Cremers, F. P. M. (2019). Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splicing defects by antisense oligonucleotides. *Genet Med*. doi:10.1038/s41436-018-0414-9

Sangermano, R., Khan, M., Cornelis, S. S., Richelle, V., Albert, S., Garanto, A., . . . van den Born, L. I. (2018). ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome research, 28*(1), 100-110.

Schulz, H. L., Grassmann, F., Kellner, U., Spital, G., Rüther, K., Jägle, H., . . . Weber, B. H. (2017). 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. *Investigative ophthalmology & visual science, 58*(1), 394-403.

Shapiro, M. B., & Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic acids research, 15*(17), 7155-7174.

Sironi, M., Menozzi, G., Riva, L., Cagliani, R., Comi, G. P., Bresolin, N., . . . Pozzoli, U. (2004). Silencer elements as possible inhibitors of pseudoexon splicing. *Nucleic acids research, 32*(5), 1783-1791.

Smith, P. J., Zhang, C., Wang, J., Chew, S. L., Zhang, M. Q., & Krainer, A. R. (2006). An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Human molecular genetics, 15*(16), 2490-2508.

Stargardt, K. (1909). Über familiäre, progressive Degeneration in der Maculagegend des Auges. *Graefe's Archive for Clinical and Experimental Ophthalmology, 71*(3), 534-550.

Sullivan, L. S., & Daiger, S. P. (1996). Inherited retinal degeneration: exceptional genetic and clinical heterogeneity. *Molecular medicine today, 2*(9), 380-386.

Sun, H., Molday, R. S., & Nathans, J. (1999). Retinal stimulates ATP hydrolysis by purified and reconstituted ABCR, the photoreceptor-specific ATP-binding cassette transporter responsible for Stargardt disease. *Journal of Biological Chemistry, 274*(12), 8269-8281.

Tayebi, N., Akinrinade, O., Khan, M. I., Hejazifar, A., Dehghani, A., Cremers, F. P., & Akhlaghi, M. (2019). Targeted next generation sequencing reveals genetic defects underlying inherited retinal disease in Iranian families. *Molecular vision, 8*(25), 106-117.

Vaz-Drago, R., Custódio, N., & Carmo-Fonseca, M. (2017). Deep intronic mutations and human disease. *Human genetics, 136*(9), 1093-1111.

Wang, Z., Rolish, M. E., Yeo, G., Tung, V., Mawson, M., & Burge, C. B. (2004). Systematic identification and analysis of exonic splicing silencers. *Cell, 119*(6), 831-845.

Wu, Y., Zhang, Y., & Zhang, J. (2005). Distribution of exonic splicing enhancer elements in human genes. *Genomics, 86*(3), 329-336. doi:10.1016/j.ygeno.2005.05.011

Yeo, G., & Burge, C. B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Journal of computational biology, 11*(2-3), 377-394.

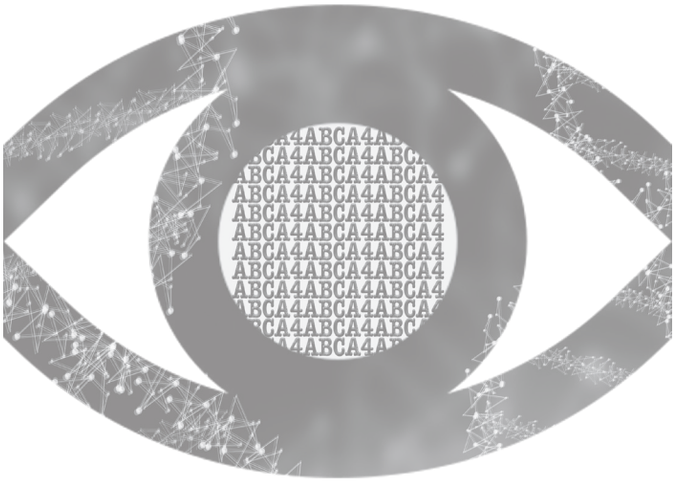
Zernant, J., Lee, W., Collison, F. T., Fishman, G. A., Sergeev, Y. V., Schuerch, K., . . . Allikmets, R. (2017). Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration. *Journal of medical genetics, 54*(6), 404-412.

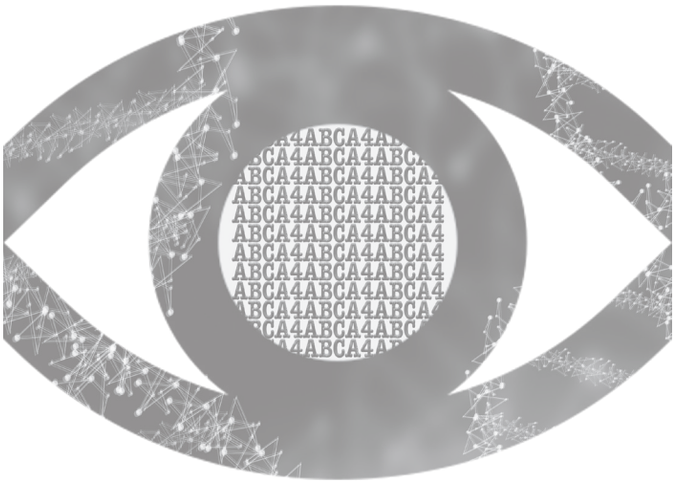
Zernant, J., Xie, Y., Ayuso, C., Riveiro-Alvarez, R., Lopez-Martinez, M.-A., Simonelli, F., . . . Bertelsen, M. (2014). Analysis of the ABCA4 genomic locus in Stargardt disease. *Human molecular genetics, 23*(25), 6797-6806.

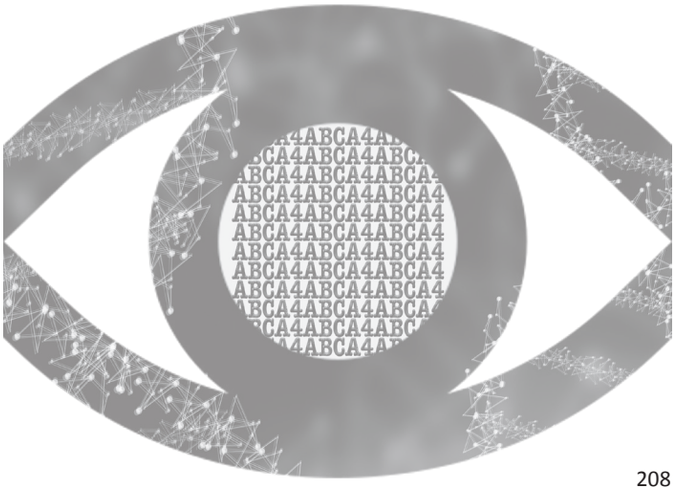
Zhang, X. H.-F., Arias, M. A., Ke, S., & Chasin, L. A. (2009). Splicing of designer exons reveals unexpected complexity in pre-mRNA splicing. *Rna, 15*(3), 367-376.

Chapter 7

Resolving the dark matter of *ABCA4* for 1,054 Stargardt disease probands through integrated genomics and transcriptomics







Resolving the dark matter of *ABCA4* for 1,054 Stargardt disease probands through integrated genomics and transcriptomics

Mubeen Khan, Stéphanie S. Cornelis, Marta del Pozo-Valero, Laura Whelan, Esmee H. Runhart, Ketan Mishra, Femke Bults, Yahya AlSwaiti, Alaa AlTalbishi, Elfride De Baere, Sandro Banfi, Eyal Banin, Miriam Bauwens, Tamar Ben-Yosef, Camiel J.F. Boon, L. Ingeborgh van den Born, Sabine Defoort, Aurore Devos, Adrian Dockery, Lubica Dudakova, Ana Fakin, G. Jane Farrar, Juliana Maria Ferraz Sallum, Kaoru Fujinami, Christian Gilissen, Damjan Glavač, Michael B. Gorin, Jacquie Greenberg, Takaaki Hayashi, Ymkje Hettinga, Alexander Hoischen, Carel B. Hoyng, Karsten Hufendiek, Herbert Jägle, Smaragda Kamakari, Marianthi Karali, Ulrich Kellner, Caroline C.W. Klaver, Bohdan Kousal, Tina Lamey, Ian M. MacDonald, Anna Matynia, Terri McLaren, Marcela D. Mena, Isabelle Meunier, Rianne Miller, Hadas Newman, Buhle Ntozini, Monika Oldak, Marc Pieterse, Osvaldo L. Podhajcer, Bernard Puech, Raj Ramesar, Klaus Rüther, Manar Salameh, Mariana Vallim Salles, Dror Sharon, Francesca Simonelli, Georg Spital, Marloes Steehouwer, Jacek P. Szaflik, Jennifer A. Thompson, Caroline Thuillier, Anna M. Tracewska, Martine van Zweeden, Andrea L. Vincent, Xavier Zanlonghi, Petra Liskova, Heidi Stöhr, John De Roach, Carmen Ayuso, Lisa Roberts, Bernhard H.F. Weber, Claire-Marie Dhaenens[#], Frans P.M. Cremers[#]

Genet Med. in press and BioRxiv. 2019; **doi**:
http://biorxiv.org/cgi/content/short/817767v1

[#]Shared senior authors

Abstract

Purpose: Missing heritability in human diseases represents a major challenge and this is particularly true for *ABCA4*-associated Stargardt disease (STGD1). We aimed to elucidate the genomic and transcriptomic variation in 1,054 unsolved STGD and STGD-like probands.

Methods: Sequencing of the complete 128-kb *ABCA4* gene was performed using single molecule molecular inversion probes (smMIPs), based on a semi-automated and cost-effective method. Structural variants (SVs) were identified using relative read coverage analyses and putative splicing defects were studied using *in vitro* assays.

Results: In 448 bi-allelic probands 14 known and 13 novel deep-intronic (DI) variants were found, resulting in pseudoexon (PE) insertions or exon elongations in 105 alleles. Intriguingly, intron 13 variants c.1938-621G>A and c.1938-514G>A resulted in dual PE insertions consisting of the same upstream, but different downstream PEs. The intron 44 variant c.6148-84A>T resulted in two PE insertions and flanking exon deletions. Eleven distinct large deletions were found, two of which contained small inverted segments. Uniparental isodisomy of chromosome 1 was identified in one proband.

Conclusions: Deep sequencing of *ABCA4* and midigene-based splice assays allowed the identification of SVs and causal deep-intronic variants in 25% of bi-allelic STGD1 cases, which represents a model study that can be applied to other inherited diseases.

Introduction

High throughput whole-genome sequencing (WGS) has made a huge impact in biology and is considered the most powerful genetic test to elucidate inherited human diseases (Carss, et al. 2017). It allows the unbiased detection of a wide spectrum of genetic variants including coding and non-coding single nucleotide variants (SNVs), as well as structural variants (SVs). However, sequencing and data storage costs as well as the possibility of secondary genetic findings hamper the use of genome sequencing.

Based on the advantages and limitations mentioned above, WGS is not the best method to perform sequence analysis of one or a few genes that are associated with a clinically distinct condition. This is illustrated by autosomal recessive Stargardt disease (STGD1) due to variants in the *ABCA4* gene. STGD1 is the most frequent inherited macular dystrophy with an estimated prevalence of 1/10,000 (Blacharski 1988). Thus far, 1,180 unique *ABCA4* variants have been reported in 8,777 alleles of 6,684 cases (www.lovd.nl/ABCA4) (Cornelis et al. 2017). A large proportion of the variants affect noncanonical splice site (NCSS) sequences, with variable effects on mRNA processing, (Schulz et al. 2017; **Chapters 2, 4** and **5**) and several deep-intronic (DI) variants have been identified (Braun et al. 2013; Zernant et al. 2014; Bauwens et al. 2015; Bax et al. 2015; Bauwens et al. 2019; **Chapters 3, 4** and **6**). Most of these DI variants strengthen cryptic splice sites resulting in the insertion of pseudoexons (PEs) in the mature *ABCA4* mRNA. SVs seem to be rare in *ABCA4*, (Maugeri et al. 1999; Zernant et al. 2014; Bax et al. 2015; Bauwens et al. 2019) although systematic copy number variant analyses have not been performed in most STGD1 cases.

Due to the relatively large size of the *ABCA4* gene (50 exons; 128,313 bp), mutation screening initially was restricted to mutation-scanning of the exons and flanking splice sites, by using several targeted approaches, but these scanning techniques showed poor sensitivity, leaving 50-70% of STGD1 probands genetically unsolved (Maugeri et al. 1999; Rivera et al. 2000; Jaakson et al. 2003; Maia-Lopes et al. 2009). More recently, sequence analysis of the entire 128-kb gene and up- and downstream DNA segments was performed using next-generation sequencing platforms after enrichment of *ABCA4* sequences using Raindance microdroplet-PCR target enrichment or Illumina TruSeq Custom Amplicon target enrichment, (Zernant et al. 2014) Haloplex-based sequence enrichment, (Bauwens et al. 2019; **Chapter 3**) or WGS (Carss et al. 2017; **Chapter 3**).

Identification of two pathogenic alleles is important to confirm the clinical diagnosis because several promising clinical trials are underway based on RNA modulation with antisense oligonucleotides, (Albert et al. 2018; Bauwens et al. 2019; **Chapter 3**) drug based therapies, such as compounds inhibiting lipofuscin

accumulation, (Charbel Issa et al. 2015) as well as gene augmentation (Allocca et al. 2008) and stem cell therapies (Lu et al. 2009). STGD1 cases will only be eligible for one of these therapies if both causal alleles are known. In addition, recent studies have identified alleles carrying a coding variant in *cis* with a deep-intronic variant, and only these combinations represented fully penetrant alleles, (Bauwens et al. 2019; **Chapter 3**) pointing towards the importance of analyzing non-coding regions in the STGD1 cases.

Recently, we reported the design and use of 483 single-molecule molecular inversion probes (smMIPs) to sequence the 50 coding exons and 12 intronic regions carrying 11 pathogenic DI variants of 412 genetically unsolved STGD1 cases (**Chapter 4**). In this study, we aimed to design a semi-automated, high throughput, cost-effective and comprehensive sequence analysis of the entire *ABCA4* gene which could serve as a model study to investigate human inherited diseases due to variants in one or a few genes. Using 3,866 smMIPs we sequenced 1,054 genetically unsolved STGD or STGD-like probands and 138 bi-allelic controls carrying known *ABCA4* variants. Novel NCSS and DI variants were tested *in vitro* for splicing defects. Additionally, a very high and reproducible read coverage allowed us to perform copy number variant analysis.

Materials and Methods

Samples

Written informed consent was obtained prior to participation in the study, and adhered to the Declaration of Helsinki. More information regarding the control samples and patients cohort and details of the Materials and Methods used, are provided in the Supplemental Material and Methods.

smMIPs design and *ABCA4* sequence analysis

Detailed information of the smMIPs-based *ABCA4* sequencing, selection of candidate splice variants and inclusion criteria is provided in the Supplemental Material and Methods.

Midigene-based splice assay

The effect of nine NCSS variants and 58 deep-intronic variants was assessed by midigene-based splicing assays employing 23 wild-type (WT) BA clones previously described (**Chapter 2**) and the newly designed BA32, BA33, BA34 and BA35. WT and mutant constructs were transfected in HEK293T cells and the extracted total RNA was

subjected to reverse transcription (RT)-PCR as described previously (**Chapter 2**). Details are provided in Supplemental Material and Methods.

Identification of copy number variants and assessment of the underlying mechanism

An excel script was employed to detect copy number variants (CNVs) using smMIP read depth. Only large (>400 bp) deletions and duplications in principle can be identified, but not insertions and inversions. Therefore the term CNV instead of SV was used in the following section. The read coverage of a single target was first normalized in comparison to all 3,375 targets by dividing the number of reads for a single target by the total number of reads for a given DNA sample. The coverage of each target of a patient was subsequently compared with the coverage of each target in all probands of that series using the following ratio: R = average coverage of each target of a proband/average coverage of each target of all probands in the run. In order to identify large deletions and duplications, the following ranges were defined: homozygous deletions, ratios (R) <0.3; heterozygous deletions, 0.3≤R<0.7; no CNV, 0.8≤R≤1.2; heterozygous duplications, 1.2<R≤1.7; homozygous duplications, R>1.7. To discriminate actual CNVs from artifacts, ratios per target were compared within each patient and among all the patients. Candidate CNVs were considered if at least 6 consecutive targets were deleted or duplicated.

Microhomology at the breakpoints was assessed using ClustalW, breakpoint regions were analysed for non-B motifs by tool (nBMST and QGRS Mapper) (for details see Supplemental Materials and Methods).

Semi-quantification of RT-PCR products

To quantify the ratios between correct and aberrant RT-PCR products, densitometric analysis was performed using Image J software.

Uniparental disomy detection

To test the presence of uniparental disomy (UPD), haplotype analysis was performed in one STGD1 case (DNA14-33085) using whole exome sequencing data.

Results

smMIPs performance and *ABCA4* sequencing

A pilot sequencing study was conducted using 15 STGD1 samples and five DNA samples of control individuals, revealing all previously identified 34 variants (**Table S1**). The average number of reads for the 20 DNA samples ranged from 10 to 152,500 per smMIP, with an overall average coverage of 933x for each smMIP.

In total 1,192 DNA samples were analyzed for variants in *ABCA4* using six NextSeq500 runs. The average number of reads of the 3,866 smMIPs was 377x. As most nucleotide positions are targeted with two smMIPs, the effective average coverage was ~700x. To determine the coverage of *ABCA4* in more detail, we calculated the average coverage of each nucleotide position for runs 1 to 5 combined (**Table S2**). To visualize the results, nucleotide positions that were not or poorly covered (≤10 reads), moderately covered (11-49 reads) or well covered (≥50 reads) are depicted in **Figure S1**. From the 128,366 nt of *ABCA4*, 1,980 nt (1.5%) were not or poorly covered, 1,410 nt (1.1%) were moderately covered, and 124,976 nt (97.4%) were well covered. Although *ABCA4* introns carry several repetitive elements (**Figure S1**), they only had a small effect on smMIPs design. Several larger repeats are present in up- and downstream regions of *ABCA4*, which resulted in the absence - or poor performance - of smMIPs. Sequencing of 1,192 samples yielded a total of 7,756 unique variants in *ABCA4* that are listed in **Table S3**.

Sensitivity and specificity of the smMIPs-based sequencing

To assess the sensitivity of the new smMIPs sequencing platform, we tested 123 previously genotyped samples (**Chapters 3** and **4**) in three series (runs 2, 3 and 6) (**Table S4**) as well as 15 control DNA samples carrying 13 different SVs spread throughout the *ABCA4* gene (run 6) (**Table S5**). All previously known SNVs (n=300) could be identified, yielding a sensitivity of 100%. Six additional variants were found due to low coverage in the previous studies, and three variants had not been annotated correctly previously.

***ABCA4* gene sequencing and identification of variants**

ABCA4 sequencing was performed for 1,054 genetically unsolved STGD and STGD-like patients. This revealed 323 unique (likely) pathogenic SNVs and 11 SVs in 1,144 alleles. Sixty-four of 323 SNVs (26%) and all 11 SVs were novel (**Table S6**). Detailed *in silico* analysis of novel SNVs is provided in **Table S7**. Thirteen percent of these alleles were represented by DI variants and SVs and another 10% accounted for NCSS variants (**Figure 1A**). All variants and the respective cases were uploaded into the *ABCA4* variant and STGD1 cases database LOVD, at www.lovd.nl/ABCA4**.**

Two (likely) pathogenic variants were found in 326 probands, three of them carrying p.Asn1868Ile in a homozygous manner, and one (likely) causal variant in *trans* with p.Asn1868Ile was found in 125 probands. Only one (likely) causal variant was identified in 174 probands. Additionally, in 65 probands, the p.Asn1868Ile variant was the only identified variant (**Table S8**). No (likely) causal variants were found in 364 cases.

Among the SNVs, the most common causal alleles were c.5603A>T (n=134), c.5882G>A (n=84), c.[5461-10T>C;5603A>T] (n=44), c.[1622T>C;3113C>T] (n=30), c.[4469G>A;5603A>T] (n=27), c.4539+2001G>A (n=26), c.6079C>T (n=23) and c.4253+43G>A (n=21) (**Table S6**). To visualize the relative frequency of causal STGD1-causing alleles, we excluded 65 heterozygous c.5603A>T alleles that were found as the only *ABCA4* allele in these cases, as they were most likely present because of its high allele frequency (0.06) in the general population (**Figure S2**) (Zernant et al. 2017; Runhart et al. 2018).

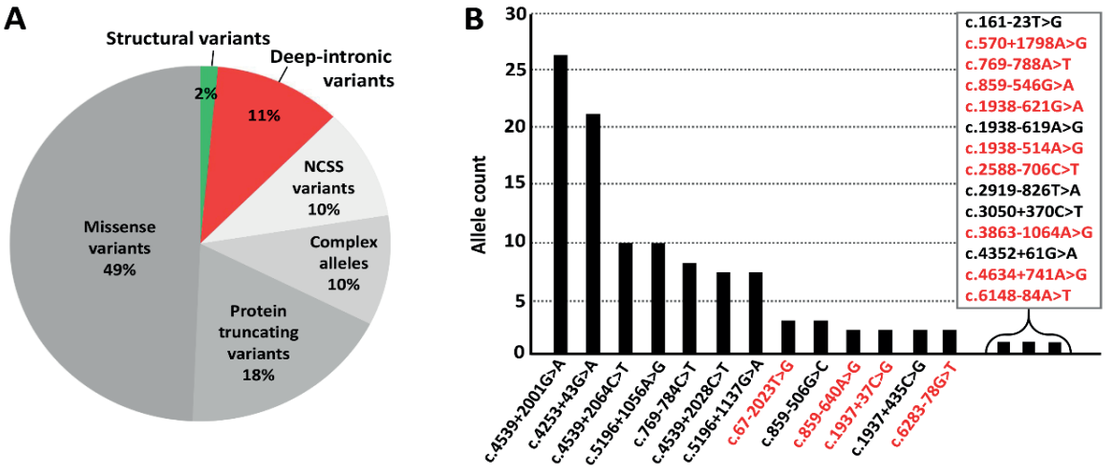


Figure 1. Distribution of different types of alleles and deep-intronic variants in *ABCA4*. A. The contribution of each type of variant or allele in bi-allelic, and mono-allelic cases except those carrying c.5603A>T, is represented. Protein truncating variants comprise nonsense, frameshift and canonical splice site variants. The 10% complex alleles only consist of combinations of missense variants, the most frequent of which were c.[1622T>C;3113C>T] (n=30; 27% of all complex alleles) and c.[4469G>A;5603A>T] (n=27; 25% of all complex alleles). They do not include the complex alleles which contain noncanonical splice site (NCSS) variants, deep-intronic variants or protein truncating variants, when present in *cis* with other variants. If these would have been included, 16% of the alleles would consist of complex alleles. **B.** Deep-intronic variant allele count in this study. Novel deep-intronic variants are highlighted in red. One hundred and seventeen causal deep-intronic variants were identified. The deep-intronic variants c.4539+2001G>A (n=26) and c.4253+43G>A (n=21) were found most frequently. Most of the novel deep-intronic variants were found in single STGD1 probands.

Splicing defects due to noncanonical splice site variants

The effect on splicing of nine NCSS variants was tested in nine wild-type splicing constructs previously described (**Figure S3** and **Chapter 2**). All of the nine tested novel NCSS variants showed a splicing defect when tested in HEK293T cells. Severity was assigned according to the percentage of remaining WT mRNA, as described previously

(**Chapter 2**). Five NCSS variants were deemed severe as they showed ≤30 of WT mRNA, three were considered to have a moderate effect with WT RNA present between >30 and ≤70% correct RNA and only one was mild as it showed >70% of WT RNA (**Table S9** and **Figure S4**).

Deep-intronic variants identification and functional characterization

Based on the defined selection criteria, 58 DI variants were selected for splice assays. To test their effects, 27 WT midigenes splicing constructs were employed, 23 of which were described previously, (**Chapter 2**) and four of which were new (**Figure S3**). Thirteen of 58 tested DI variants showed a splicing defect upon reverse transcription (RT)-PCR and Sanger validation (**Figure 2**). For the variants that did not show any splicing defect, RT-PCR results are shown in **Figure S5**.

Six of the novel DI variants, i.e., variants c.570+1798A>G, c.769-788A>T, c.859-640A>G, c.1938-514A>G, c.2588-706C>T and c.4634+741A>G, resulted in out of frame PE inclusions in the RNA and were deemed severe (**Figures 2** and **3**). Variants c.67-2023T>G and c.859-546G>A were classified to have a moderate effect as 33% and 36% of the WT RNA products were present, respectively. As predicted due to the presence of a downstream cryptic SDS, variant c.1937+37C>G led to an elongation of exon 13 by 36 nucleotides, which resulted in the introduction of a premature stop codon (p.Phe647*). Moreover, two intron 13 variants, c.1938-621G>A and c.1938-514A>G, showed a complex splicing pattern which led to the generation of two mutant transcripts each (**Figure 3A-C**). Each of these products contained a shared PE of 134 nt (PE1) as well as mutation-specific PEs, denoted PE2 (174 nt) or PE3 (109 nt) for c.1938-621G>A and c.1938-514A>G, respectively (**Figure S6**). For variant c.1938-621G>A only 7% of the total cDNA product showed PE inclusion whereas for c.1938-514A>G, 87% of the cDNA products were mutant. To investigate the nature of the PE1 insertions, we studied the exon 12-17 segment of the mRNA obtained from PPCs derived from a control person. As depicted in **Figure S7**, transcripts containing PE1 or PE1 and PE2, were identified when photoreceptor progenitor cells (PPCs) were grown under nonsense-mediated decay-suppressing conditions. The sum quantity of these two products was 2.9% of total mRNA suggesting that there are small amounts of PE insertions involving PE1 in the healthy retina.

Intriguingly, DI variant c.6148-84A>T showed four RNA splicing products, namely a normal spliced RNA, the skipping of exon 45, the insertion of a 221-nt PE (pe1a) coupled with the deletion of exon 44, and finally, the insertion of a 173-nt PE (pe1b) that consist of the same splice donor site as pe1a but a different splice acceptor site (**Figure 3D-F**). Finally, variant c.3863-1064A>G showed a complex splicing pattern compared to the WT and variant c.6283-78G>T led to the insertion of a 203-nt PE in

intron 45 (**Figure S6**). However, the exact boundaries of the presumed PE for variant c.3863-1064A>G could not yet be determined due to technical difficulties.

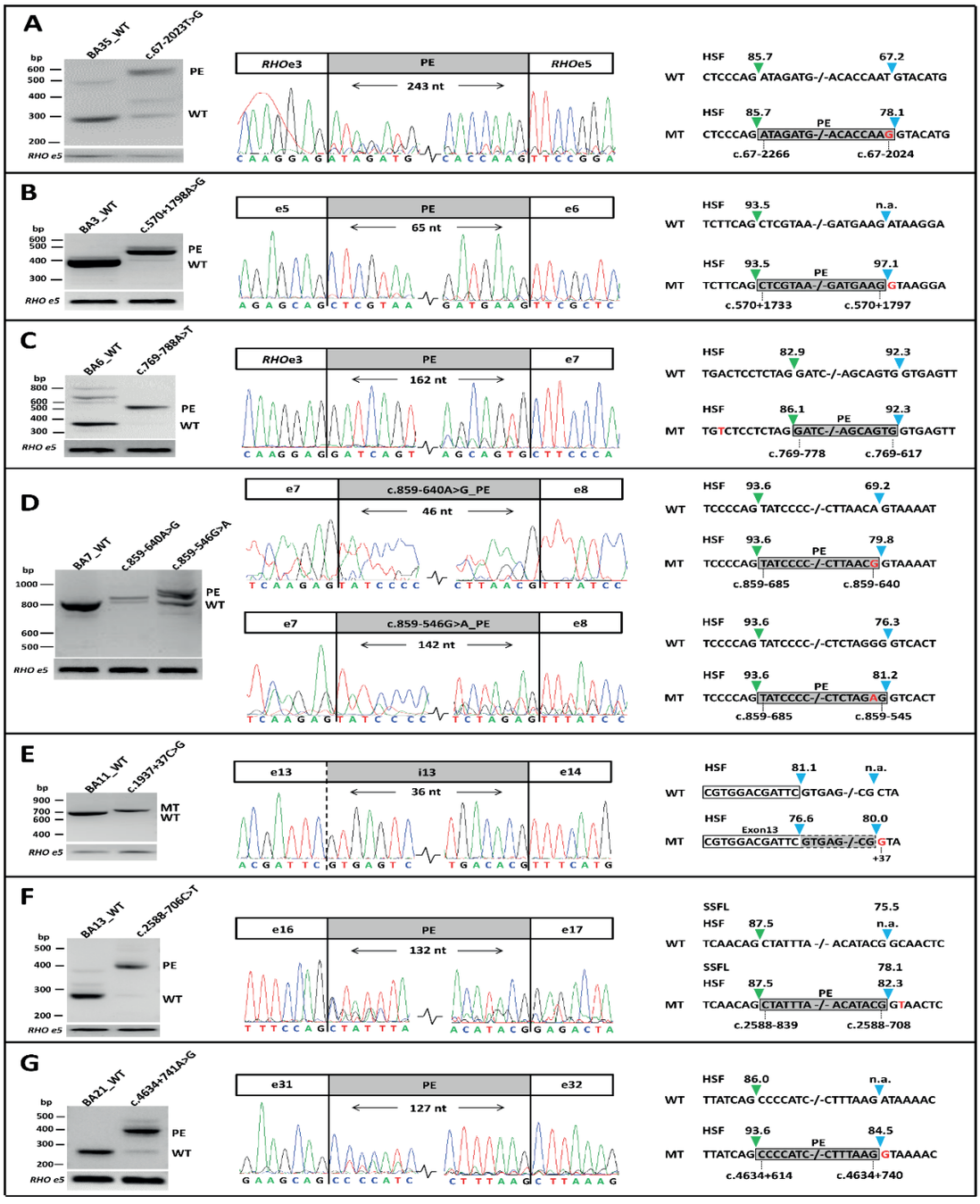


Figure 2. Novel splicing defects due to deep-intronic *ABCA4* variants. Wild-type (WT) and mutant (MT) midigenes were transfected in HEK293T cells and the extracted RNA was subjected to RT-PCR. Left panels show the *ABCA4*-specific RT-PCR products with Rhodopsin exon 5 (*RHO* e5) RT-PCR as a transfection efficiency control. In the middle panels, Sanger sequencing results of the RT-PCR products are given. At the right side pseudoexons (PE) and an exon elongation are depicted with splice site strength predictions for WT and MT, with green rectangles representing the splice acceptor sites and blue rectangles representing the

splice donor sites. Red highlighted nucleotides represent the mutations. Except for c.1937+37C>G (**D**), which resulted in a 36-nt exon 13 elongation, all deep-intronic variants lead to PEs. The intron 7 variants in part **C** result in partially overlapping PEs that share the same splice acceptor site at position c.859-685. HSF, human splicing finder; SSFL, slice site finder like; PE, pseudoexon; n.a., not applicable.

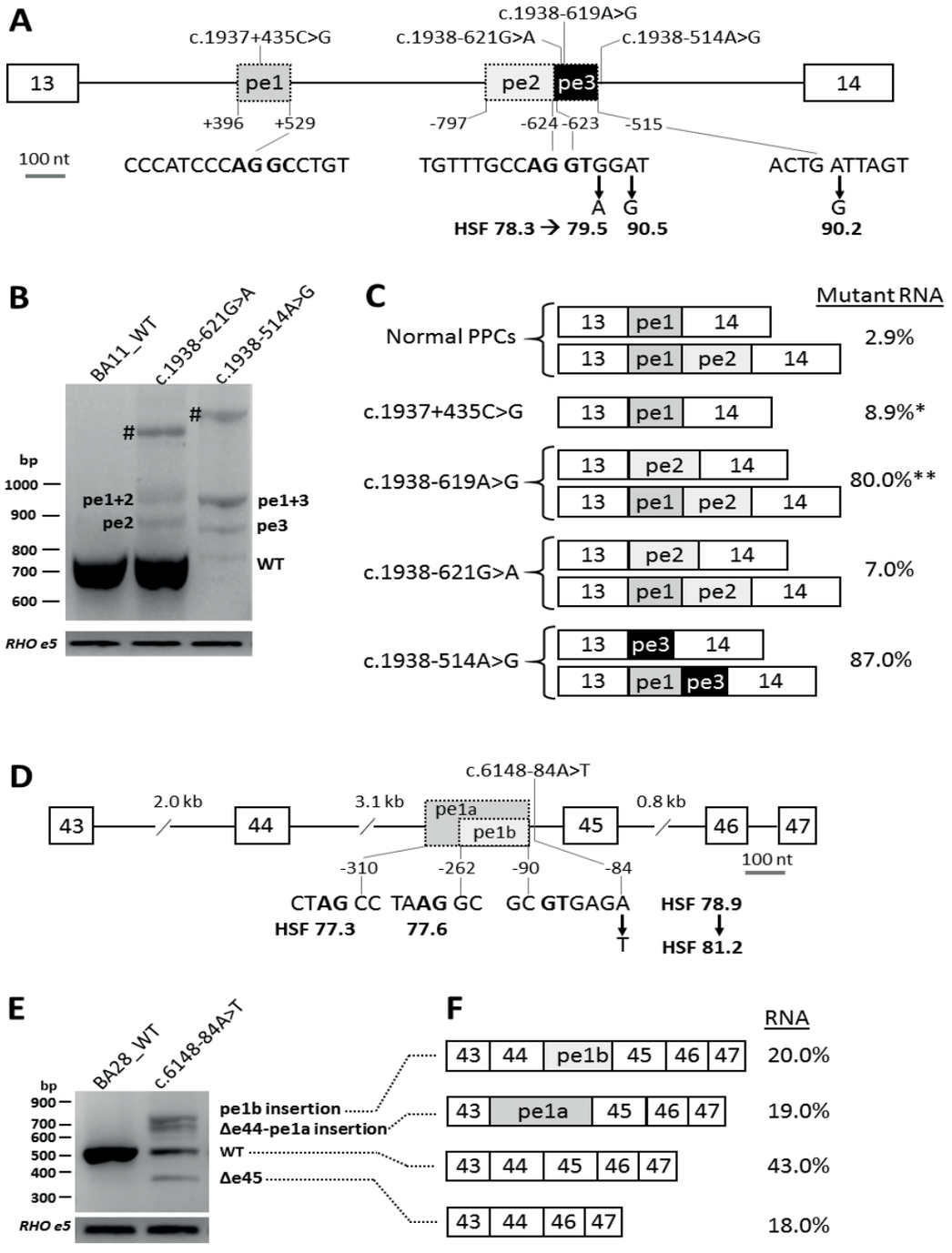


Figure 3. Splicing defects due to variants in *ABCA4* intron 13 and 44. A. Genomic structure of intron 13 containing three pseudoexons (PEs) due to four deep-intronic variants. PE2 and PE3 share a splice donor site (for PE2) and splice acceptor site (for PE3). Variants c.1938-621G>A and c.1938-619A>G strengthen the same cryptic splice donor site of PE2 slightly or strongly,

respectively, as based in the Human Splicing Finder (HSF). Variant c.1938-514A>G creates a new strong splice donor site of PE3. The canonical and putative canonical splice sequences are given in bold lettering. The first and last positions of the PEs are provided. **B.** Agarose gel analysis of RT-PCR products for intron 13 variants upon HEK293T cell splice assays. PE2 and PE3 were observed as single insertions, but also in combination with PE1. #Heteroduplex fragments of the lower bands. **C.** Schematic representation of all mutant transcripts identified upon RT-PCR in HEK293T cell splice assays and of PE1 and PE1/PE2 observed as naturally occurring PEs when analyzing photoreceptor progenitor cells (PPCs) derived from a healthy individual. Interestingly, PE1 was previously shown to be induced by variant c.1937+435C>G (**Chapter 3**) and also can be part of mutant transcripts, together with PE2 or PE3. This is surprising as it is located far upstream of the other causal variants. **Reported in **Chapter 6**. **D.** Variant c.6148-84A>T strengthens a splice donor site and results in PE1a or PE1b by employing upstream or downstream splice acceptor sites, respectively. These splice acceptor sites are comparable in predicted strength based on HSF. The canonical splice sequences are given in bold. **E.** Agarose gel analysis of RT-PCR products due to c.6148-84A>T. The largest fragment shows a 173-nt PE insertion between exons 44 and 45. The second largest band contains a 221-nt PE insertion (PE1a) and skipping of exon 44. The third-largest fragment represents the WT mRNA and the smallest fragment misses exon 45. The relative amounts of the products are listed at the right side.

Overall, 13 novel DI variants were found in 18 alleles. Next to the novel variants, 14 previously reported pathogenic DI variants (Braun et al. 2013; Albert et al. 2018; Bauwens et al. 2019; **Chapters 3** and **6**) were found in a total of 99 alleles, details of which are shown in **Figure 1B** and **Table S6**.

Identification of novel structural variants in STGD1 cases

Among 1,054 STGD and STGD-like patients analyzed, we identified 11 unique novel heterozygous SVs, all exon-spanning deletions, in 16 patients. The corresponding deletions encompass between 1 and 33 exons, ranging from 411 bp to 55.7 kb (**Figure 4**, **Tables S10-S15**). All deletions were found in a heterozygous state in single cases, except the smallest (c.699_768+341del), which encompassed the 5' part of exon 6 and 341 bp of intron 6, and was found in six unrelated patients of Spanish origin. Deletion breakpoints were determined employing genomic PCR and Sanger sequencing for 9 of the 11 deletions. Two deletion junctions (deletions 7 and 11) could not be amplified as the 3' breakpoints were located downstream of the gene beyond the regions targeted by smMIPs. Surprisingly, Sanger sequencing revealed two complex rearrangements as deletions 5 and 6 carried inverted fragments of 279 and 224 bp respectively, residing between large deletions. These small inversions could not be identified with the copy number variant (CNV) detection tool.

Microhomologies, repetitive elements and non-B DNA conformation at deletion breakpoints

The breakpoints of the deletions were subjected to bioinformatic analysis to find elements underlying their formation. The presence of microhomology, repetitive elements and non-B DNA conformations was investigated except for deletions 7 and 11 as exact boundaries could not be determined by Sanger sequencing. All other studied SVs presented microhomology at the breakpoint junctions, ranging in size from 1 to 6 bp (**Figure S8**), four of which presented short insertions (**Table S16**). In eight of 11 (72.7%) of the deletion breakpoints, a known repetitive element was observed, including seven non-Long Terminal Repeats (non-LTR) retrotransposons, among which there were one SINE and four LINEs, three DNA transposons from the *hAT* superfamily and two retrotransposons from the LTR superfamily. However, none of the breakpoints were part of a known element belonging to the same class and no *Alu* sequence was observed at the breakpoint junctions. Finally, the most prevalent non-B conformations observed among our breakpoints are Oligo(G)n tracts as 21 of these repeats were found in seven SVs (**Tables S16**-**S17**). Inverted repeats were observed in five breakpoint regions. No direct repeats or mirror repeats have been detected, excluding therefore triplex and slipped hairpin structure formation, respectively.

Uniparental isodisomy of chromosome 1

In STGD1 proband DNA14-33085, a causal homozygous DI variant, c.859-506G>C (p.[Phe287Thrfs*32,=]), was identified. Segregation analysis revealed this variant to be present in his unaffected father, but not in his unaffected mother. To test the possibility that the mother carried a deletion spanning this variant, we performed CNV analysis in the proband's *ABCA4* gene. No deletion was identified (**Table S15**, column AU) and no heterozygous SNPs were observed in or near *ABCA4* in the proband's DNA. To test whether the chromosome 1 of the father carrying the c.859-506G>C *ABCA4* variant was passed on to the proband as two copies (uniparental isodisomy, UPD), whole exome sequencing was conducted for the proband's DNA. As shown in **Figure S9**, chromosome 1 of the proband carries only homozygous SNPs, strongly suggesting the occurrence of UPD.

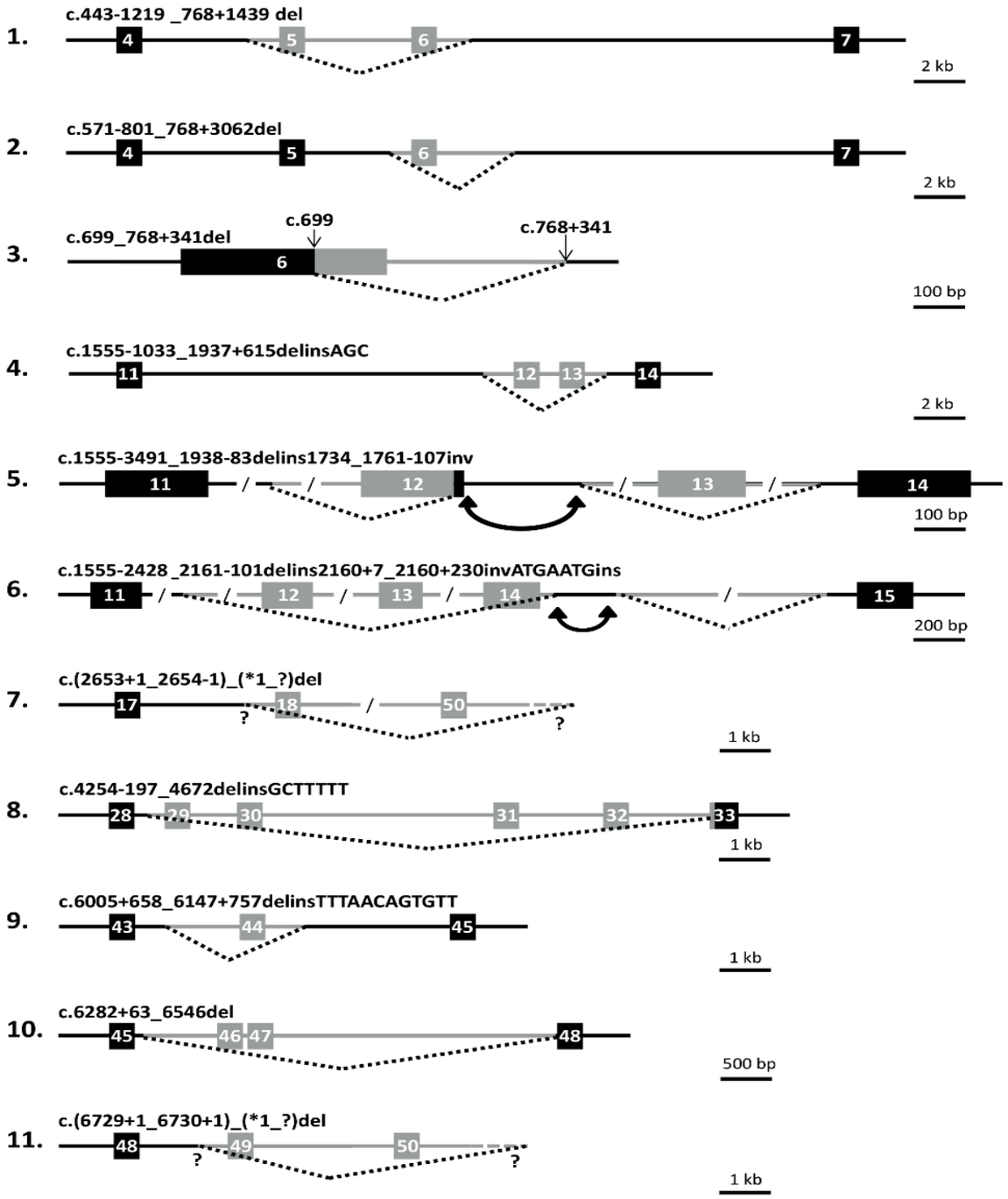


Figure 4. Novel heterozygous structural variants in *ABCA4*. Schematic representation of the 11 structural variants identified. Exons are represented as boxes, black when they are not deleted and grey when they are deleted. Introns are represented as continuous lines, whereas stippled lines depict the deleted regions. Question marks denote that the exact location of breakpoints were unknown. Inverted double arrows represent inverted sequences.

Discussion

Employing 3,866 smMIPs, 97.4% of the 128-kb *ABCA4* gene could be sequenced robustly in 1,054 genetically unsolved probands with a STGD or a STGD-like phenotype. In this way, 448 (42.5%) of the probands could be genetically solved. We not only identified nine novel NCSS variants and 13 novel DI variants, but also facilitated by a reproducible high average read coverage, 11 novel heterozygous SVs. The large set-up of this study allowed us to provide a 'landscape' overview of the different types of mutations causing STGD1. As depicted in **Figure 1A**, we can appreciate that DI variants constitute a significant cause of STGD1, i.e., 11.7% of the alleles in bi-allelic cases, identified in 22.5% of bi-allelic probands. Deletions constitute 1.8% of alleles and were found in 3.5% of bi-allelic cases. Seven probands carried two DI variants or one DI variant and one SV. Taken together, 'dark matter' alleles were found in 113/448 (25.2%) bi-allelic STGD1 probands. Together, these results strongly argue for a complete sequence analysis of the *ABCA4* gene to fully appreciate its mutational landscape.

Complex splicing defects due to intron 13 and 44 variants

Interestingly, the two intron 13 DI variants, i.e., c.1938-621G>A and c.1938-514G>A, were in close vicinity of two previously described variants, c.1937+435C>G (**Chapter 3**) and c.1938-619A>G (Zernant et al. 2014; **Chapter 6**). As shown in **Figure 3A-C** the PE resulting from c.1937-514A>G (PE3) is located adjacent to PE2 as they share a dual SAS/SDS (**Figure 3A-C**). The involvement of PE1, located 491, 493 and 775 nt upstream of variants c.1937-621G>A, c.1937-619A>G and c.1937-514A>G, respectively, is very surprising. Control PPCs also show a small percentage (2.9%) of mRNAs containing PE1 or PE1-PE2. Interestingly, the SDS of PE1 also can be employed as a SAS which, in theory could render this intronic SAS/SDS a target for recursive splicing (Hafez and Hausner 2015). Together, these findings suggest that there is a 'natural sensitivity' for PE1 to be recognized as a PE even if the splicing defect is located far downstream. To our knowledge, this has not been described thus far. Intron 44 variant c.6148-84A>T interestingly resulted in three abnormal splicing products involving different PE insertions with or without flanking exon 44 or 45 deletions. Follow-up studies employing patient-derived retinal-like cells are required to validate these complex splicing patterns.

In **Table S18**, we listed all published 353 DI variant alleles (Braun et al. 2013; Zernant et al. 2014; Bauwens et al. 2015; Bax et al. 2015; Schulz et al. 2017; Albert et al. 2018; Zernant et al. 2018; Bauwens et al. 2019; Nassisi et al. 2019; **Chapters 3, 4** and **6**). The three most frequent are c.4253+43G>A (n=100), c.4539+2001G>A (n=64) and

c.5196+1137G>A (n=47). For some DI variants, the splicing defects in HEK293T cells or patient-derived PPCs are very small (c.769-784C>T, c.1937+435G>C, c.1937-621G>A) (Runhart et al. 2019; **Chapters 3** and **7**) or smaller than expected (c.4539+2001G>A, c.4539+2028C>T). We hypothesize that retina-specific splicing motifs and factors play a role which are largely missing (HEK293T cells) or underrepresented (PPCs) compared to the normal retina.

Current state of knowledge on structural variants in *ABCA4*

In this study, 11 unique SVs with sizes ranging from 411 bp to 55.7 kb, were readily identified employing an easy-to-use visual detection tool taking advantage of the high number of reads obtained from smMIPs based-sequencing. Although this tool needs further automation to increase its performance for the detection of smaller deletions or duplications, it demonstrated its efficiency for deletions as small as 411 bp. To our knowledge, 47 different SVs have been identified in STGD1 patients (**Table S19**), 25 of which have been published elsewhere. Forty are deletions, ranging in size from 23 bp to complete deletion of the *ABCA4* gene. There are six duplications, ranging from 24 bp to 26 kb, two indels and one small insertion of 24 bp. As shown in **Figure S10**, these SVs are spread over the entire gene. All SVs are rare, except for a 23-bp deletion affecting the splicing of exons 28 and 29 in 15 Israeli probands, as well as deletions spanning exons 20-22 and exon 6, both found in 6 probands, in Belgium/Germany/Netherlands and from Iberic origin, respectively, suggesting founder effects.

This genomic instability could be explained by the local genomic architecture (the presence of microhomology, repetitive elements, sequences forming non-B DNA conformations and sequence motifs), leading to genomic rearrangements by impairing the replication process. For example, a microhomology of 1–4 bp may facilitate non-homologous end-joining (Lieber 2010) and longer microhomologies of between 5-25 bp may favor MMEJ (McVey and Lee 2008). The assessment of the local architecture of deletions identified in this study lead us to rule out the NAHR hypothesis (as no *Alu* sequence or L1 at any breakpoint was observed) and to propose the NHEJ or replication slippage models as the main implicated mechanisms (**Table S16** and **Figure S8**). Indeed, the presence of microhomologies <5 bp in most of the junctions, and of scars characterized by insertion of several random nucleotides, could be a signature for NHEJ. Alternatively, several examples of an impaired replication fork have been noted that supports the replicative-based repair model. Indeed, despite the absence of repetitive elements of the same class at both sides of the breakpoints, their presence may initiate the formation of secondary structures, as repetitive elements could be more difficult to replicate, leading to an increased chance of replication fork stalling or

collapsing (Vissers et al. 2009). Finally, Oligo(G)n tracts displayed a significant overrepresentation in the breakpoint regions. Such structures can induce tetraplex formation (Bacolla and Wells 2004) and could also trigger rearrangement.

Uniparental isodisomy chromosome 1

UPD was found in one STGD1 case in this study, which represents the third STGD1 case showing UPD thus far reported (Fingert et al. 2006; Riveiro-Alvarez et al. 2007). UPD is a rare event, with an estimated occurrence of 1 in 5,000 or even less individuals (Liehr 2010). UPD was also described in six other inherited retinal dystrophy patients in which chromosomes 1, 2 and 6 were implicated (Thompson et al. 2000; Rivolta et al. 2002; Thompson et al. 2002; Wiszniewski et al. 2007; Roosing et al. 2013; Souzeau et al. 2018). We cannot exclude that there are additional UPD cases in our cohort as segregation analysis was not performed for all homozygous cases. Our finding stresses the importance of segregation analysis in the parents' DNAs as the recurrence risk for future offspring is very low in UPD families.

Missing heritability

In 174/1,054 (16.5%) of probands, we identified only one (likely) causal allele. In view of the high carrier frequency of *ABCA4* variants in the general population, estimated to be ~5%, (Maugeri et al. 1999; Cornelis et al. 2017) about one-third of these mono-allelic cases may be explained in this way. This may even be higher as we intentionally recruited mono-allelic STGD and STGD-like probands for this study. Some causal variants may have escaped our attention. First, we have not focused on variants affecting transcription regulation. Thus far, there is limited evidence for *ABCA4* variants affecting transcription (Bauwens et al. 2019), but the reported putative regulatory variants were not found in this study. As *in silico* tools (Alamut algorithms, SpliceAI) (Jaganathan et al. 2019), may not predict retina-specific splicing defects, we may have missed some causal variants. Also, smMIPs-based sequencing may miss heterozygous deletions smaller than ~400 bp and will not detect insertions or inversions larger than ~40 bp. In addition, more refined functional tests of coding and non-coding *ABCA4* variants are needed to understand the full genetic landscape of STGD1.

The major advantages of smMIPs-based *ABCA4* sequencing compared to WGS are that: i) it is at least an order of magnitude cheaper than WGS, ii) results in much smaller data storage, and iii) requires no separate informed consent regarding incidental findings. Disadvantages of smMIPs are that: i) it is restricted to one or a few genes if including introns, ii) it is more cost-effective when large series are analyzed, iii) the analysis is suitable for the detection of CNVs but not for inversions and

insertions, and finally, iv) the sequencing procedure and variant calling requires a specialized set-up.

In our study a significant fraction of probands carried one (likely) causal variant or c.5603A>T as a single allele (239; 22.7%) or no causal variant (364; 34.5%). A more comprehensive smMIPs-based screening platform for these STGD-like cases would likely require the sequence analysis of an additional ~80 genes associated with inherited central vision defects. As shown in this study, smMIPs-based analysis of the complete sequence(s) of one or a few genes implicated in clinically well-defined human diseases may allow the (re)analysis of hundreds to thousands of samples, in particular by targeting cohorts in developing countries in which low-cost analysis is crucial. A similar approach can be applied to all other frequent monogenic disorders to find missing variants in non-coding regions to provide a genetic diagnosis.

In conclusion, comprehensive sequence analysis of *ABCA4* in 1,054 unsolved STGD and STGD-like probands, splice assays in HEK293T cells and SV analysis, resulted in the identification of 'dark matter' variants in 25% of bi-allelic STGD1 probands. Novel complex types of splicing defects were identified for intron 13 and 44 variants. Together with published causal DI variants and SVs, a detailed genomic and transcriptomic landscape of *ABCA4*-associated STGD1 was thereby established.

Acknowledgements

We thank Ellen Blokland, Duaa Elmelik, Emeline Gorecki, Marlie Jacobs-Camps, Charlene Piriou, Mariateresa Pizzo, Saskia van der Velde-Visser for technical assistance. We thank Béatrice Bocquet, Dominique Bonneau, Krystyna H. Chrzanowska, Hélene Dollfus, Isabelle Drumare, Monika Heusipp, Takeshi Iwata, Beata Kocyła-Karczmarewicz, Atsushi Mizota, Nobuhisa Nao-i, Adrien Pagin, Valérie Pelletier, Rafal Ploski, Agnieszka Rafalska, Rosa Riveiro, Malgorzata Rydzanicz, Blanca Garcia Sandoval, Kei Shinoda, Francesco Testa, Kazushige Tsunoda, Shinji Ueno and Catherine Vincent-Delorme for their cooperation and ascertaining STGD1 cases. We thank Rolph Pfundt for his assistance in WES data analysis. We thank ERN-EYE and ERDC networks, the Japan Eye Genetics Consortium, and the East Asian Inherited Retinal Disease Society.

Supplemental data

Supplemental data includes only Figures S1, S2 and S10, and Table S10. Complete supplemental files can be found at https://www.biorxiv.org/content/10.1101/817767v1.supplementary-material

Supplemental materials and methods

Control individuals and STGD1 probands in pilot and main sequencing studies

To assess the performance of the smMIPs pool, a pilot study was conducted using five control individuals and 15 STGD1 cases (**Table S1**). To further evaluate the sensitivity of the smMIPs-based sequencing method in subsequent sequencing studies, 138 bi-allelic STGD1 samples were included in the study, 123 of which carried *ABCA4* SNVs and 15 of which carried a deletion or duplication in *trans* with other alleles (**Table S4** and **S5**).

Unsolved and partially solved STGD1 and macular dystrophy probands

Twenty-one international and four national centers ascertained 1,054 genetically unsolved or partially solved probands with a STGD or STGD-like phenotype. Two hundred twenty-one of these cases were not previously molecularly studied (**Table S21**). We discerned two patient groups. The first patient group consisted of 993 genetically unsolved probands that carried one (n=345) or no (n=648) *ABCA4* pathogenic allele. For two subjects, DNA was not available and both parents of the probands were studied, assuming autosomal recessive inheritance. The second patient group consisted of 61 'partially solved' probands, carrying the c.5603A>T (p.Asn1868Ile) variant in *trans* with other alleles. This last group was also investigated as it was suspected that there could be unidentified deep-intronic variants in *cis* with c.5603A>T, as the penetrance of c.5603A>T, when in *trans* with a severe *ABCA4* variant, was ~5% in the population (Cremers et al. 2018; Runhart et al. 2018). Samples were collected according to the tenets of the Declaration of Helsinki and written informed consent was obtained for all patients participating in the study. DNA samples were quantified, diluted to 15 ng/µl and visually inspected after agarose gel electrophoresis.

smMIPs design

For the targeted sequencing of the *ABCA4* gene, as well as 23.9 kb up- and 14.9 kb downstream sequences, 3,866 smMIPs were designed using MIPgen pipeline, as described previously (Boyle et al. 2014). Four hundred and ninety-six of the 3,866 smMIPs spanned known single nucleotide polymorphisms (SNPs) at the extension or ligation arm. Therefore, 3,370 smMIPs effectively capture 110-nt regions. Details of the smMIPs can be found in **Table S22**. All smMIPs were designed in a way that each nucleotide position in principle is targeted by two different smMIPs to increase sensitivity and avoid dropout of alleles due to variants in the smMIP arms.

Sequencing set-up

First, a pilot study was conducted to assess the performance of smMIPs and coverage per sample. Twenty DNA samples were sequenced on a NextSeq500 platform using an Illumina mid-output sequencing kit (~130 million reads). Thereafter, 1,192 STGD and STGD-like samples were sequenced using 100 ng DNA per sample. Sequencing libraries were prepared in a fully automated manner as described previously (**Chapter 4**). To sequence these 1,192 cases, six sequencing runs were performed on an Illumina NextSeq500 platform using high-output kits (~400 million reads). The DNA of 21 patients was sequenced twice. In the first attempt, not enough DNA was used which resulted in less than 100,000 reads per sample, whereas the average number of reads for all samples was 1,355,833. These 21 samples were successfully sequenced in the second run with between 641,635 and 2,074,344 reads per sample.

Data processing

Raw sequencing data (FASTQ) files were used for mapping and variant annotation by an in-house bioinformatics pipeline. Raw reads were aligned to the reference genome (human genome build GRCh37/hg19). Variant calling and annotation were performed as described previously (**Chapter 4**). Variants were annotated with predicted effects, gene component information as well as frequency information from various population databases.

Data analysis and variant classification

To select the causative variants, we applied several filters. First, we retained variants with allele frequencies (AFs) <0.005 in different population databases such as dbSNP, the Genome Aggregation Database (gnomAD; http://gnomad.broadinstitute.org/) and Genome of the Netherlands (GoNL). We then applied a filter based on a 'quality score' of >200, based on the average number of reads targeting each genomic position in a single run and other standard filters based on the position of SNVs (e.g., exons, canonical splice sites and known NCSS variants). A number of previously listed causal variants with relatively high AFs such as c.2588G>G, c.5461-10T>C, c.5603A>T and c.5882G>A (AFs in gnomAD: 0.00430, 0.00022, 0.04219 and 0.00456 respectively) were selected separately (Cremers et al. 1998; Cornelis et al. 2017; Schulz et al. 2017; Zernant et al. 2017). To assess the variants, *in silico* analysis was performed using online pathogenicity assessment tools such as Sorting Intolerant From Tolerant (SIFT; https://sift.bii.a-star.edu.sg) (Kumar et al. 2009), Polymorphism Phenotyping v2 (PolyPhen-2) (Adzhubei et al. 2010), and Mutation Taster (Schwarz et al. 2014). Among others, Grantham and PhyloP scores were also assessed to predict the physiochemical

effect of nucleotide changes on ABCA4 protein and conservation of nucleotide positions, respectively.

For the selection of known coding and non-coding causal variants, *ABCA4* variant databases, such as the Leiden Open Variation database (LOVD; www.lovd.nl/*ABCA4*) (Cornelis et al. 2017), and the Human Genome Mutation Database (HGMD; http://www.hgmd.cf.ac.uk/ac/index.php), were used. We supplemented the list of SNVs and small indels with additional variants reported in the literature and employed a script to identify these in an automated manner. Novel candidate variants were selected based on AFs and *in silico* predictions. Thereafter, variants were classified according to the guidelines of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) (Richards et al. 2015).

Similarly, we detected previously reported causal splice variants (NCSS variants and DI variants) by listing all 165 variants previously tested using *in vitro* splice assays (Braun et al. 2013; Bauwens et al. 2015; Sangermano et al. 2016; Bauwens et al. 2019; **Chapters 2, 3, 4** and **6**), or RT-PCR analysis of patient-derived photoreceptor progenitor cells (PPCs) (Sangermano et al. 2016; Albert et al. 2018).

Candidate splice defect variant selection

Intronic variants located at positions -14 to -3 upstream of exons and at positions +3 to +6 downstream of exons were considered NCSS variants, and variants outside these positions were considered DI variants. Exonic *ABCA4* NCSS variants can also be found within the first or the last two nucleotides of an exon (**Chapter 2**). To select novel NCSS and DI variants for *in vitro* splice assays, the following selection criteria were used: i) minor allele frequency <0.005 in general population databases such as dbSNP, gnomAD and GoNL, ii) a relative difference in strength of wild-type (WT) and mutant splice site scores by at least 75% obtained by *in silico* analysis performed by using five different algorithms including SpliceSiteFinder-like (SSFL), MaxEntScan, NNSPLICE, GeneSplicer, Human Splicing Finder (HSF) via Alamut Visual software version 2.7 (Biosoftware, 2014; Interactive Biosoftware, Rouen, France; www.interactivebiosoftware) (Reese et al. 1997; Pertea et al. 2001; Cartegni et al. 2003; Yeo and Burge 2004; Desmet et al. 2009), iii) additional splice site score comparison between WT and mutant was made by using the SpliceAI program, which unlike other programs searches for altered splice sites up to 5 kb upstream and 5 kb downstream of the variant (Jaganathan et al. 2019), and iv) enrichment of variants' allele frequencies in the patient cohort compared to general population databases such as gnomAD and GoNL. A list of all the selected candidates for *in vitro* splice assays and details of *in silico* analysis are provided in **Table S23**.

Midigene-based splice assay

The splice effect of nine NCSS variants and 58 deep-intronic variants was assessed by midigene-based splice assays employing WT constructs described previously (**Chapter 2**) as well as four newly designed midi- and minigenes, namely BA32 (exons 35-38), BA33 (hg19: intron 1, g.94,585,917-94,584,820), BA34 (hg19: intron 1, g.94,582,386-94,581,151) and BA35 (hg19: intron 1, g.94,581,151-94,579,996) (**Figure S3** and **Table S24**). All constructs contained *ABCA4* genomic sequences amplified from DNA of the bacterial artificial chromosome clone CH17-325O16 (hg19: insert g.94,434,639-94,670,492). The WT clones served as templates to generate mutant constructs by site-directed mutagenesis, which subsequently were validated by Sanger sequencing. Details of mutagenesis primers and Sanger sequencing primers are given in **Table S25**. WT and mutant constructs were transfected in parallel in HEK293T cells. Total RNA was extracted 48 hours post-transfection and subjected to RT-PCR using *ABCA4* exonic primers when possible, as previously described (**Chapter 2**). Rhodopsin exon 5 amplification was used as transfection and loading control. RT-PCR primer sequences can be found in the **Table S26**.

Copy number variant analysis

To assess the accuracy of the CNV detection tool, 13 different CNVs spread throughout the *ABCA4* gene were analyzed in a single run in 15 control DNA samples (**Tables S4** and **S5**). Since 1,192 probands were sequenced in six NextSeq500 runs, CNV analysis was performed for each run independently (**Tables S10-S15**). CNVs were confirmed by PCR amplification using *Taq* DNA polymerase (Thermofisher Scientific, Waltham, MA, USA) or Q5 High-Fidelity DNA polymerase (New England Biolabs, Ipswich, MA, USA) with an extension time of 45s. Primer details are shown in **Table S27**.

Identification of repetitive elements and presence of microhomology at deletion breakpoints

For a better understanding, breakpoint (BP) numbers were assigned according to their 5'-3' localization. To characterize the flanking regions of the deletion breakpoints and to identify the underlying mechanisms of the SV formation, 150-bp segments flanking each breakpoint were analyzed and submitted to RepeatMasker (http://www.repeatmasker.org) and Censor (https://www.girinst.org/censor/) online software tools. The presence of microhomology at the breakpoints was assessed by a multiple sequence alignment between the junction fragment and the proximal and distal breakpoint regions, using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). To investigate the possible involvement of non-B DNA conformation elements on SV formation, they were searched within the

same 150-bp regions surrounding the breakpoint. DNA sequences leading to non-B DNA conformations were assessed with two different online software tools. Quadruplex forming G-Rich Sequences (QGRS) Mapper (http://bioinformatics.ramapo.edu/QGRS/analyze.php) allowed the identification of (Oligo)n tracts forming tetraplex structures and non-B DNA motif search tool (nBMST) (https://nonb-abcc.ncifcrf.gov/apps/nBMST/default) provided location for direct repeats and slipped motifs, G quadruplex forming repeats, inverted repeats and cruciform motifs, mirror repeats and triplex motifs, Z-DNA motifs and short tandem repeats.

Semi-quantification of RT-PCR products

To quantify the ratios between correct and aberrant RT-PCR products, densitometric analysis was performed using Image J software after gel electrophoresis as described elsewhere (Schneider et al. 2012). Only peaks corresponding to *ABCA4* bands were taken into account (**Table S28**).

Uniparental disomy detection

In one STGD1 case (DNA14-33085), we previously identified an apparent homozygous causal DI variant (c.859-506G>C) in *ABCA4*, but only one of the parents was shown to carry this variant. Therefore, a deletion search and segregation analysis were performed. Furthermore, to test the hypothesis of uniparental disomy (UPD), a chromosome 1 haplotype was constructed using whole exome sequencing.

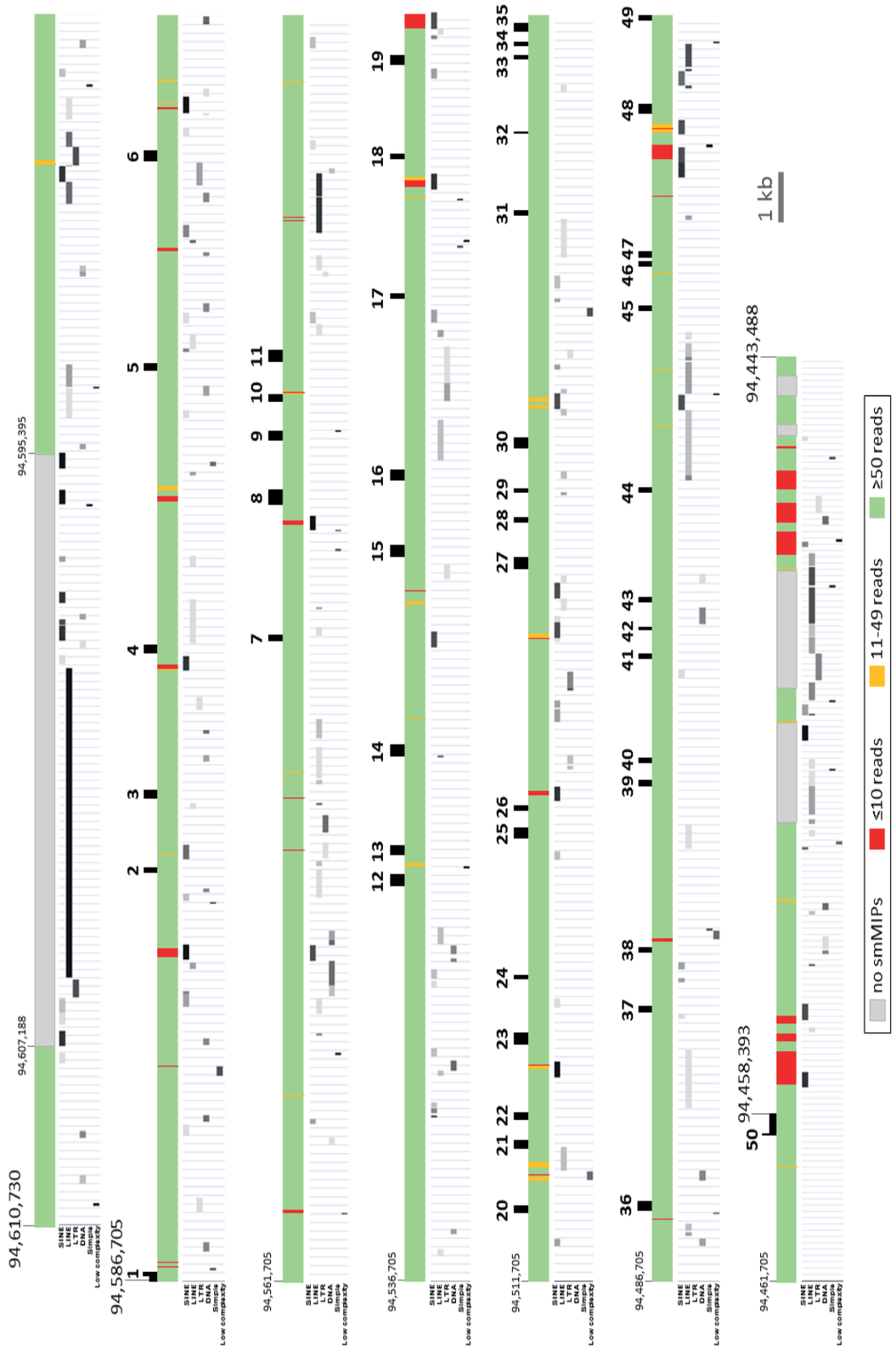

94,610,730
94,607,188
94,595,395
94,586,705
94,561,705
94,536,705
94,511,705
94,486,705
94,461,705
94,458,393
94,443,488
1 kb
no smMIPs
≤10 reads
11-49 reads
≥50 reads

Figure S1. *ABCA4* exon-intron structure and coverage upon smMIPs-based sequence analysis. The average number of reads covering each nucleotide position was calculated in runs one through five, which contained sequence data of 1,022 persons. Regions that were refractory to smMIPs design due to their repetitive nature are depicted as grey boxes. Based on an average read number of 377 for regions that were sequenced using smMIPs, we color-coded poorly covered (≤10 reads, red boxes), moderately covered (11-49 reads, orange boxes), and well covered regions (≥50 reads, green regions). From the 128,366 bp of *ABCA4*, 1,980 bp (1.5%) were not or poorly covered, 1,410 bp (1.1%) were moderately covered, and 124,976 bp (97.4%) were well covered. Although *ABCA4* introns carry several small repetitive elements, they only had a small effect on smMIPs design. Both up- and downstream of *ABCA4*, several larger repeats are present, which resulted in the absence of - or poorly performing - smMIPs. Genomic positions are according to hg19 with 25-kb segments depicted for the *ABCA4* gene transcriptional unit.

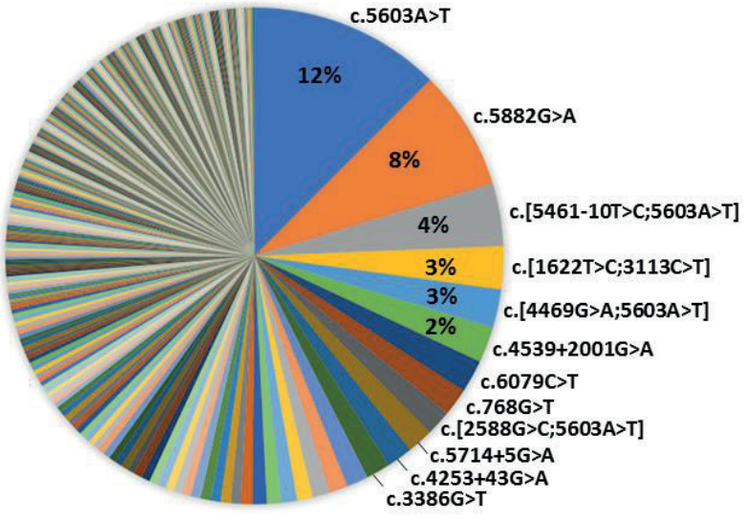


Figure S2. Percentage of different single nucleotide variants in bi-allelic Stargardt probands.

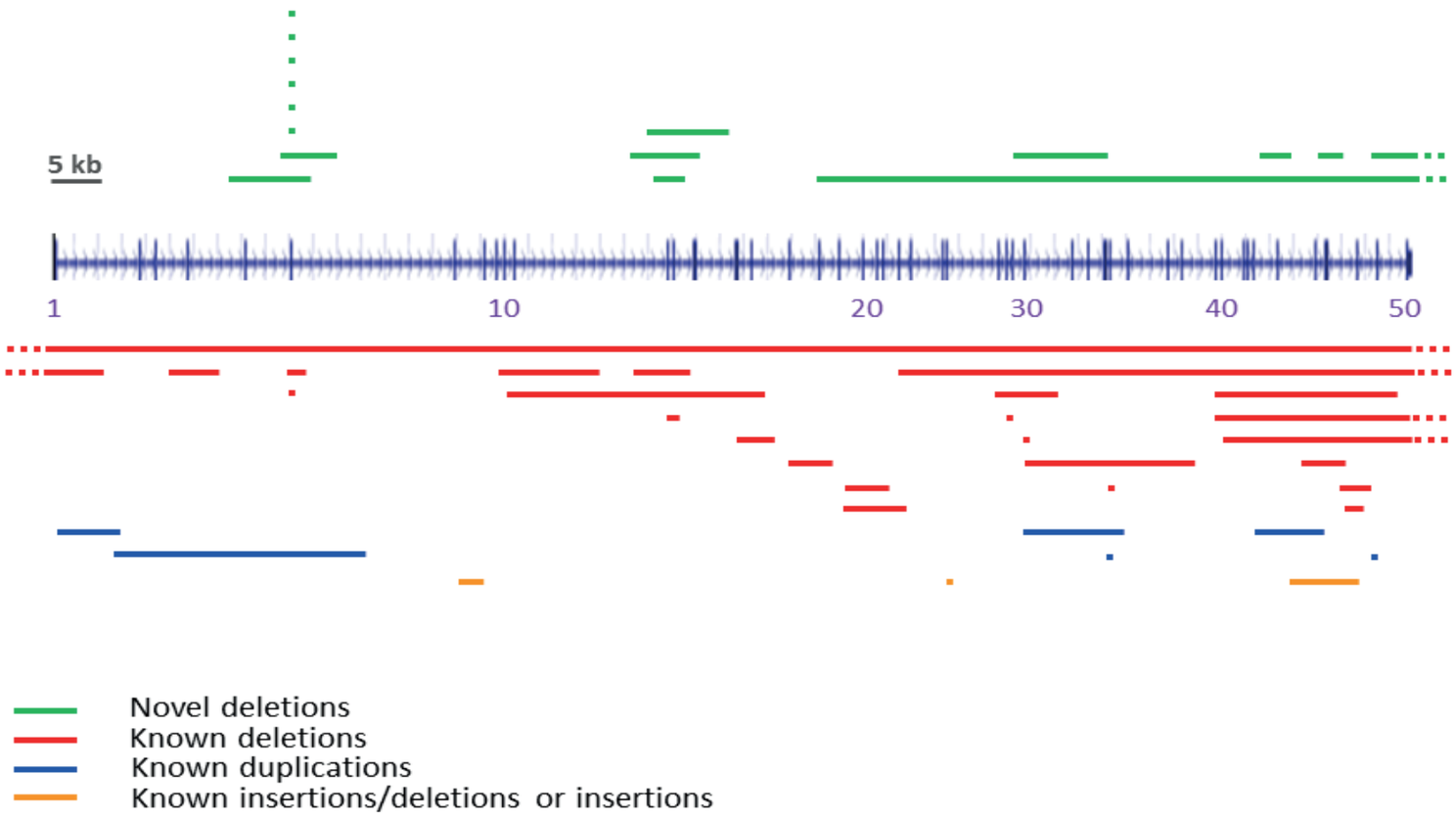


Figure S10. Schematic overview of all known causal Stargardt disease-associated structural variants in *ABCA4*.

Table S9. Novel noncanonical splice site variants in *ABCA4*

DNA variant	RNA effect	Protein effect	Splice defect	% WT mRNA	Proposed severity
c.1554+3A>T	r.[=,1357_1554del]	p.[=,Asp453_Glu518del]	Exon 11 skipping	51	Moderate
c.2654-8T>G#	r.[2653_2654ins2654-40_2654-1,=]	p.[Gly863Valfs*47,=]	Exon 18 elongation	13	Severe
c.3862G>A	r.[=,3863g>a,3814_3862del]	p.[=,Gly1288Ser,Ile1272Valfs*101]	Exon 26 skipping	69	Moderate
c.4129-3C>T	r.[=,3864_4128del,4129_4253del,3864_4253del]	p.[=,Ile1377Hisfs*3,Gly1288Aspfs*45]	Exon 28, 27 and 27 & 28 deletion	76	Mild
c.4540-8T>A	r.4539_4540ins4540-6_4540-1	p.(Gln1513insProGln)	Exon 31 elongation	0	Severe
c.4848+3A>G	r.[4774_4848del,=]	p.[Gly1592_Lys1616del,=]	Exon 34 skipping	10	Severe
c.5461-6T>G	r.5461_5714del	p.(Thr1821Aspfs*6)	Exon 39 & 40 skipping	0	Severe
c.5898+5G>A	r.[5898_5899ins_5899+1_5890-1,5898_5899ins5899+1_5899+170,=]	p.[Cys1967Valfs*24,=]	Exon 42 elongation and intron 42 retention	48	Moderate
c.6386+3A>G	r.[6386_6387ins6386+1_6387-1,6340_6386del,=]	p.[Ser2129Serfs*29, Val2114Hisfs*4,=]	Exon 46 elongation and 47-nt reduction	26	Severe

For the novel noncanonical splice site variants identified in this study, data are provided regarding their cDNA positions (hg19), RNA and protein effects. Consequences of the splice defects on exons are indicated. Pathogenicity of the variants was assigned according to the percentage of wild-type mRNA obtained by semi-quantification of both wild-type and mutant transcripts as described by **Chapter 2**, i.e., **severe** if the correctly splice RNA is present in <30% of the products, **moderate** between 30% and 70%, and **mild** >70%). WT, wild-type; # variant already reported, but not tested by Zernant et al, 2014.

References

Adzhubei IA, Schmidt S et al. 2010. A method and server for predicting damaging missense mutations. *Nat Methods* **7**: 248-249.

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

Allocca M, Doria M et al. 2008. Serotype-dependent packaging of large genes in adeno-associated viral vectors results in effective gene delivery in mice. *J Clin Invest* **118**: 1955-1964.

Bacolla A, Wells RD. 2004. Non-B DNA conformations, genomic rearrangements, and human disease. *J Biol Chem* **279**: 47411-47414.

Bauwens M, De Zaeytijd J et al. 2015. An augmented ABCA4 screen targeting noncoding regions reveals a deep intronic founder variant in Belgian Stargardt patients. *Hum Mutat* **36**: 39-42.

Bauwens M, Garanto A et al. 2019. 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* doi:10.1038/s41436-018-0420-y.

Bax NM, Sangermano R et al. 2015. Heterozygous deep-intronic variants and deletions in ABCA4 in persons with retinal dystrophies and one exonic ABCA4 variant. *Hum Mutat* **36**: 43-47.

Blacharski PA, Newsome DA. 1988. Bilateral macular holes after Nd:YAG laser posterior capsulotomy. *American journal of ophthalmology* **105**: 417-418.

Boyle EA, O'Roak BJ et al. 2014. MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. *Bioinformatics* **30**: 2670-2672.

Braun TA, Mullins RF et al. 2013. Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. *Hum Mol Genet* **22**: 5136-5145.

Carss KJ, Arno G et al. 2017. Comprehensive Rare Variant Analysis via Whole-Genome Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease. *Am J Hum Genet* **100**: 75-90.

Cartegni L, Wang J et al. 2003. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* **31**: 3568-3571.

Charbel Issa P, Barnard AR et al. 2015. Rescue of the Stargardt phenotype in Abca4 knockout mice through inhibition of vitamin A dimerization. *Proc Natl Acad Sci U S A* **112**: 8415-8420.

Cornelis SS, Bax NM et al. 2017. In Silico Functional Meta-Analysis of 5,962 ABCA4 Variants in 3,928 Retinal Dystrophy Cases. *Hum Mutat* **38**: 400-408.

Cremers FP, van de Pol DJ 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.

Cremers FPM, Cornelis SS, Runhart EH, Astuti GDN. 2018. Author Response: Penetrance of the ABCA4 p.Asn1868Ile Allele in Stargardt Disease. *Invest Ophthalmol Vis Sci* **59**: 5566-5568.

Desmet FO, Hamroun D et al. 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* **37**: e67.

Fadaie Z, Khan M et al. 2019. Identification of splice defects due to noncanonical splice site or deep-intronic variants in ABCA4. *Hum Mutat* doi:10.1002/humu.23890.

Fingert JH, Eliason DA et al. 2006. Case of Stargardt disease caused by uniparental isodisomy. *Arch Ophthalmol* **124**: 744-745.

Hafez M, Hausner G. 2015. Convergent evolution of twintron-like configurations: One is never enough. *RNA Biol* **12**: 1275-1288.

Jaakson K, Zernant J et al. 2003. Genotyping microarray (gene chip) for the ABCR (ABCA4) gene. *Hum Mutat* **22**: 395-403.

Jaganathan K, Kyriazopoulou Panagiotopoulou S et al. 2019. Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**: 535-548 e524.

Khan M, Cornelis SS et al. 2019. Cost-effective molecular inversion probe-based ABCA4 sequencing reveals deep-intronic variants in Stargardt disease. *Hum Mutat* doi:10.1002/humu.23787.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**: 1073-1082.

Lieber MR. 2010. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* **79**: 181-211.

Liehr T. 2010. Cytogenetic contribution to uniparental disomy (UPD). *Mol Cytogenet* **3**: 8.

Lu B, Malcuit C et al. 2009. Long-term safety and function of RPE from human embryonic stem cells in preclinical models of macular degeneration. *Stem Cells* **27**: 2126-2135.

Maia-Lopes S, Aguirre-Lamban J et al. 2009. ABCA4 mutations in Portuguese Stargardt patients: identification of new mutations and their phenotypic analysis. *Mol Vis* **15**: 584-591.

Maugeri A, van Driel MA 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. *American Journal of Human Genetics* **64**: 1024-1035.

McVey M, Lee SE. 2008. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet* **24**: 529-538.

Nassisi M, Mohand-Said S et al. 2019. Prevalence of ABCA4 Deep-Intronic Variants and Related Phenotype in An Unsolved "One-Hit" Cohort with Stargardt Disease. *Int J Mol Sci* **20**.

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

Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. *J Comput Biol* **4**: 311-323.

Richards S, Aziz N et al. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**: 405-424.

Riveiro-Alvarez R, Valverde D et al. 2007. Partial paternal uniparental disomy (UPD) of chromosome 1 in a patient with Stargardt disease. *Mol Vis* **13**: 96-101.

Rivera A, White K 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.

Rivolta C, Berson EL, Dryja TP. 2002. Paternal uniparental heterodisomy with partial isodisomy of chromosome 1 in a patient with retinitis pigmentosa without hearing loss and a

missense mutation in the Usher syndrome type II gene USH2A. *Arch Ophthalmol* **120**: 1566-1571.

Roosing S, van den Born LI et al. 2013. Maternal uniparental isodisomy of chromosome 6 reveals a TULP1 mutation as a novel cause of cone dysfunction. *Ophthalmology* **120**: 1239-1246.

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

Runhart EH, Valkenburg D et al. 2019. Late-Onset Stargardt Disease Due to Mild, Deep-Intronic ABCA4 Alleles. *Invest Ophthalmol Vis Sci* **60**: 4249-4256.

Sangermano R, Bax NM et al. 2016. Photoreceptor progenitor mRNA analysis reveals exon skipping resulting from the ABCA4 c. 5461-10T→ C mutation in Stargardt disease. *Ophthalmology* **123**: 1375-1385.

Sangermano R, Garanto A et al. 2019. Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. *Genet Med* doi:10.1038/s41436-018-0414-9.

Sangermano R, Khan M et al. 2018. ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Res* **28**: 100-110.

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* **9**: 671.

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

Schwarz JM, Cooper DN, Schuelke M, Seelow D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature Methods* **11**: 361-362.

Souzeau E, Thompson JA et al. 2018. Maternal uniparental isodisomy of chromosome 6 unmasks a novel variant in TULP1 in a patient with early onset retinal dystrophy. *Mol Vis* **24**: 478-484.

Thompson DA, Gyurus P et al. 2000. Genetics and phenotypes of RPE65 mutations in inherited retinal degeneration. *Invest Ophthalmol Vis Sci* **41**: 4293-4299.

Thompson DA, McHenry CL et al. 2002. Retinal dystrophy due to paternal isodisomy for chromosome 1 or chromosome 2, with homoallelism for mutations in RPE65 or MERTK, respectively. *Am J Hum Genet* **70**: 224-229.

Vissers LE, Bhatt SS et al. 2009. Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. *Hum Mol Genet* **18**: 3579-3593.

Wiszniewski W, Lewis RA, Lupski JR. 2007. Achromatopsia: the CNGB3 p.T383fsX mutation results from a founder effect and is responsible for the visual phenotype in the original report of uniparental disomy 14. *Hum Genet* **121**: 433-439.

Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* **11**: 377-394.

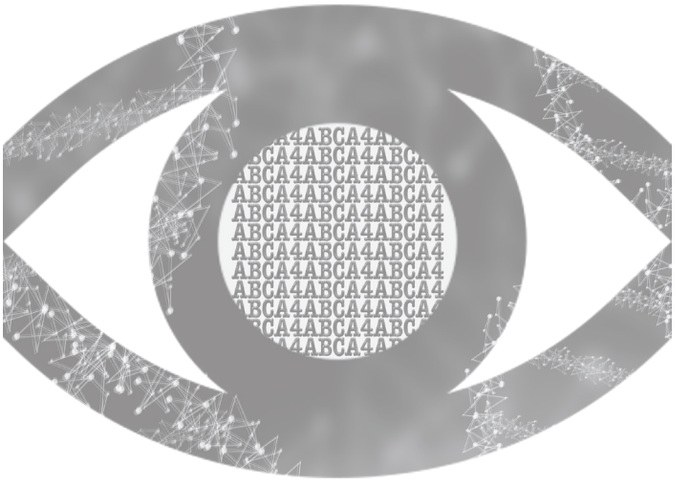
Zernant J, Lee W et al. 2017. Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration. *J Med Genet* **54**: 404-412.

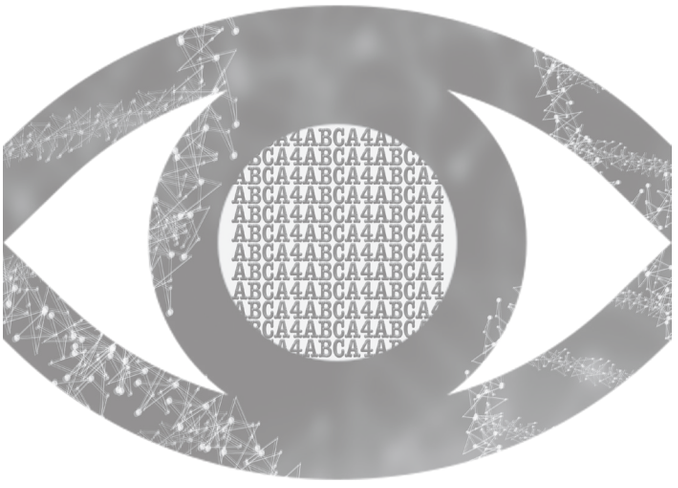
Zernant J, Lee W et al. 2018. Extremely hypomorphic and severe deep intronic variants in the ABCA4 locus result in varying Stargardt disease phenotypes. *Cold Spring Harb Mol Case Stud* **4**.

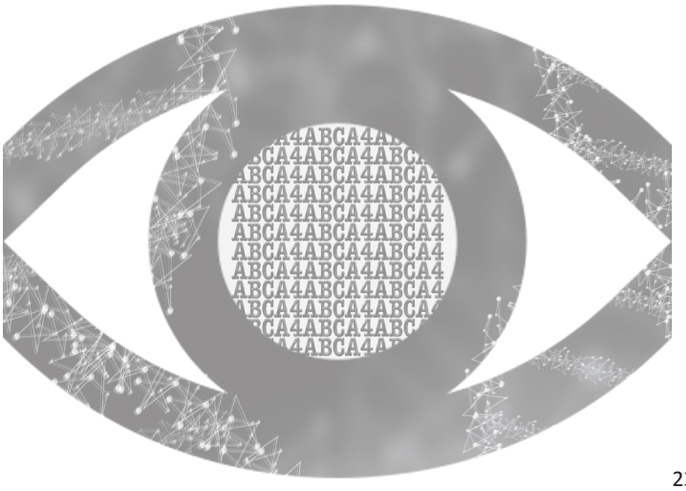
Zernant J, Xie YA et al. 2014. Analysis of the ABCA4 genomic locus in Stargardt disease. *Hum Mol Genet* **23**: 6797-6806.

Chapter 8

In or Out? New insights on exon recognition through splice-site interdependency







In or Out? New insights on exon recognition through splice-site interdependency

Mubeen Khan*, Stéphanie S. Cornelis*, Riccardo Sangermano, Iris J.M. Post, Amber Janssen Groesbeek, Jan Amsu, Christian Gilissen, Alejandro Garanto**, Rob W.J. Collin**, Frans P.M. Cremers

Int J Mol Sci. 2020**; 21**: 2300

*These authors contributed equally to this work

**These authors contributed equally to this work

Abstract

Noncanonical splice-site (NCSS) mutations are an important cause of inherited diseases. Based on *in vitro* and stem cell-based studies, some splice-site variants show a stronger splicing defect than expected based on their predicted effects, suggesting that other sequence motifs influence the outcome. We investigated whether splicing defects due to human inherited disease-associated variants in NCSS sequences in *ABCA4*, *DMD*, and *TMC1*, could be rescued by strengthening the splice-site on the other side of the exon. NCSS splice donor and acceptor site variants were selected. Rescue variants were introduced based on an increase in splice site strength prediction, and the effects of these variants were analyzed using *in vitro* splice assays in HEK293T cells. Exon skipping due to five variants in NCSS of exons in *ABCA4*, *DMD* and *TMC1* could be partially or completely rescued by increasing the predicted strengths of the other splice site of the same exon. We coined this mechanism 'splicing interdependency', which likely is based on exon recognition by the splicing machinery. Awareness of this interdependency is of importance in the classification of NCSS variants associated with disease and may open new opportunities for treatments.

Introduction

Pre-mRNA splicing of multi-exon genes is performed by the spliceosome which recognizes specific sequence motifs at exon boundaries (Brody and Abelson 1985; Frendewey and Keller 1985; Grabowski et al. 1985). Variants in these motifs can change the binding strength of the spliceosome, which can alter splicing (Pagani and Baralle 2004). This can result in disease when necessary (parts of) exons are skipped or (parts of) introns are retained, causing frameshifts, premature stop codons or disruptions in protein folding (Wang and Cooper 2007).

In order to predict the effect of genetic variants on splicing, many splice-site prediction programs have been created. Five splicing prediction programs are used in the commonly used program Alamut Visual (Interactive Biosoftware, e.g. version 2.7.1) to increase the reliability of this *in silico* analysis. Splicing prediction scores are provided for both the reference and alternative sequence, not only for the splice sites indicating the known exon-intron boundaries but also for alternative (cryptic) splice sites in nearby sequences. Decreased splicing scores often result in exon skipping, but exon elongation and partial exon truncation are also observed when alternative splice sites are recognized by the splicing machinery (Koul et al. 1999; Bonnet et al. 2008; Kergourlay et al. 2014). Different algorithms are used per splicing prediction program. SpliceSiteFinder-like (Shapiro and Senapathy 1987; Leman et al. 2018), MaxEntScan (Yeo and Burge 2004), NNSPLICE (Reese et al. 1997), GeneSplicer (Pertea et al. 2001) and Human Splicing Finder (HSF) (Desmet et al. 2009), are incorporated into the Alamut software and take into account a maximum of 80 nucleotides on either side of the variant. The exons in human DNA are relatively small (average 145-150 bp; mean 120 bp) (Consortium 2004; Bruce Alberts 2008) and introns are generally much larger (average 5.5 kb; mean 3.3 kb) (Consortium 2001; Abebrese et al. 2017), which likely forms the basis for an initial 'exon recognition' by the spliceosome. Surprisingly, the above-mentioned splicing algorithms assess the strengths of the splice acceptor site (SAS) and splice donor site (SDS) independently, not taking into account that exon recognition may also depend on the strength of the splice site at the other side of an exon.

In 2002, Hefferon et al. showed how exon skipping due to a short polypyrimidine tract upstream of *CFTR* exon 9 SAS could be rescued by a +4 adenine insertion strengthening the SDS of exon 9. *In vivo* studies of mice and sheep supported this finding (Hefferon et al. 2002). Later it was found that natural skipping of exon 6 of the *CHRNE* gene could be reduced by improving either the SAS or SDS of exon 6 by changing at least four nucleotides in the sequence (Ohno et al. 2003), indicating that if one of

the splice sites of *CHRNE* exon 6 is weak and the other is strong, normal splicing can still take place.

By studying the effect of a predicted very mild SAS variant (c.5461-10T>C) in *ABCA4* (NM_000350.2) in individuals with Stargardt disease (STGD1), we observed a very strong splicing defect (exon 39 or exon 39 and 40 skipping) in mRNA isolated from patient-derived photoreceptor precursor cells (Sangermano et al. 2016). We hypothesize that the large effect of the very mild SAS variant was due a weak SDS of exon 39.

In this study we aimed to study whether splice site interdependency represents a more general phenomenon. Next to the noncanonical splice site (NCSS) variant (c.5461-10T>C) in *ABCA4*, we selected additional predicted weak SAS NCSS variants in introns of *DMD* (NM_004006.2) and *TMC1* (NM_138691.2), that were accompanied by relatively weak SDSs (Bovolenta et al. 2012; Ganapathy et al. 2014). Each of them was previously shown to result in exon skipping and was associated with inherited human disease. Employing *in vitro* midigene splice assays to strengthen the SDS site by changing their sequence towards the SDS NCSS consensus sequence (Brent and Guigo 2004), enabled us to study the effect on splicing of the concerned exon as well as adjacent exons. Similarly, we investigated whether the exon skipping of two *ABCA4* exons due to NCSS variants at the SDS, could be corrected by strengthening the SAS site to investigate whether the dependency between splice sites works both ways.

Results

Rescuing the exon skipping effect of splice acceptor site variants

The SAS variant c.5461-10T>C constitutes the most frequent severe *ABCA4* variant in STGD1. When present alone in a midigene splice construct and transfected into HEK293T cells, it resulted in different exon 39/40 skipping RNA products (**Figure 1**). Densitometric analysis showed that RNA transcribed from the wild-type (WT) construct also shows a fair amount of exon skipping (1% of exon 39 skipping and 31% of exon 39-40 skipping, **Figure 1C**, **Table S1**), also shown previously (Sangermano et al. 2016). The BA26 midigene construct carrying c.5461-10T>C was mutagenized to also contain the alternative nucleotides A, C or T at position c.5584+4. Based on HSF the +4A variant strengthens this SDS significantly (**Figure 1E**, from 83.8 to 92.1). The +4C and +4T variants, also according to HSF, do not strengthen this splice site, but were inserted to investigate another hypothesis. The +4 to +6 nucleotides are G'GT, which are part of a weak predicted SDS (HSF: 69.0) which could negatively influence proper exon 39 recognition. The +4G>C and +4G>T changes potentially could remove this effect and thereby result in less exon 39 skipping in the presence of the WT SAS sequence c.5461-

10T (**Figure 1D**), or the mutant variant c.5461-10C. The c.5584+4G>A variant but not c.5584+4G>C or c.5584+4G>T, clearly rescued the splicing defect due to c.5461-10T>C (**Figures 1D-F**, **Figure S1**). Interestingly, the c.[5461-10T>C;5584+G>A] construct shows more correctly spliced mRNA than the WT construct. However, compared to the WT mRNA, a new band of ~400 nucleotides is present too, but this band could not be sequenced, which suggests that it probably is a heteroduplex fragment. As shown in **Figure 1A** and **1D**, the +4G>C and +4G>T changes do not result in a clear rescue of erroneous splicing in the presence of the WT or mutant SAS, respectively, suggesting that the G'GT sequence at positions +4 to +6 does not significantly affect the strength of the SDS. Thereby, we have shown that exon skipping due to a weak variant in the SAS of *ABCA4* exon 39, could be fully rescued by strengthening the SDS of the same exon.

The next variant we investigated for splice-site interdependency is the pathogenic NCSS mutation c.1705-5T>G at the SAS of *DMD* exon 15 that is associated with Becker muscular dystrophy (www.lovd.nl/DMD) (Bovolenta et al. 2012). Midigenes of *DMD* exons 14-16 were used to examine the effect of this mutation and a putative rescuing variant c.1812+4T>A on the splicing of these exons. Variant c.1705-5T>G leads to skipping of exon 15 in the midigene construct, as shown by RT-PCR (**Figure 2A-C**, **Figures S2** and **S3**). The exon skipping is largely rescued by introducing the c.1812+4T>A variant at the SDS downstream of the exon in a construct containing c.[1705-5T>G;1812+4T>A]. The densitometric analysis of RT-PCR products showed a rescue of ~73% of WT splicing as a result of the construct containing the splice-site variant and the rescue variant (**Figure 2C**, **Table S1**). Therefore, the partial rescuing effect of increasing an SDS of an exon with a weak SAS is also possible in the human *DMD* gene.

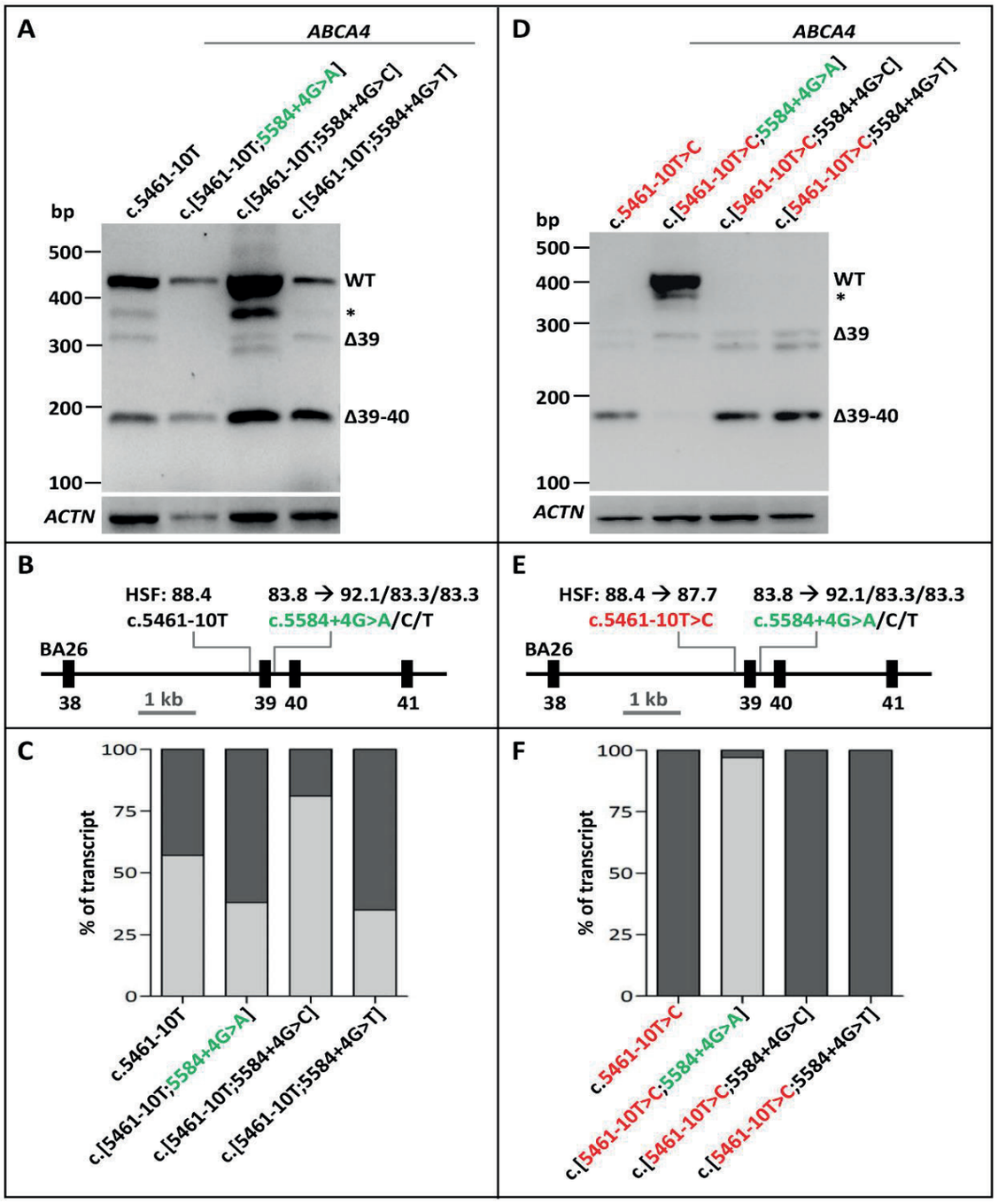


Figure 1. Rescue of exon 39/40 skipping due to *ABCA4* variant c.5461-10T>C by splice donor site strengthening. A. The effect of altered nucleotides at position c.5584+4 in the WT BA26 construct. None of the variants have a significant effect on the exon 39/40 splicing. Natural skipping of exons 39/40 is observed for all. **B.** Schematic overview of the BA26 construct used in A. annotated with the Human Splicing Finder (HSF) scores [0-100]. **C.** Semi quantification of the ratio of correctly (light gray rectangles) and aberrantly spliced (dark gray rectangles) RT-PCR products due to altered nucleotides at position c.5584+4 in the WT BA26 constructs. When multiple aberrant products were observed the percentages were summed up. **D.** The effect of altered nucleotides at c.5584+4 in the BA26 construct containing the c.5461-10T>C variant c.5584+4G>A, but not +4G>C or +4G>T, rescued the exon skipping due to c.5461-10T>C alone. **E.** Schematic overview of the BA26 construct used in B. **F.** Semi quantification of the

ratio of correctly (light gray rectangles) and aberrantly spliced (dark gray rectangles) RT-PCR products in BA26 construct containing c.5461-10T>C alone and together with rescue variants. *Band is identified as heteroduplex by Sanger sequencing. Red lettering indicates pathogenic variant. Green lettering indicates variant that rescued exon skipping.

The third pathogenic NCSS variant that we investigated was c.237-6T>G at the SAS of exon 8 in *TMC1*, which has been associated with autosomal recessive hearing impairment. Midigenes of *TMC1* containing exons 7 and 8 were generated with either the WT sequence, with the mutation alone, or one of the potential rescuing variants (i.e., c.362+4T>A or c.362+6G>T), or combinations of the mutation with the potential rescuing variants. RT-PCR of the WT construct led to a fragment of 459 bp containing *TMC1* exons 7 and 8 as well as *RHO* exon 5 (**Figure 2D**, **Figures S2** and **S4**). The c.237-6T>G variant alone led to skipping of exon 8. The c.[237-6T>G;362+4T>A] and the c.[237-6T>G;362+6G>T] constructs both result in WT mRNA, but also to some skipping of exon 8. They also resulted in an additional sequence (marked with an asterisk in **Figure 2D**), which turned out to be a heteroduplex fragment. Furthermore, all the constructs apart from the mutant construct led to a small fraction of aspecific product (marked with a hashtag) mapping to chromosome 17. The c.[237-6T>G;362+4T>A] combined construct led to more WT product (52%) than the c.[237-6T>G;362+6G>T] combined construct (29.5%) (**Figure 2F**, **Table S1**). This is not unexpected as, based on the human splice-site consensus sequence, the SDS +4A nucleotide is more prevalent than the +6T nucleotide (Brent and Guigo 2004). Altogether, we show that SAS NCSS intronic variants that decrease or completely disrupt splicing, can be partially or fully rescued by strengthening the NCSS intronic region of the SDS in multiple genes.

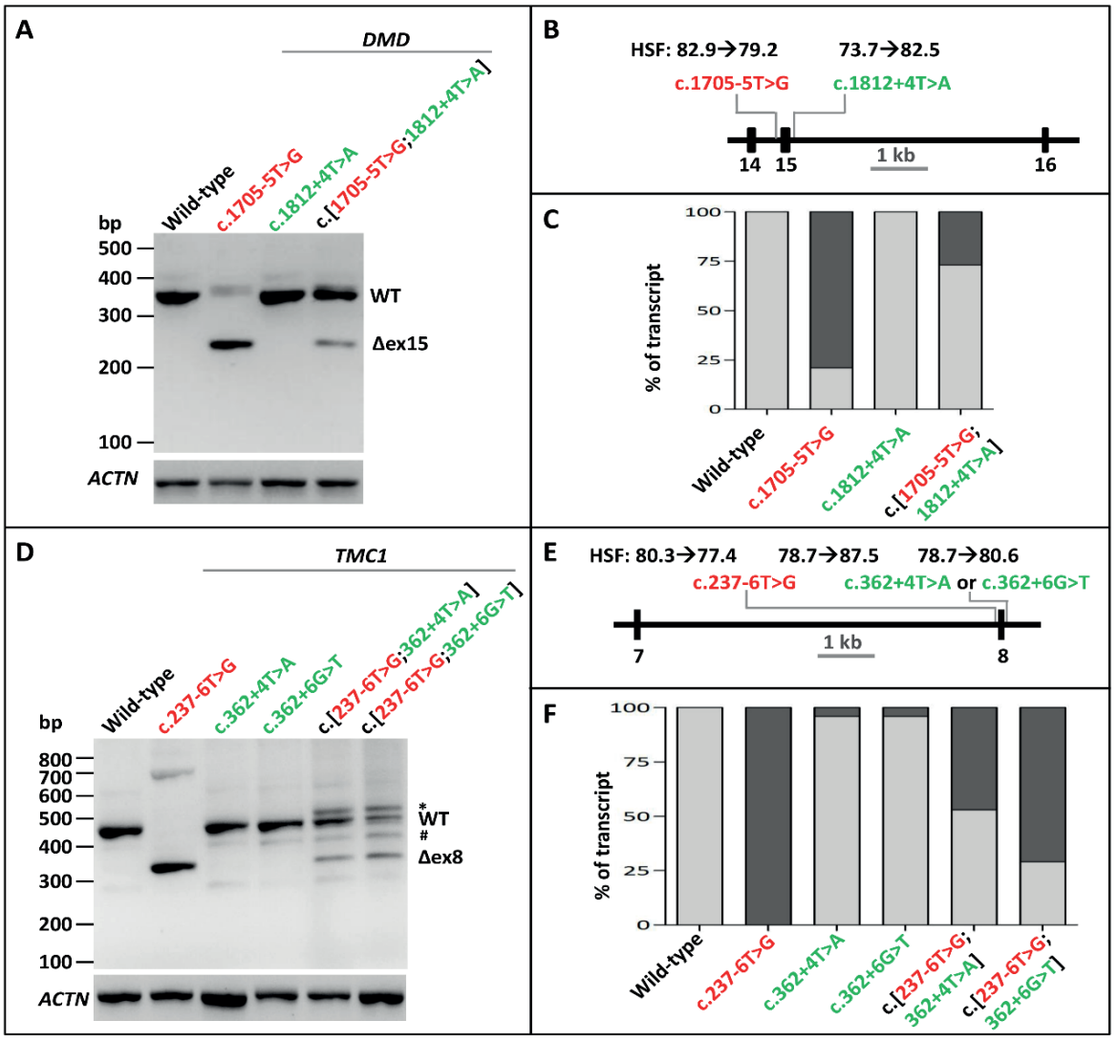


Figure 2. Splice defect rescue of pathogenic splice acceptor site variants in *DMD* and *TMC1* through splice donor site strengthening. A. The effect of the pathogenic variant c.1705-5T>G alone, the rescuing variant c.1812+4T>A alone, as well as their combination in *DMD* midigenes. Exon 15 skipping is almost completely rescued when c.1812+4T>A is introduced into the midigene. **B.** Schematic overview of the *DMD* midigene containing exons 14-16 used in A. annotated with the Human Splicing Finder (HSF) scores [0-100] of both variants. **C.** Semi quantification of the ratio of correctly (light gray rectangles) and aberrantly spliced (dark gray rectangles) RT-PCR products of WT splicing as well as exon 15 skipping in the *DMD* midigene. **D.** The effect of the pathogenic variant c.237-6T>G alone, the rescuing variants c.362+4T>A and c.362+4T>G alone, as well as the combinations of the pathogenic variant with rescuing variants in the *TMC1* midigene. c.362+4T>A shows a slightly better splice defect rescue than c.362+6G>T. **E.** Schematic overview of the *TMC1* midigene containing exons 7 and 8 used in D. annotated with the HSF scores of all three variants. **F.** Semi quantification of the ratio of correctly (light gray rectangles) and aberrantly spliced (dark gray rectangles) RT-PCR products of WT splicing as well as exon 8 skipping in the *TMC1* midigene. When multiple aberrant products were observed the percentages were summed up. *Band is identified as heteroduplex by Sanger sequencing. Red lettering denotes pathogenic variants. Green lettering indicates the variants that (partially) rescued exon skipping. #Aspecific band corresponding to a region on chromosome 17 with high similarity to *TMC1*.

Rescuing the exon skipping effect of NCSS variants at the splice donor site

After showing that NCSS intronic variants at the SAS splice site of exons in several genes can be rescued by strengthening the SDS *in vitro*, we investigated whether pathogenic NCSS variants at the SDS of exons in the *ABCA4* gene might be rescued similarly by introducing a noncanonical variant at the SAS of the same exon. The first variant we analysed was the intronic *ABCA4* c.302+4A>C variant downstream of exon 3, which led to complete skipping of exon 3 in the BA1 construct. Introducing the c.161-3A>C variant at the SAS of the exon in the mutant construct completely rescued the exon skipping to WT ratios (**Figure 3A-C**, **Table S1**, **Figures S2** and **S5**).

The exonic *ABCA4* variant c.6478A>G residing at the last nucleotide position of exon 47, which has been shown to diminish correct splicing of exon 47 with almost 50%, was previously introduced into midigene BA29 (**Chapter 2**). Mutagenizing the cytosine into a guanine at position c.6387 at the SDS of exon 47 fully rescued the exon skipping effect of variant c.6478A>G (**Figure 3D-F**, **Figures S2** and **S6**). This indicates that also weak SDS can be rescued by strengthening the SAS of the same exon.

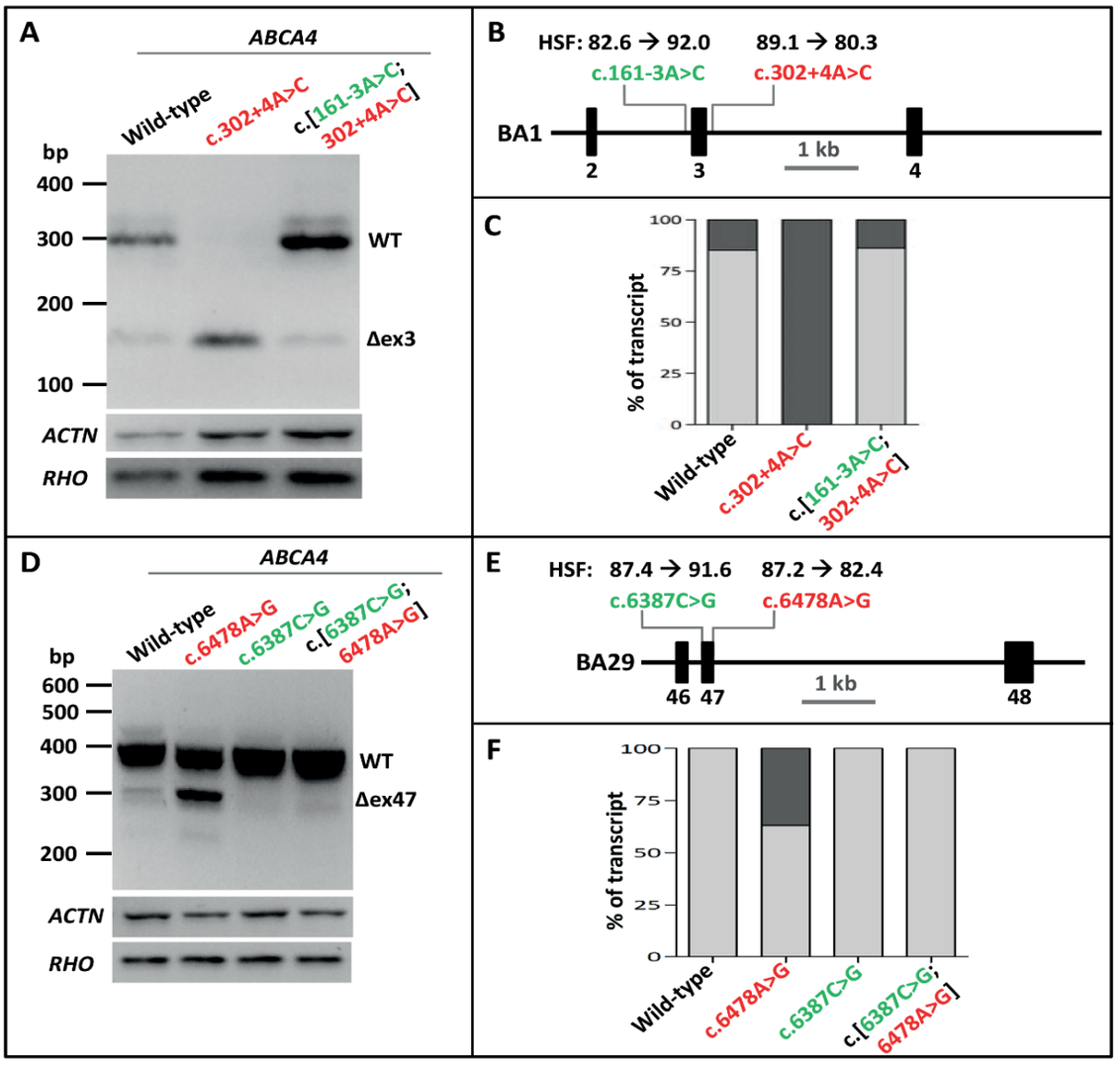


Figure 3. Strengthening splice acceptor sites rescues splice defects due to pathogenic splice donor site variants in *ABCA4* exons 3 and 47. A. The effect of the pathogenic variant c.302+4A>C as well as the combination of this variant with the rescuing variant c.161-3A>C in construct BA1. The exon 3 skipping is fully corrected. **B.** Schematic overview of the *ABCA4* BA1 construct containing exons 2-4 used in A. annotated with the Human Splicing Finder (HSF) scores [0-100] of both variants. **C.** Semi quantification of the ratio of correctly (light gray rectangles) and aberrantly spliced (dark gray rectangles) RT-PCR products of WT splicing as well as exon 3 skipping in the BA1 construct. **D.** The effect of the pathogenic variant c.6478A>G, the rescuing variants c.6387C>T as well as their combination in the BA29 construct. The partial exon 47 skipping is fully rescued after the introduction of the SAS variant. **E.** Schematic overview of the BA29 construct containing exons 46-48 used in D. annotated with the HSF scores of both variants. **F.** Semi quantification of the ratio of correctly and aberrantly spliced RT-PCR products of WT splicing as well as exon 47 skipping in the BA29 construct. Red lettering indicates pathogenic variants. Green lettering indicates variants that rescued exon skipping. When multiple aberrant products were observed the percentages were summed up.

Discussion

In this study we showed that in exons of different genes, exon skipping due to mutations of the NCSS sequences of either the SDS or the SAS can be undone by strengthening the SDS and SAS, respectively. This phenomenon, which we coined 'splicing interdependency', can be widespread, which warrants the development of new splice-site strength algorithms that take into account the strengths of both splice sites of exons. Jaganathan et al. recently reported on a new splice site prediction tool, coined 'SpliceAI', which is based on a larger sequence context of an investigated variant (Jaganathan et al. 2019). SpliceAI provides splice-site predictions for SDS and SAS of exons and pseudoexons (PEs), and thereby confirms that our findings are relevant for all multi-exon genes.

In **Figure 4**, we provide a model for splicing interdependency. A NCSS variant results in less splicing factors (SFs) binding to one end of the exon (**Figure 4A-C**) in pre-mRNA which has an effect on the exon recognition by the splicing machinery as the 'cooperation' between SFs at the exon boundaries is lost (compare **Figure 4B** and **4C**). Strengthening the opposite splice-site (**Figure 4D**) recruits extra SFs and thereby compensates for the loss of SFs binding at the NCSS variant site.

In the past, Zeniou et al. already tried to rescue exon skipping in a similar way by increasing the polypyrimidine stretch with a TCTC insertion in the SAS of exon 6 of *RPS6KA3*, formerly known as *RSK2*, when the SDS of that exon was known to be weakened due to a +3A>G change (Zeniou et al. 2004). The exact location of the TCTC insertion is not mentioned, but a TCTC insertion on the logical positions (-3_-2ins, -5_-4ins or -9_-8ins) does not lead to an increased SAS score according to all prediction programs in Alamut Visual, which might explain why the exon skipping was not rescued. As mentioned before, Hefferon et al. and Ohno et al. showed that aberrant splicing of the *CFTR* gene or reduced splicing of the *CHRNE* gene could be rescued by strengthening the alternative splice sites (Hefferon et al. 2002; Ohno et al. 2003). Similarly, our results showed that this is a common mechanism in multiple genes. Scalet et al. (2017) showed that strengthening of the SAS or SDS of *ATR* exon 9 could partially or fully rescue, respectively, the effect of an exonic variant that introduced a cryptic SDS (Scalet et al. 2017).

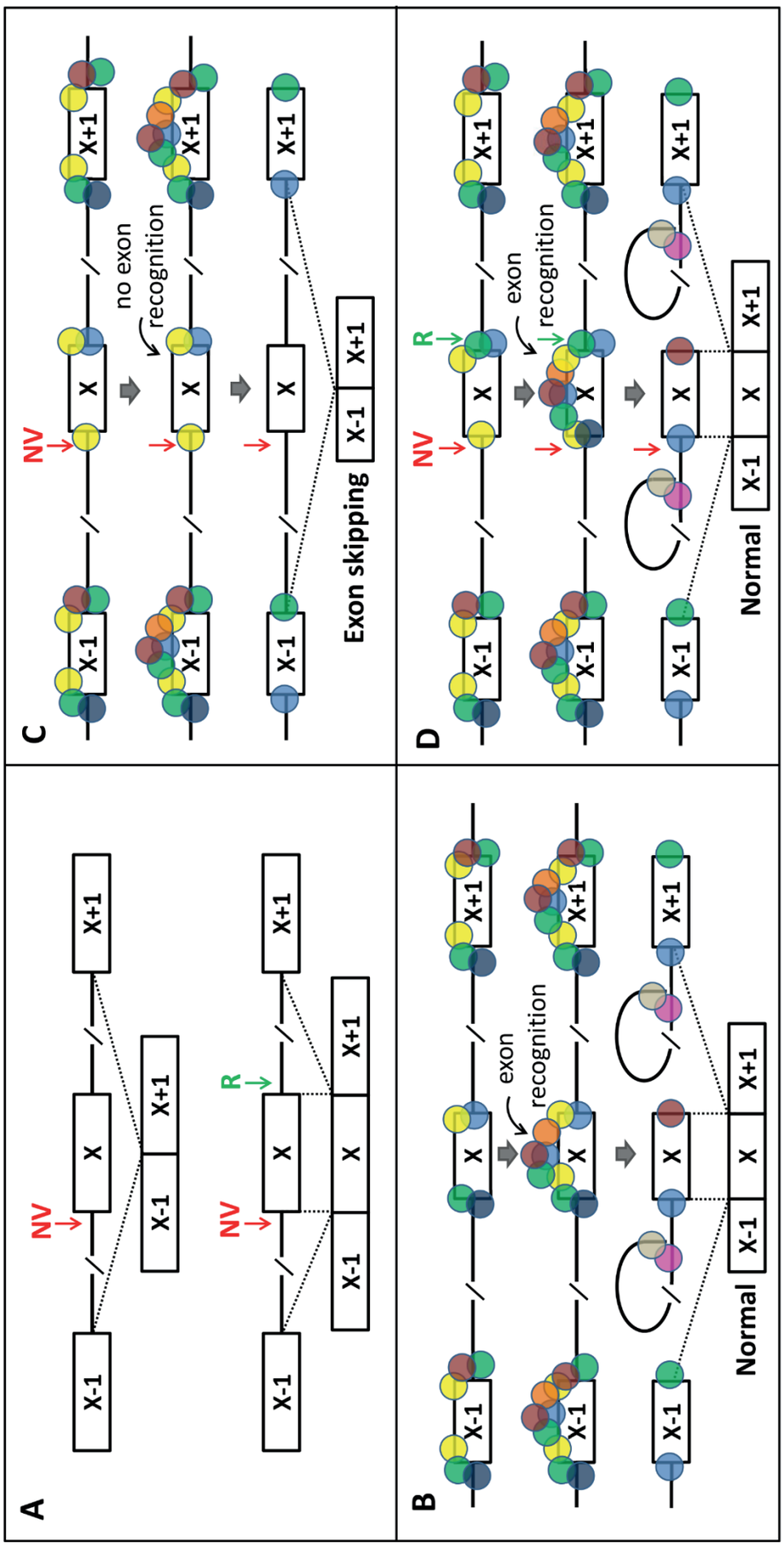

A
B
C
D
NV
R
X-1
X
X+1
no exon
recognition
exon
Exon skipping
Normal

Figure 4. Splice-site interdependency-based rescue of exon skipping. A. Principle of rescue of exon X skipping due to a noncanonical splice site (NCSS) variant (NV) at the SAS (upper panel) by strengthening the SDS with a 'Rescue variant' (R, lower panel). **B.** In normal splicing of exon X, upon binding of splice junction sequence motifs by splice factors, additional splice factors will bind and define the exon, which is followed by the lariat configuration and normal splicing. Absence of one or more splice enhancing motifs in exon X makes this exon vulnerable for exon skipping when a NV variant is present. **C.** In the presence of a NV in the SAS insufficient splice factors will bind and there is no exon definition. **D.** By strengthening the SDS with the 'R', additional splice proteins can bind and the exon definition and normal splicing are restored.

Our data clearly show that a splice-site strength is not solely dependent on its nearby sequence context – for which splicing prediction programs usually take into account less than 100 nucleotides – but also on the broader context of the pre-mRNA. It is likely that because the splicing machinery requires both a SDS and SAS, the loss of a SAS makes the upstream SDS – which it is normally spliced to – compete with the SDS of the affected exon to splice together with the next downstream SAS. Increasing the strength of the SDS of the affected exon will increase its chance to bind to the downstream SAS. The SDS of the exon upstream of the affected exon is then forced to splice with the weakened SAS by lack of a stronger SAS.

Furthermore, another contextual factor that might affect the splicing efficiency of splice sites, is the length of the adjacent exon and introns. Here, we have investigated the effect of NCSS mutations of exons with lengths between ~100-130 bp between introns of ~100-8,000 bp. Our results fit the exon definition model of Susan Berget, in which splice sites of one exon communicate with each other and thereby depend on each other (Robberson et al. 1990). However, as it is possible that intron definition may also occur in the human genome (De Conti et al. 2013), exons and introns that are respectively bigger or smaller than those investigated here, might show a different splice-site interdependency and thereby a different mechanism underlying exon skipping. Especially larger exons might be more subject to flanking intron definition versus exon definition, which might alter the mode of splicing. Interestingly, intron length is not related to the speed of intron removal in humans (Hollander et al. 2016), although this is not the case in flies (Pai et al. 2017). This could mean that the intron length has no or little effect on the splice-site recognition in humans.

In this study, predicted exon and intron splice enhancers and silencers have not been taken into account. However, prediction scores have been consulted (**Figures S1** and **S3-S6**). The importance of also considering exonic splicing enhancers (ESEs) is illustrated by a study of Jin et al (Jin et al. 1996a). They showed in a midigene construct

containing the 134-bp long exon 9 of the *ITGB3* gene (a.k.a. *GPIIIa*), that a C>A polymorphism 13 nt downstream of the exon 9 SAS (c.1138) or a G>A mutation 6 bp upstream of the exon 9 SDS (c.1255) did not lead to exon skipping, while a combination of these exonic variants did result in exon skipping (Jin et al. 1996b). Splicing predictions of the canonical splice sites only show a slight decrease for the SAS due to c.1138C>A predicted by GeneSplicer. On the contrary, ESE predictions show that both the former as well as the latter variant disrupt more strong ESEs than they create, while the ESS predictions show that ESSs of the c.1138A variant slightly increase in total strength. This, as well as other studies (Zandberg et al. 1995; Baralle et al. 2006), suggests that splicing interdependency does not only depend on the SDS and SAS , but also on ESE and ESS motifs.

Additional studies are needed to design an exon skipping prediction program that will be important for the identification of disease-causing variants by taking into account both involved splice-site strengths as well as ESEs and ESSs motifs.

Future studies will be needed to also investigate the effect of adjacent splice sites on exon skipping and determine whether these splice sites should also be taken along in exon skipping predictions.

Finally, our findings are of therapeutic importance. We previously found significant natural exon skipping events leading to frameshifts and likely mRNA degradation for *ABCA4* exons 3 and exons 39/40 (**Chapter 2**). For both, we showed correction of NCSS variant-induced exon skipping by increasing the strength of the other splice site. This means that RNA splicing modulation can potentially be applied to motifs that influence the functionality of either the SDS and SAS, irrespective of which of the two is mutated at NCSS sequences. The prime example for the therapeutic potential constitutes the steric hindrance of an intronic *SMN2* silencer element using an antisense oligonucleotide that boosts the inclusion of exon 7 in *SMN2* mRNA and effectively treats infantile-onset spinal muscular atrophy (Porensky and Burghes 2013; Finkel et al. 2016). Similar strategies can now be employed for a wide range of inherited diseases, which are due to erroneous RNA splicing.

In conclusion, the exon-skipping effects of several NCSS variants associated with inherited diseases could be corrected by increasing the strength of the other splice site of the affected exon. This shows that exon inclusion strongly depends on the combined strengths of its SAS and SDS. However, most of the utilized splice-site prediction programs do not incorporate the strength of the opposite splice site, but a recently developed program (SpliceAI) does take that into account. In addition, these findings have important repercussions for the design of RNA splicing modulation therapeutics, since RNA-binding molecules designed to promote exon inclusion or PE exclusion can now be designed to much broader target sequences.

Materials and Methods

Noncanonical splice site variant selection

To assess the role of SDS and SAS strength interdependency in proper splicing of exons, NCSS variants in inherited disease-associated genes were selected based on the following criteria i). found at least once in patients (www.lovd.nl), ii). present in a gene frequently mutated in an autosomal recessive inherited disease, iii). leading to a reduction in the strength of the splice site when calculated by at least one of the five algorithms (SpliceSiteFinder-like, MaxEntScan, NNSPLICE, GeneSplicer, HSF (Shapiro and Senapathy 1987; Reese et al. 1997; Pertea et al. 2001; Yeo and Burge 2004; Desmet et al. 2009) via Alamut Visual software version 2.7 (Interactive Biosoftware; www.interactive-biosoftware.com), and iv). the other splice site of the affected exon had a score of 80 or lower by SpliceSiteFinder-like and a score of 90 or lower by HSF. NCSS variants located at the SAS of the genes *TMC1* (c.237-6T>G), *DMD* (c.1705-5T>G), and *ABCA4* (c.5461-10T>C), and NCSS variants located at the SDS of two exons of the *ABCA4* gene (c.302+4A>C and c.6478A>G) were selected for further analysis (Bovolenta et al. 2012; Ganapathy et al. 2014; Sangermano et al. 2016).

Selection of splice donor site rescue variants

Downstream of the SDS position c.5584 (HSF 75.8) in *ABCA4* there is a GGT sequence at c.5584+4G_5584+6T, which forms a cryptic splice site (HSF 69.0 at +5) and might diminish the correct splicing of exon 39. Therefore, the position c.5584+4G was selected to be mutagenized to C, A or T, both in the WT and the SAS mutant (c.5461-10T>C) construct to assess their effect on splicing. In this way the cryptic splice site decreased in splicing strength prediction and the canonical splice site either increased or decreased in splicing strength prediction as shown by Alamut Visual (**Figure S1**). We hypothesized that a decreased splicing prediction of the canonical splice site might still lead to increased correct splicing due to the loss of the cryptic splice site.

To increase the overall strength of the SDS of the two remaining selected variants, noncanonical SDS positions between +4 and +10 were assessed for their effect on the splice site scores. Based on the splice-site consensus sequence (Brent and Guigo 2004), certain nucleotides are more conserved than others; i.e., at position +4 an adenine is more prevalent than other nucleotides. Therefore, in *DMD* the SDS sequence c.1812+4T (HSF 73.7) was selected to be mutagenized into c.1812+4A (HSF 82.5) in both WT and SAS mutant (c.1705-5T>G) constructs to increase the SDS strength (**Figure S3**).

Similarly, to decrease the effect of the *TMC1* SAS mutant (c.237-6T>G), the SDS position c.362+4T (HSF 78.7) was selected to be mutagenized into c.362+4A (HSF 82.5)

(**Figure S4).** This strengthened a cryptic splice site at +6. Therefore, we also mutagenized position c.362+6G into c.362+6T, both separately as well as combined (c.[362+4T>A;362+6G>T]), leading to three different rescue constructs.

Selection of splice acceptor site rescue variants

Two NCSS SDS variants in the *ABCA4* gene that are known to cause exon skipping (**Chapter 2**) were selected to investigate if the skipping effect of these types of mutations could also be rescued by strengthening the upstream SAS of the relevant exon. The first variant was c.302+4A>C, which was shown to cause complete skipping of exon 3 in BA1 construct. The c.161-3A>C variant was introduced into the mutant construct to increase the SAS strength (HSF 82.6 to 92.0) (**Figure S5**). The second variant was c.6478A>G causing partial exon 47 skipping in BA29 construct. The variant c.6387C>G (HSF 87.4 to 91.6) (**Figure S6**) was also introduced into the mutant construct, resulting in c.[6387C>G;6478A>G].

Generation of wild-type midigenes

To generate the WT midigene constructs for each gene (*ABCA4, DMD* and *TMC1*) bacterial artificial chromosome (BAC) clones were used. BACs were obtained from the BACPAC Resource Center (BPRC) at the Children's Hospital Oakland Research Institute (Oakland, CA, USA). Using NucleoBond Xtra Midi EF (cat. no. 740420.250, Macherey-Nagel), BAC DNA was isolated and used as a template to generate large WT fragments for each gene such that the fragments contained multiple exons when possible. For *TMC1* a region of 7.2 kb containing exons 7 and 8, and for *DMD* a 9.4 kb region containing exons 14, 15, and 16 were amplified by using Gateway-tagged PCR primers (**Table S2**) designed as described previously (**Chapter 2**). Inclusion of exon 9 into the *TMC1* construct was technically impossible as it lies 9.5 kb downstream of exon 8. Together with other sequences this would exceed the maximal insert size (11.7 kb) that we were able to clone in the Gateway system thus far (**Chapter 2**). For *ABCA4*, the WT constructs BA1 (exon 1-3), BA26 (exon 38-41) and BA29 (exon 46-48) were used (**Chapter 2**). Purification of all the amplified fragments was performed using the PCR cleanup kit (cat. no.740609.250, Macherey-Nagel) according to the manufacturer's protocol prior to Gateway cloning that was performed as described in **Chapter 2**. All the WT constructs were previously Sanger sequenced (**Chapter 2**).

Mutant midigene generation

To introduce the variants at the SAS of the WT constructs of the respective gene, site-directed mutagenesis primers were designed by using the online Quick primer tool (http://www.genomics.agilent.com/primerDesignProgram.jsp?toggle=uploaNow&mu

tate=true&_requestid=1039517). For the generation of double mutants, SAS mutant constructs and WT constructs were used as a template to introduce the rescuing variant at the selected SDS positions. Mutagenesis PCR was performed as described previously (**Chapter 2**). Presence of the introduced mutation was confirmed by Sanger sequencing. Details of the site-directed mutagenesis primers and Sanger sequencing validation primers are provided in **Table S3**.

In vitro splice assay and RT-PCR assessment

WT and mutant constructs for *ABCA4*, *DMD*, and *TMC1* were transfected in Human Embryonic Kidney 293T (HEK293T) cells and the extracted total RNA was subjected to reverse transcription (RT)-PCR, as previously described (**Chapter 2**).

Quantification of RT-PCR products

To assess the quantity of the mis-spliced and correctly spliced RT-PCR products, densitometric analysis was performed by using Image J software after gel electrophoresis as described elsewhere (Schneider et al. 2012).

Supplemental data

Supplemental data are not included in this thesis and they can be found online at the following link: https://www.mdpi.com/1422-0067/21/7/2300

References

Abebrese EL, Ali SH et al. 2017. Identification of human short introns. *PloS one* **12**: e0175393.

Baralle M, Skoko N et al. 2006. NF1 mRNA biogenesis: effect of the genomic milieu in splicing regulation of the NF1 exon 37 region. *FEBS letters* **580**: 4449-4456.

Bonnet C, Krieger S et al. 2008. Screening BRCA1 and BRCA2 unclassified variants for splicing mutations using reverse transcription PCR on patient RNA and an ex vivo assay based on a splicing reporter midigene. *Journal of medical genetics* **45**: 438-446.

Bovolenta M, Scotton C et al. 2012. Rapid, comprehensive analysis of the dystrophin transcript by a custom micro-fluidic exome array. *Human mutation* **33**: 572-581.

Brent MR, Guigo R. 2004. Recent advances in gene structure prediction. *Current opinion in structural biology* **14**: 264-272.

Brody E, Abelson J. 1985. The" spliceosome": yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science* **228**: 963-967.

Bruce Alberts AJ, Julian Lewis, Martin Raff, Keith Roberts, Peter Walter. 2008. *Molecular Biology of the Cell*.

Consortium IHGS. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860.

Consortium IHGS. 2004. Finishing the euchromatic sequence of the human genome. *Nature* **431**: 931.

De Conti L, Baralle M, Buratti E. 2013. Exon and intron definition in pre-mRNA splicing. *Wiley Interdisciplinary Reviews: RNA* **4**: 49-60.

Desmet F-O, Hamroun D et al. 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic acids research* **37**: e67-e67.

Finkel RS, Chiriboga CA et al. 2016. Treatment of infantile-onset spinal muscular atrophy with nusinersen: a phase 2, open-label, dose-escalation study. *The Lancet* **388**: 3017-3026.

Frendewey D, Keller W. 1985. Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. *Cell* **42**: 355-367.

Ganapathy A, Pandey N et al. 2014. Non-syndromic hearing impairment in India: high allelic heterogeneity among mutations in TMPRSS3, TMC1, USHIC, CDH23 and TMIE. *PLoS One* **9**: e84773.

Grabowski PJ, Seiler SR, Sharp PA. 1985. A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell* **42**: 345-353.

Hefferon TW, Broackes-Carter FC, Harris A, Cutting GR. 2002. Atypical 5' splice sites cause CFTR exon 9 to be vulnerable to skipping. *The American Journal of Human Genetics* **71**: 294-303.

Hollander D, Naftelberg S et al. 2016. How Are Short Exons Flanked by Long Introns Defined and Committed to Splicing? *Trends Genet* **32**: 596-606.

Jaganathan K, Panagiotopoulou SK et al. 2019. Predicting splicing from primary sequence with deep learning. *Cell*.

Jin Y, Dietz HC et al. 1996a. Glanzmann thrombasthenia. Cooperation between sequence variants in cis during splice site selection. *J Clin Invest* **98**: 1745-1754.

Kergourlay V, Raï G et al. 2014. Identification of splicing defects caused by mutations in the dysferlin gene. *Human mutation* **35**: 1532-1541.

Koul A, Nilbert M, Borg Å. 1999. A somatic BRCA2 mutation in RER+ endometrial carcinomas that specifically deletes the amino-terminal transactivation domain. *Genes, Chromosomes and Cancer* **24**: 207-212.

Leman R, Gaildrat P et al. 2018. Novel diagnostic tool for prediction of variant spliceogenicity derived from a set of 395 combined in silico/in vitro studies: an international collaborative effort. *Nucleic acids research*.

Ohno K, Milone M, Shen X-M, Engel AG. 2003. A frameshifting mutation in CHRNE unmasks skipping of the preceding exon. *Human molecular genetics* **12**: 3055-3066.

Pagani F, Baralle FE. 2004. Genomic variants in exons and introns: identifying the splicing spoilers. *Nature Reviews Genetics* **5**: 389.

Pai AA, Henriques T et al. 2017. The kinetics of pre-mRNA splicing in the Drosophila genome and the influence of gene architecture. *Elife* **6**.

Pertea M, Lin X, Salzberg SL. 2001. GeneSplicer: a new computational method for splice site prediction. *Nucleic acids research* **29**: 1185-1190.

Porensky PN, Burghes AH. 2013. Antisense oligonucleotides for the treatment of spinal muscular atrophy. *Human gene therapy* **24**: 489-498.

Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. *Journal of computational biology* **4**: 311-323.

Robberson BL, Cote GJ, Berget SM. 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol* **10**: 84-94.

Sangermano R, Bax NM et al. 2016. Photoreceptor progenitor mRNA analysis reveals exon skipping resulting from the ABCA4 c. 5461-10T→ C mutation in Stargardt disease. *Ophthalmology* **123**: 1375-1385.

Sangermano R, Khan M et al. 2018b. ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Res* **28**: 100-110.

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* **9**: 671.

Shapiro MB, Senapathy P. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic acids research* **15**: 7155-7174.

Wang G-S, Cooper TA. 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nature Reviews Genetics* **8**: 749.

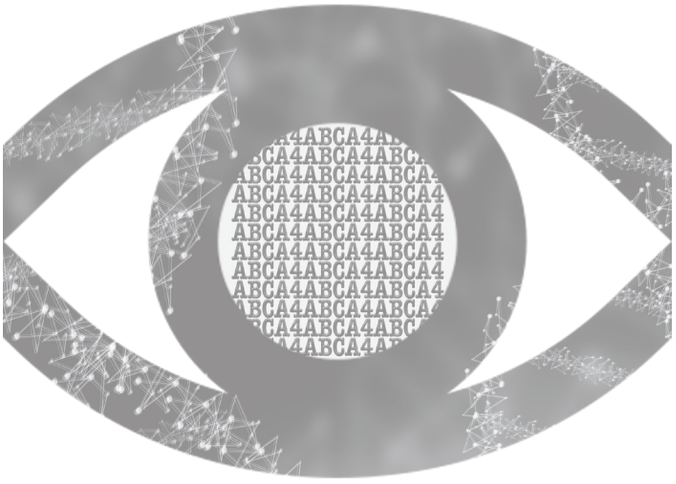
Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Journal of computational biology* **11**: 377-394.

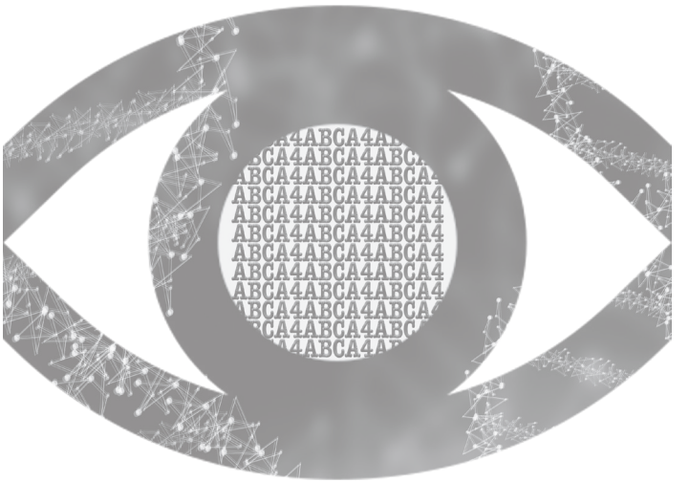
Zandberg H, Moen T, Baas P. 1995. Cooperation of 5'and 3'processing sites as well as intron and exon sequences in calcitonin exon recognition. *Nucleic acids research* **23**: 248.

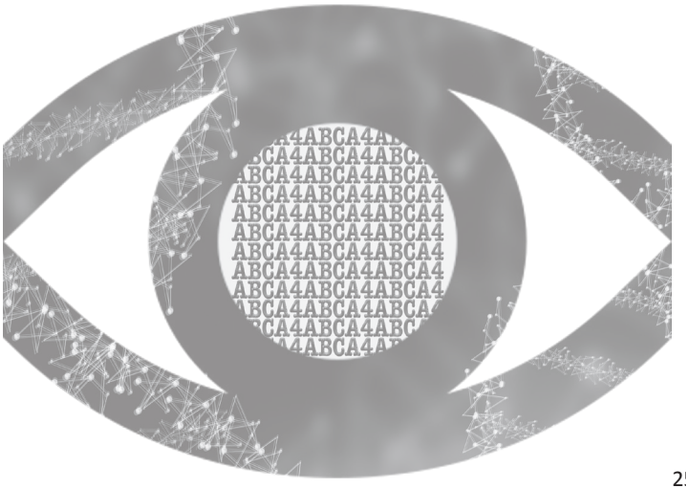
Zeniou M, Gattoni R, Hanauer A, Stévenin J. 2004. Delineation of the mechanisms of aberrant splicing caused by two unusual intronic mutations in the RSK2 gene involved in Coffin–Lowry syndrome. *Nucleic acids research* **32**: 1214-1223.

Chapter 9

General Discussion







The gene underlying Stargardt disease (STGD1), *ABCA4,* has been discovered 23 years ago (Allikmets et al. 1997). A proper genetic diagnosis for many probands still remained a challenge due to many factors, i.e., the large size of the gene (128 kb; 50 coding exons), missing coding variants due to insensitive mutation scanning approaches, the lack of a proper classification of variants residing in noncanonical splice-site (NCSS) sequences, the lack of non-coding sequence information, and the absence of proper functional testing of variants. To resolve this issue, we focused our research on the development of cost-effective sequencing of the entire 128-kb *ABCA4* gene, combined with the functional testing of both NCSS and deep-intronic (DI) variants to establish their causality. Finally, the ultimate aim of this study was not only to identify the genetic causes of *ABCA4*-associated STGD1, but also to provide a basis for the development of potential therapies. For this purpose, we used antisense oligonucleotides (AONs) which successfully blocked the insertion of pseudoexons (PEs) in *ABCA4* mRNA when tested in HEK293T cells, and in selected patient-derived fibroblasts and photoreceptor precursors cells (PPCs).

ABCA4 targeted sequencing approaches

Missing heritability is one of the major issues in inherited retinal diseases (IRDs), including *ABCA4*-associated STGD1, due to - until recently - selected sequencing of coding regions using conventional methods such as Sanger sequencing or targeted next generation sequencing (NGS). However, a few years ago, non-coding variants in *ABCA4* have been identified after a thorough analysis of the retinal *ABCA4* mRNA and targeted sequencing of selected genomic segments (Braun et al. 2013) or sequencing of the complete *ABCA4* gene (Zernant et al. 2014).

Thus, to find the missing causal variants in the non-coding regions, we sequenced the complete *ABCA4* gene locus in a total of 1,245 Stargardt and Stargardt-like probands. For this purpose, we employed two different targeted DNA fragment enrichment methods, i.e., Haloplex (**Chapter 3**) and single molecule molecular inversion probes (smMIPs) (**Chapters 4** and **7**). These studies enabled us to compare the utility of different sequencing methods for the detection of the missing variants. High average read coverage (>300x) and relatively low DNA input (100 to 200 ng) renders both smMIPs- and Haloplex-based methods more suitable sequencing approaches compared to other targeted methods, e.g., TruSeq and SureSelect. In terms of ease of use, sensitivity, average coverage, multiplexing and turnaround time, Haloplex- and smMIPs-based methods are comparable.

Nevertheless, smMIPs-based sequencing has some advantages over Haloplex-based sequencing, i) ~700x average coverage per nucleotide (**Chapter 7**) with 10-fold less costs per nucleotide compared to ~300x coverage for Haloplex sequencing

(**Chapter 3**), ii) the presence of ''unique single molecule tags'' allows to distinguish between artefacts and actual variants by improved alignment, and iii) greater depth per DNA sample which makes them suitable for the detection of low level genetic mosaicism in cancer-related genes compared to array-based conventional methods.

Despite these advantages, smMIPs-based sequencing also has some disadvantages, i) large inversions and insertions cannot be detected due to probe specificity and orientation, ii) analysis is only cost-effective when restricted to one or a few genes with large numbers of patients to be analyzed, and iii) the sequencing procedure and variant calling requires a specialized set-up.

Finally, smMIPs-based sequencing can be applied to all other frequent genetic diseases following mendelian inheritance to find missing variants in non-coding regions.

Midigene-based splice assays for NCSS and DI variants

Variants with a potential effect on RNA splicing, namely NCSS and DI variants, can be tested directly by RT-PCR of mRNA extracted from accessible human tissues such as lymphoblasts or fibroblasts. But this is not always possible, as the majority of the ocular disease-associated genes show retina-specific expression. Alternatively, *in vitro* splice assays can be performed employing minigenes transfected in HEK293T cells as a read-out system. 'Classical' minigenes however do not always reflect the actual *in vivo* situation. For instance, *ABCA4* NCSS variant c.5714+5G>A was considered moderately severe based on genotype-phenotype correlations and statistical analysis (Cremers et al. 1998; Cornelis et al. 2017). However, a mutant minigene containing only exon 40, showed full exon 40 skipping, predicted to be fully deleterious, which was not in line with previous observations. A similar finding was observed while studying NCSS variant c.5461-10T>C (Sangermano et al. 2016). We hypothesized that this could be due to two main reasons, i) relatively strong donor and acceptor sites of *RHO* exon 3 and exon 5 (Human splicing finder (HSF) scores 87.5 and 90.7, respectively) used in the Gateway adapted expression vector (pCI-NEO-*RHO*) that can lead to erroneous splicing, and/or ii) absence of a larger genomic context.

To test this hypothesis, we used a larger ''minigene'' containing *ABCA4* exons 39 through 41. Indeed, the c.5714+5G>A mutant construct showed a normal RNA product in addition to the exon 40 skipping product, as expected for a moderately severe variant. Similar results were also shown for *NF1*, where skipping of exons 36 and 37 was detected due to a variant in exon 37, which was determined by the presence of exons 34–38 in the splicing construct (Baralle et al. 2006).

We therefore designed a complete library of 32 overlapping *ABCA4*-splicing vectors coined midigenes to test *ABCA4* variants in their proper genomic context, i.e.,

flanking exons and *cis*-acting elements that influence exon recognition. Inclusion of a larger genomic context led to the identification of many complex splicing defects which would have been missed using minigenes. For example, NCSS variants c.5898+5del (**Chapter 2**) and c.5898+5G>A (**Chapter 7**) showed multiple exon 42 elongations as well as full intron 42 retention. Even more interestingly, DI variants located in intron 13, c.1938-619A>G (**Chapter 6**), c.1938-621G>A and c.1938-514G>A (**Chapter 7**), along with variant-specific PE inclusions, led to the inclusion of an upstream PE, located 493, 491 and 775 nt upstream of these variants, respectively. Similarly, DI variant c.6148-84A>T showed four RNA splicing products, namely a normal spliced RNA, the skipping of exon 45, the insertion of a 221-nt PE (PE1a) coupled with the deletion of exon 44, and finally, the insertion of a 173-nt PE (PE1b) that consist of the same splice donor site as PE1a, but a different splice acceptor site. All these findings suggest the importance of including a larger genomic context.

In this way, we provided a landscape of *ABCA4* non-coding pathogenic variants residing both in NCSS (n=60) and DI (n=26) sequences of the gene. However, it was not possible to include the entire intron 6 (15.3 kb) or intron 11 (14.3 kb) along with flanking exons in one splicing construct respectively, due to the exponentially decrease in efficacy of the Gateway cloning system for large fragments (>12 kb).

What do we miss in midigene-based splice assays and *in silico* splicing prediction tools?

Despite the detection of large numbers of pathogenic variants, many of the tested DI variants did not show any splicing defect in HEK293T cells. Absence of a splicing defect could be partially explained by the criteria employed for the selection of candidate variants. For instance, a cut-off of the allele frequency (AF) ≤0.005 gnomAD non-Finnish European (nFE) was used for candidate variant selection, but as a result some frequent yet causal DI variants may have been missed. In fact, there is strong evidence for pathogenicity of the frequent DI variant c.4253+43G>A (nFE-AF=0.006), associated with late-onset STGD1, or c.769-784C>T (nFE-AF=0.004), which is only pathogenic when present *in cis* with variant c.5603A>T (p.N1868Ile, nFE-AF=0.06) (Zernant et al. 2018; Runhart et al. 2019; **Chapter 3**). Secondly, the threshold assigned for splicing prediction scores (≥75%) was primarily based on the relative differences between the wild-type (WT) and mutant splice-site strengths, calculated by SpliceSiteFinder-like (SSFL), MaxEntScan, NNSPLICE, GeneSplicer and HSF programs. But these tools only predict the local splicing effect (40-nt up- and downstream) of these variants and disregard the proper exon recognition definition, which requires the inclusion of both splice acceptor site (SAS) and splice donor sites (SDS) as discussed below.

New insights in splice-site interdependency: a lot more to learn

Most of the splicing prediction tools do not take into account the relative strength of SAS and SDS in proper splicing. Our **chapter 8** studies have shown that there is SAS and SDS splicing interdependency (SID), which is crucial for proper exon recognition. We investigated SID by rescuing the exon skipping due to SAS NCSS variants in *ABCA4*, *DMD* and *TMC1* by strengthening the SDS. Similar results were obtained for SDS NCSS variants *in ABCA4*. These findings advocate for the development of a new splice site strength algorithm that takes into account the role of SID and other splicing motifs for *in silico* assessment. Recently, an improved algorithm, SpliceAI, was reported (Jaganathan et al. 2019), which does correlate the strengths of the SAS and SDS for the inclusion or exclusion of exons and PEs due to splicing disrupting and splicing enhancing variants, respectively. Similarly, the potential role of other splicing motifs such as exonic splicing enhancers (ESEs) and exonic splicing silencers (ESSs) cannot be ruled out. The Alamut Visual algorithms do not include these motifs in their predictions. SpliceAI was trained on RNAseq data of 35 different human tissues and theoretically will take the effect of these motifs into account. However, SpliceAI was not trained on RNAseq data of the human retina and therefore also may not predict retina-specific splicing defects due to changes in retina-specific ESE or ESS motifs.

In addition, some of the NCSS variants described in **chapters 2, 4** and **6**, not only affect the splicing of the respective exon but showed an effect on the flanking exons, i.e., variant c.5461-10T>C, located at SAS of exon 39, led to the skipping of exon 39 or exon 39 and 40 together. But none of the aforementioned splicing prediction tools could predict these complex splicing defects as they lack the ability to assess the potential role of RNA secondary structures in pre-mRNA splicing (Eperon et al. 1988; Graveley 2005; Hiller et al. 2007; Lovci et al. 2013; Vaz-Drago et al. 2017). For instance, a pathogenic role of nucleotide changes affecting RNA structure has been identified in neurodegenerative diseases (Grover et al. 1999; Varani et al. 1999; Jiang et al. 2000; Yasuda et al. 2000). Despite the existence of many RNA secondary structure predictors like the mfold web server (http://unafold.rna.albany.edu/?q=mfold), current splicing programs still lack the integration of this kind of information in their algorithms.

Alternative cellular systems

Although we have shown experimentally that the splicing of *ABCA4* exons in WT midigene constructs transfected in HEK293T cells most often capture what is observed in the human retina (**Chapter 2**), we cannot exclude that some of the negative results obtained after transfection of mutant splicing constructs (**Chapters 3, 4, 5, 6** and **7**) could be due to the absence of retina-specific splicing factors in these cells. Tissue-specific splicing has been reported in neuronal tissues (Calarco et al. 2009), and retina-

specific splicing factors were identified by transcriptome analysis of normal human retina and by comparing mouse retinae with and without photoreceptor cells (Farkas et al. 2013; Murphy et al. 2016). Thereby, it requires to investigate the splicing defects in their natural cellular environment, i.e., photoreceptors. The tissue-specific expression of *ABCA4* hampered the investigation of splicing defects in accessible cellular types such as lymphoblasts, fibroblasts and keratinocytes. Fortunately, the introduction of stem cell technology has revolutionized the way to perform functional studies for retina-specific expressed genes. Starting from a patient skin biopsy or peripheral blood cells, induced pluripotent stem cells (iPSCs) can be generated by using the four transcription factors (Oct3/4, Sox2, Klf4, and c-Myc) described by Takahashi and colleagues (Takahashi et al. 2007).

The iPSCs can be differentiated into any cell lineage including PPCs by using selected growth factors. It is important to note that new differentiation protocols are published regularly. We employed the protocol described by Zhou et al, which enabled us to obtain an early cone (PPCs) and RPE cell population giving a robust *ABCA4* expression in PPCs after 30 days of differentiation (Zhou et al. 2015).

Employing PPCs, we discovered that two neighboring DI variants in intron 30 of *ABCA4*, c.4539+2001G>A and c.4539+2028C>T (Braun et al. 2013) led to a retina-specific partial inclusion of a 345-nt PE (Albert et al. 2018). Based on genotype-phenotype correlations, the observed 50% and 30% of mutant RNA (from the relevant alleles), respectively, are probably only half of the mutant RNAs expected in the patients' retinae. These findings suggest that the PPCs, after 30 days, have not yet fully completed their differentiation. Unlike most of the DI variants, both variants do not affect the strength of flanking splicing sites of a PE, but alter and/or create ESE motifs inside the PE. Additionally, we also identified a retina-specific splicing defect for variant c.5196+1137G>A, the 3rd most frequent DI variant in our cohort (**Chapters 3, 4, 5** and **7**). It showed almost negligible PE inclusion in HEK293T cells whereas ~45% of RNA product of the patient-derived PPCs contained a 73-nt PE (**Chapter 5**). As the RT-PCR was not allele-specific and the patient was heterozygous for the mutant allele, these results are in agreement with genotype-phenotype correlations which consider that c.5196+1137G>A has a moderate to severe effect on ABCA4 activity. Similarly, *CEP290* variant c.2991+1655A>G showed a much stronger PE insertion in patient-derived organoids (mutant vs correct RNA ~4:1) (Parfitt et al. 2016), compared to the ~1:1 ratio in fibroblasts (Gerard et al. 2012; Garanto et al. 2016). These findings demonstrate the power of using a relevant cellular system to study the splicing defects.

However, it is not always possible to study splicing defects in the relevant tissue, due to i) absence of STGD1 patient donor eyes, ii) non-availability of patient samples to generate PPCs, and iii) huge time and costs investments. Alternatively, *in vitro*

splicing assays represent a valuable method to experimentally validate the *in silico* results. But as already mentioned, HEK293T cells may not always mimic the actual splicing defects in photoreceptors. This requires to further investigate other suitable cell lines exhibiting retina-like characteristics, i.e., Y-79, WERI-Rb1, R28 and 661W. Of these, Y-79 (rod precursor) and WERI-Rb1 (retinal stem-cell like) are derived from human retinoblastoma cells, while R28 (a retinal cell line) and 661W (cone precursor cells) are derived from mouse retina precursor cells. In addition, two other lines, RPE1-hTERT and ARPE19, originated from human cell lines that exhibit RPE characteristics and could be used as a read-out system, as all of these lines are transfectable.

The selection of the appropriate cell line should mainly be based upon the ultimate aim of the study, i.e., RNA splicing analysis, expression profiles or cellular morphology. In this regard, for studying the splicing defects for retina-specific genes, Y-79 or WERI-Rb1 cells-lines seems to be most closely mimicking the PPCs as an alternative read-out system. However, to determine the transfection efficiency of these cell lines before the complete replacement of the HEK293T cells, additional *in vitro* studies are required.

Pseudoexons as potential targets for splicing modulation

Since the identification of the first PE in β-Thalassemia patients (Dobkin et al. 1983), PE inclusions are frequently detected as an underlying cause of genetic diseases (Dhir and Buratti 2010). Most of the PE formations disrupt the reading frame by introducing premature stop codons (Popp and Maquat 2013). Based upon the literature reported till 2017, ~185 DI variants in 77 different genes resulting in PE formation have been described (Liquori et al. 2016; Vaz-Drago et al. 2017). At the molecular level, creation or strengthening of a cryptic donor or acceptor site explains the great majority of these events. In contrast, variants that create or disrupt splicing enhancer or silencer elements, have been found less frequently. Interestingly, it has been shown that almost 50% of all identified PEs contain transposable elements, in particular short interspersed nuclear elements such as *Alu* repeats (Vorechovsky 2010).

Recently, numerous experimental and clinical studies have focused on the development of promising approaches to reverse incorrect splicing. The most effective is the use of AONs and we successfully corrected splicing defects in HEK293T cells, patient-derived fibroblasts and/or PPCs (**Chapters 3** and **5**). Several studies have shown that the efficacy of AONs to achieve splicing modulation merely depends on the accessibility of the target mRNA molecule. Similarly, in our studies all variants showed complete rescue of the erroneous splicing products (**Chapters 3** and **5**) by at least one of the tested AONs, except for variant c.4253+43G>A, where only a partial rescue could be achieved. This variant resulted in partial exon 28 skipping and its close vicinity to

exon 28 restricts the design of AONs to a small genomic sequence (**Chapter 3**). AONs located close to an exon may interfere with normal splicing processes, thereby complicating the development of safe and effective therapeutic strategies.

Alternatively, CRISPR-based genome editing holds a lot of potential in treating IRDs (Yu and Wu 2018). Proof of concept studies are now being conducted on rodents. Disruption of the murine *Rho* p.Ser334* allele results in delayed retinal degeneration (Bakondi et al. 2016). A similar approach was used for *Cep290* in the mouse retina, to delete the intronic fragment harboring the most frequent LCA-associated DI variant, c.2991+1655A>G, by using AAV5-mediated delivery of CRISPR components (Ruan et al. 2017). This kind of approach can also be beneficial for STGD1 patients carrying intronic mutations, as some intronic regions (introns 7, 13, 30, 36; **Chapter 7**) showed a clustering of DI variants. Following CRISPR treatment, promising therapeutic benefits have been achieved in a few rodent models of retinal diseases but off-target events and immune responses still remain a major concern.

Missing heritability in STGD1

Upon complete Haloplex-based *ABCA4* sequence analysis of 36 clinically well-defined Dutch cases displaying the ophthalmologic hallmarks of STGD1, we found two pathogenic *ABCA4* variants in 24/36 probands (**Chapter 3**). Follow-up studies (e.g., **Chapter 7**) resolved another seven cases leaving us with five mono-allelic cases. This means that in our total cohort, we have genetically solved 295/300 (98%) Dutch cases. When sequencing the complete *ABCA4* gene of 1,054 probands using smMIPs, 448/1,054 probands carried two alleles, 239/1,054 probands from other centers carried one allele and 364/1,054 probands carried no *ABCA4* allele. We believe that this much higher missing heritability could be due to many reasons; some of the most plausible ones are discussed below.

First of all, enrichment methods employed in these studies can pose a technical risk of missing pathogenic variants. As shown in **chapters 3** and **7**, both Haloplex- and smMIPs-based enrichment methods showed ~97% coverage, leaving ~3% of the sequences uncovered. This lack of information is inevitable and is largely due to the presence of interspersed repeated elements across the locus, and the use of short read-based sequencing as the target capture size is 220 nt for Haloplex- and 110 nt for smMIPs-based sequencing. In the future, most of these problems can be solved by employing long-read real-time sequencing technologies that are able to generate reads with an average size larger than 10,000 bp (Eid et al. 2009).

smMIPs-based sequencing is semi-automated and experimental variability could be introduced by the operator. Other factors are DNA quantity, as read coverage is directly related to a sample's DNA quantity compared to the other samples in one

run, and the suboptimal quality of reagents. The operator-induced variation can even be more pronounced for Haloplex target enrichment as all experiments are performed manually. Poor coverage of the targeted fragments can lead to low-confidence variants which do not pass quality control steps and subsequently are erroneously discarded from the analysis.

Secondly, the presence of one *ABCA4* variant in 239/1,054 probands could be explained by the high carrier frequency of *ABCA4* variants in the general population, estimated to be ~5% (Maugeri et al. 1999; Cornelis et al. 2017), excluding c.5603A>T. These 239 cases most likely can be explained by the variants in other maculopathy-associated genes.

Most importantly, as shown by the huge differences of bi-allelic cases between the cohorts from the Netherlands (98%) (Bax et al. 2015; Runhart et al. 2018 and **Chapter 3**) and elsewhere (43%) (**Chapter 7**), we consider that the international samples often were wrongly diagnosed as 'Stargardt disease'. This is mainly due to the overlapping clinical expression of disease phenotypes beginning in the central macula (maculopathies). True '*ABCA4*-associated Stargardt disease' exhibits three main clinical features which include macular cone degeneration, the formation of flecks, and peripapillary sparing. Maculopathies associated with ~38 other genes, e.g., *BEST1, CRB1, CRX, ELOVL4*, *PROM1*, *PRPH2, RAB28, RDH11* and *ROM1* mostly exhibit only one or two of these three characteristics, and often also show additional features (Apfelstedt-Sylla et al. 1995; Xie et al. 2014; Riveiro-Alvarez et al. 2015; Zaneveld et al. 2015; Zernant et al. 2017; Imani et al. 2018; Ma et al. 2019). Among these genes, several harbor heterozygous variants that are inherited in an autosomal dominant manner. As several variants show reduced penetrance, the probands may appear as sporadic or autosomal recessive cases. The phenotype that mimics *ABCA4*-associated STGD1 most closely is *PRPH2*-associated macular disease that often exhibits all three *ABCA4* disease features. To investigate this further, we also employed 39 smMIPs to sequence all three *PRPH2* exons in parallel with the *ABCA4* smMIPs-based sequencing, which led to the identification of 26 unique variants in 33 probands recruited as STGD-like cases (Khan M, Dhaenens C-M, Cremers F.P.M, unpublished data). These findings suggest that the majority of the 606 probands with one or no *ABCA4* allele carry variants in other maculopathy-associated genes.

Additionally, some causal variants residing in non-coding regions may have escaped our attention due to the stringent candidate variant selection criteria or the absence of a selection criteria for putative regulatory variants. Thus far, there is limited evidence for *ABCA4* variants affecting transcription (Bauwens et al. 2019), and these were not detected in our cohorts (**Chapters 3** and **7**). The impact of genetic *cis*- or *trans*-

acting modifiers can also not be excluded, as potential modifier alleles in *AHI1* were found in *CEP290*-associated cases (Coppieters et al. 2010).

Moreover, complex alleles are not uncommon in STGD1 (Lewis et al. 1999; Maugeri et al. 1999) and they have been detected in approximately 12% of all STGD1 patients (Cornelis et al. 2017). In the context of 'missing heritability', c.[2588G>C;5603A>T] is very relevant. Independently, these two variants, when in trans with a severe ABCA4 variant, are not causing STGD1 in most individuals (Zernant et al. 2017, Runhart et al. 2018). However, the complex allele seems to be fully penetrant only when in trans with a severe variant. It cannot be excluded that other combinations of very mild ABCA4 variants occurring in cis have been overlooked as we may lack the proper in silico prediction tools to evaluate their pathogenicity.

Another critical aspect which is not fully addressed in our studies is a detailed risk evaluation of variants of uncertain significance (VUS). The inability to classify the VUS either as (likely) pathogenic or (likely) benign variations represents a challenge for proper genetic diagnosis. Despite being pathogenic these variants could have been discarded, due to lack of proper *in silico* evaluation and functional testing. Recently, VUS have been functionally characterized in *BRCA1/2* (Caleca et al. 2019; Mesman et al. 2019). Similar approaches can be applied in the absence of patient material, and VUS can be introduced in animal models using genome editing techniques to investigate their effect on the phenotype.

Finally, some of the structural variants (SVs) could have been missed due to the lack of SV detection in the Haloplex sequencing approach (**Chapter 3**) or insensitivity of the SV detection tool used in the smMIPs study (**Chapter 7**), which may need further automation to detect deletions smaller than ~400 bp.

Diagnostic implications

In this study, we identified bi-allelic *ABCA4* variants in 569 STGD1 probands (**Chapters 3, 4** and **7**), which allowed us to provide a 'landscape overview' of the different *ABCA4* variants associated with STGD1, including coding and non-coding SNVs, SVs, and a chromosome 1 uniparental isodisomy (UPD). For sequencing of the complete *ABCA4* gene, we employed smMIPs which proved to be a cost-effective method. Estimated costs per sample are ~€30 (excluding smMIPs synthesis). In addition, the high coverage also enabled the robust detection of SVs.

Selection of the pathogenic variants was performed based on the following criteria, i) gnomAD nFE-AF ≤0.005, ii) a high average read coverage ≥200, iii) coding variants pathogenicity was assessed by employing online tools, i.e., Sorting Intolerant From Tolerant (SIFT; https://sift.bii.a-star.edu.sg) (Kumar et al. 2009), Polymorphism Phenotyping v2 (PolyPhen-2) (Adzhubei et al. 2010), and Mutation Taster (Schwarz et

al. 2014), and iv) non-coding variants were selected based upon splicing predictions tool (SSFL, MaxEntScan, NNSPLICE, GeneSplicer, HSF via Alamut Visual software version 2.7 (Biosoftware, 2014; Interactive Biosoftware, Rouen, France; www.interactivebiosoftware) (Reese et al. 1997; Pertea et al. 2001; Cartegni et al. 2003; Yeo and Burge 2004; Desmet et al. 2009). All above-mentioned variant filtration criteria are compatible with the standards of diagnostic laboratories.

By employing *in vitro* splice assays using HEK293T cells we functionally characterized all reported (n=44) and newly identified (n=21) NCSS variants (**Chapters 3, 4, 6** and **7**). This allowed a proper functional classification for each variant by performing a quantitative RNA analysis and comparison of these data with phenotypic data in reported cases. Later we extended our midigene analysis to test non-coding variants (**Chapters 3, 4, 5, 6** and **7**), which led to the identification of 25 novel pathogenic DI variants. Our splice assays can be performed for any other human disease gene as bacterial artificial chromosome clones are available for all genes (https://bacpacresources.org/) (Osoegawa and de Jong 2004), and ~90% of introns are in the size range for midigenes (<11 kb) (Sakharkar et al. 2004). Despite the fact that the generation of a complete WT splicing vector library for a given gene seems quite labor-intensive, it is a one-time effort and can ensure the functional testing of any identified variant in that particular gene.

All our studies not only expanded the list of pathogenic NCSS and DI variants (n=98) (**Chapters 3** to **7**, Braun et al, 2013, Bauwens et al, 2019), but also significantly increased the number of novel coding and canonical splice-site variants (n=97). As described in **chapter 4**, sequencing of all coding and selected DI variants, increased the variant detection rate in our pre-screened German/French cohort of mono-allelic STGD1 patients. Similarly, in **chapter 7** complete *ABCA4* sequencing led to the identification of 14 previously identified DI variants (**Chapters 3** and **6**, Braun et al, 2013 and Bauwens et al, 2019), in 99 alleles. Thirteen novel DI variants were found in another 18 alleles and the most frequent of these were found in only three STGD1 probands. These results suggest that, for Caucasian STGD1 cases, future *ABCA4* sequencing will mostly reveal the 33 currently known DI variants.

DI variants leading to PE formation have been reported for several ocular disease-associated genes such as *CEP290*, *CHM*, *CNGB3*, *COL11A1*, *COL2A1*, *OFD1*, *OPA1, PROM1*, *PRPF31*, *RB1* and *USH2A* (van den Hurk et al. 2003; den Hollander et al. 2006; Dehainault et al. 2007; Rio Frio et al. 2009; Richards et al. 2012; Webb et al. 2012; Bonifert et al. 2014; Liquori et al. 2016; Mayer et al. 2016; Weisschuh et al. 2019). Through extensive sequencing and functional characterization, *ABCA4* has become the ocular disease gene with the largest number of different (n=33) and total (n=349)

pathogenic DI variants identified so far (**Chapters 3, 4, 6** and **7**; Braun et al. 2013; Albert et al. 2018; Bauwens, et al. 2019).

Additionally, we identified 11 novel SVs in 16 alleles, hence increasing the number of SVs found so far in *ABCA4* in STGD1 cases to a total of 36 (**Chapter 7**). Intriguingly, a script employed for the detection of SVs provides the information on the potential breakpoints as it shows the increase or decrease in the coverage ratio at the specific smMIPs positions.

In conclusion, we showed that comprehensive sequence analysis of the *ABCA4* gene, coupled with functional characterization of non-coding variants is crucial to resolve the missing heritability in STGD1 cases. Diagnostic centers should therefore consider the implementation of targeted sequencing approaches which could as a minimum cover all known DI variants to increase the variant detection rate in their unsolved STGD1 patients, as data of all *ABCA4* pathogenic variants is publicly available at www.lovd.nl/ABCA4. This will not only improve molecular diagnostics, but will also increase the number of genetically explained patients eligible for ongoing and future STGD1 clinical trials.

Although our studies were mainly focused on *ABCA4*-associated STGD1, all techniques employed for in-depth genomics and transcriptomics analysis of *ABCA4* have a broader implication and are applicable to any other disease gene. As depicted in Figure 1, such an integrated approach could enable clinicians and geneticists to resolve the missing heritability in many genetic diseases, as the number of novel disease genes (RETNET: after the introduction of WES 10 per year from 2010 – 2015 compared to 2,5 per year from 2016 onwards) is not growing that much anymore but there is an increased evidence of more non-coding pathogenic variants in known disease causing genes. Pathogenic variants can be identified systematically, starting from the detection of known pathogenic variants, as several databases exist to provide variant information per gene. ClinVar (Landrum et al. 2016) and Leiden Open source Variation Databases (LOVDs) (Fokkema et al. 2011) are examples of such databases, and also provide genotype-phenotype information. Furthermore, non-coding candidate variants can be selected by the criteria described for *ABCA4* splice variants selection (**Chapters 2** to **8**), followed by the functional testing in midigenes splicing assay or in patient-derived stem cells. Thereby, enabling many patients for suitable therapeutic approach.

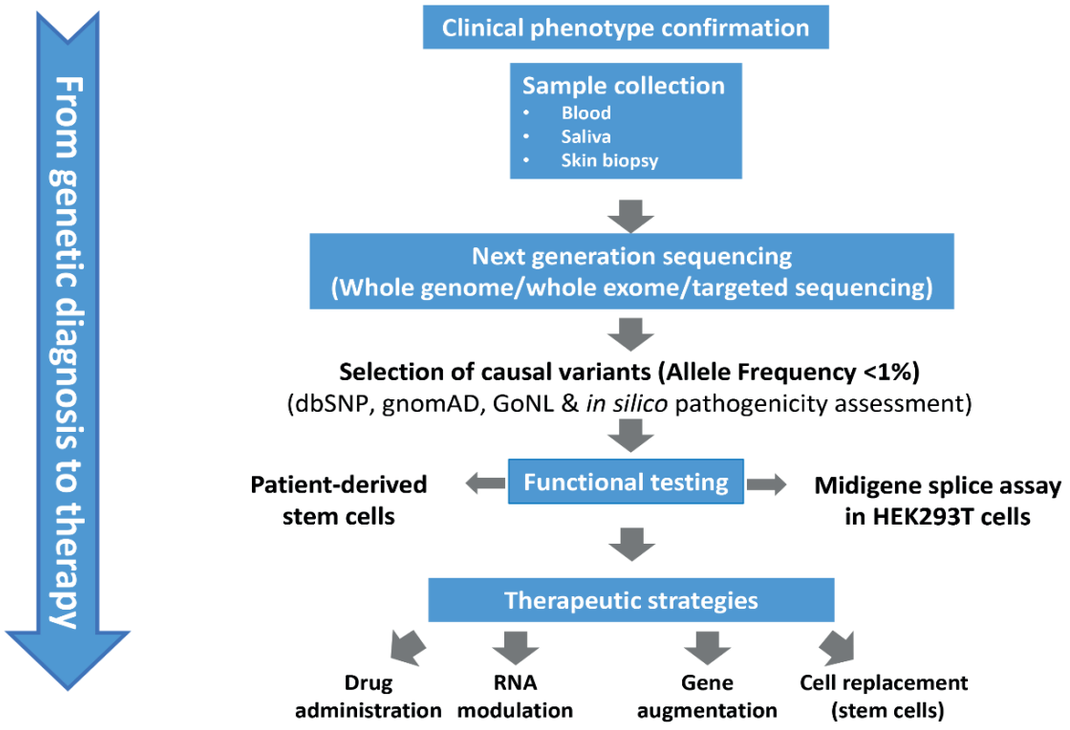


Figure 1: Schematic representation of the workflow depicting all the steps from genetic diagnosis to therapy.

Therapeutic interventions

Receiving a molecular diagnosis becomes increasingly important with the development of different therapeutic options. Available approaches mainly range from mutation-specific to more generally applicable cell replacement strategies, selection of which mainly depends on the primary genetic defect, and the disease stage at the time of treatment (Vazquez-Dominguez et al. 2019). Though no approved therapy to treat *ABCA4*-associated STGD1 exists yet, a reasonably large window of therapeutic intervention can be employed mainly due to i) a relatively small increase in ABCA4 activity in retinal cells may halt disease progression, and ii) in most of the *ABCA4*-associated phenotypes, a delayed degeneration of retinal cells is observed.

All possible therapeutic strategies available for *ABCA4*-associated STGD1 are briefly discussed here. These treatments are mainly based on RNA modulation, drug-based therapies, stem cell therapies and gene augmentation, many of which are currently in clinical trials or in preclinical development.

Splicing modulation therapies mainly employ AONs and are focussed on those variants that affect pre-mRNA splicing of *ABCA4.* AONs are small (~15-25 nt) molecules which are able to bind pre-mRNA and can induce RNA modulations (Hammond and Wood 2011). AONs successfully blocked the insertion of PEs in *ABCA4* in HEK293T cells and in patient-derived fibroblasts and photoreceptor precursors cells (Albert et al.

2018, Bauwens et al. 2019, **Chapters 3** and **5**). This could benefit a large number of STGD1 patients, since we showed in **chapter 3** and **7** that ~10% of all bi-allelic STGD1 probands carry DI mutations. Next to the PE exclusion, AONs can be employed to rescue all splice defects at the mRNA level, i.e., exon skipping and exon elongation. AONs administration can be rapidly translated to clinics as positive interim results have been observed for intravitreal AONs administration in *CEP290*-associated LCA subjects (Cideciyan et al. 2019), carrying one of the first described DI variant (c.2991+1655A>G) in IRDs. Despite being safe and effective therapeutic molecules, AONs are mutation specific and need multiple injections.

Currently some of the compounds under investigation are ALK-001 (C20-D3-vitamin A), fenretinide (N-(4-hydroxyphenyl) retinamide) and A1120 (2-(4-(2-(trifluoromethyl)phenyl)piperidine-1-carboxamido)benzoic acid). Despite their different chemical composition and target sites, drug therapies usually aim to suppress the disease progression by inhibiting the lipofuscin accumulation.

Another approach to encounter the degeneration of photoreceptors and RPE cells is cell replacement therapy. A clinical trial based on human embryonic stem cells (hESCs)-derived RPE cells administration was conducted in 2011 (Marlborough, MA, NCT01345006). Preclinical studies conducted in mice injected with hESCs-derived RPE cells, have shown an improvement in visual function (Lu et al. 2009). Importantly, no evidence of tumor formation or spread to other areas of the body was observed, indicating that cells stably recognized and joined the original tissue.

Based on the approval of gene augmentation therapy for *RPE65*-associated retinitis pigmentosa, it can be an attractive approach for *ABCA4*-associated STGD1 as well. However, the commonly used AAV-mediated gene delivery method is not very suitable for the *ABCA4* cDNA (size: 6.8 kb), due to a limited capacity of the vector (~4.7 kb). Another possibility could be the administration of lentiviral-mediated gene delivery (packaging capacity ~8 kb) (Thomas et al. 2003). The only disadvantage is their ability to randomly integrate in the genome. In the first clinical trial (NCT01367444) of a lentivirus-based drug, a STGD1 patient has been treated subretinally at the Casey Eye Institute in Portland, USA. A phase I/II clinical trial was also conducted in Paris, where 28 STGD1 cases were enrolled for SAR422459, but no results have been published.

Alternatively, a non-viral DNA nanoparticles-based approach has been introduced by Han and colleagues to deliver the human *ABCA4* cDNA to *Abca4* knockout mice (Han et al. 2012). In treated animals, persistent gene expression was observed up to eight months post-delivery, along with significant functional (dark adaptation by ERG) and structural (reduced fundus flecks) improvements.

Summarizing the clinical trials so far, it can be concluded that there are several different treatments for *ABCA4* disease under development, yet none of them so far

revealed itself to be the ideal therapeutic intervention. This is further illustrated by the many pre-clinical studies that are currently ongoing, to identify novel therapeutic strategies as well as improved clinical diagnosis.

Conclusions

In this study, we aimed to identify the underlying cause for the missing heritability in STGD1 by sequencing the complete *ABCA4* gene employing different methods. An integrated genomics and transcriptomics approach enabled us to generate a complete *ABCA4* variation landscape. Moreover, for variants leading to PEs, AON-based splice modulations provided encouraging data for further development of therapeutic studies. Overall, this *ABCA4*-oriented study can be regarded as a model for elucidating missing heritability in other autosomal recessive diseases with a recognizable phenotype and with an incomplete molecular diagnosis.

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**: 248-249.

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

Allikmets R, Singh N, Sun H, Shroyer NF, 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.

Apfelstedt-Sylla E, Theischen M, Ruther K, Wedemann H, Gal A, Zrenner E. 1995. Extensive intrafamilial and interfamilial phenotypic variation among patients with autosomal dominant retinal dystrophy and mutations in the human RDS/peripherin gene. *Br J Ophthalmol* **79**: 28-34.

Bakondi B, Lv W, Lu B, Jones MK, Tsai Y, Kim KJ, Levy R, Akhtar AA, Breunig JJ, Svendsen CN et al. 2016. In Vivo CRISPR/Cas9 Gene Editing Corrects Retinal Dystrophy in the S334ter-3 Rat Model of Autosomal Dominant Retinitis Pigmentosa. *Mol Ther* **24**: 556-563.

Baralle M, Skoko N, Knezevich A, De Conti L, Motti D, Bhuvanagiri M, Baralle D, Buratti E, Baralle FE. 2006. NF1 mRNA biogenesis: effect of the genomic milieu in splicing regulation of the NF1 exon 37 region. *FEBS Lett* **580**: 4449-4456.

Bauwens M, Garanto A, Sangermano R, Naessens S, Weisschuh N, De Zaeytijd J, Khan M, Sadler F, Balikova I, Van Cauwenbergh C et al. 2019. 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* doi:10.1038/s41436-018-0420-y.

Bonifert T, Karle KN, Tonagel F, Batra M, Wilhelm C, Theurer Y, Schoenfeld C, Kluba T, Kamenisch Y, Carelli V et al. 2014. Pure and syndromic optic atrophy explained by deep intronic OPA1 mutations and an intralocus modifier. *Brain* **137**: 2164-2177.

Braun TA, Mullins RF, Wagner AH, Andorf JL, Johnston RM, Bakall BB, Deluca AP, Fishman GA, Lam BL, Weleber RG et al. 2013. Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. *Hum Mol Genet* **22**: 5136-5145.

Calarco JA, Superina S, O'Hanlon D, Gabut M, Raj B, Pan Q, Skalska U, Clarke L, Gelinas D, van der Kooy D et al. 2009. Regulation of vertebrate nervous system alternative splicing and development by an SR-related protein. *Cell* **138**: 898-910.

Caleca L, Colombo M, van Overeem Hansen T, Lazaro C, Manoukian S, Parsons MT, Spurdle AB, Radice P. 2019. GFP-Fragment Reassembly Screens for the Functional Characterization of Variants of Uncertain Significance in Protein Interaction Domains of the BRCA1 and BRCA2 Genes. *Cancers (Basel)* **11**.

Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. 2003. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* **31**: 3568-3571.

Cideciyan AV, Jacobson SG, Drack AV, Ho AC, Charng J, Garafalo AV, Roman AJ, Sumaroka A, Han IC, Hochstedler MD et al. 2019. Effect of an intravitreal antisense oligonucleotide on vision in Leber congenital amaurosis due to a photoreceptor cilium defect. *Nat Med* **25**: 225-228.

Coppieters F, Casteels I, Meire F, De Jaegere S, Hooghe S, van Regemorter N, Van Esch H, Matuleviciene A, Nunes L, Meersschaut V et al. 2010. Genetic screening of LCA in Belgium: predominance of CEP290 and identification of potential modifier alleles in AHI1 of CEP290-related phenotypes. *Hum Mutat* **31**: E1709-1766.

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

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 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.

Dehainault C, Michaux D, Pages-Berhouet S, Caux-Moncoutier V, Doz F, Desjardins L, Couturier J, Parent P, Stoppa-Lyonnet D, Gauthier-Villars M et al. 2007. A deep intronic mutation in the RB1 gene leads to intronic sequence exonisation. *Eur J Hum Genet* **15**: 473-477.

den Hollander AI, Koenekoop RK, Yzer S, Lopez I, Arends ML, Voesenek KE, Zonneveld MN, Strom TM, Meitinger T, Brunner HG et al. 2006. Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. *Am J Hum Genet* **79**: 556-561.

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**: e67.

Dhir A, Buratti E. 2010. Alternative splicing: role of pseudoexons in human disease and potential therapeutic strategies. *FEBS J* **277**: 841-855.

Dobkin C, Pergolizzi RG, Bahre P, Bank A. 1983. Abnormal splice in a mutant human beta-globin gene not at the site of a mutation. *Proc Natl Acad Sci U S A* **80**: 1184-1188.

Farkas MH, Grant GR, White JA, Sousa ME, Consugar MB, Pierce EA. 2013. Transcriptome analyses of the human retina identify unprecedented transcript diversity and 3.5 Mb of novel transcribed sequence via significant alternative splicing and novel genes. *BMC Genomics* **14**: 486.

Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. 2011. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat* **32**: 557-563.

Garanto A, Chung DC, Duijkers L, Corral-Serrano JC, Messchaert M, Xiao R, Bennett J, Vandenberghe LH, Collin RW. 2016. In vitro and in vivo rescue of aberrant splicing in CEP290-associated LCA by antisense oligonucleotide delivery. *Hum Mol Genet* **25**: 2552-2563.

Gerard X, Perrault I, Hanein S, Silva E, Bigot K, Defoort-Delhemmes S, Rio M, Munnich A, Scherman D, Kaplan J et al. 2012. AON-mediated Exon Skipping Restores Ciliation in Fibroblasts Harboring the Common Leber Congenital Amaurosis CEP290 Mutation. *Mol Ther Nucleic Acids* **1**: e29.

Hammond SM, Wood MJ. 2011. Genetic therapies for RNA mis-splicing diseases. *Trends Genet* **27**: 196-205.

Han Z, Conley SM, Makkia RS, Cooper MJ, Naash MI. 2012. DNA nanoparticle-mediated ABCA4 delivery rescues Stargardt dystrophy in mice. *J Clin Invest* **122**: 3221-3226.

Imani S, Cheng J, Shasaltaneh MD, Wei C, Yang L, Fu S, Zou H, Khan MA, Zhang X, Chen H et al. 2018. Genetic identification and molecular modeling characterization reveal a novel PROM1 mutation in Stargardt4-like macular dystrophy. *Oncotarget* **9**: 122-141.

Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, Kosmicki JA, Arbelaez J, Cui W, Schwartz GB et al. 2019. Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**: 535-548 e524.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**: 1073-1082.

Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J et al. 2016. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* **44**: D862-868.

Lewis RA, Shroyer NF, Singh N, Allikmets R, Hutchinson A, Li Y, Lupski JR, Leppert M, Dean M. 1999. Genotype/Phenotype analysis of a photoreceptor-specific ATP-binding cassette transporter gene, ABCR, in Stargardt disease. *Am J Hum Genet* **64**: 422-434.

Liquori A, Vache C, Baux D, Blanchet C, Hamel C, Malcolm S, Koenig M, Claustres M, Roux AF. 2016. Whole USH2A Gene Sequencing Identifies Several New Deep Intronic Mutations. *Hum Mutat* **37**: 184-193.

Lu B, Malcuit C, Wang S, Girman S, Francis P, Lemieux L, Lanza R, Lund R. 2009. Long-term safety and function of RPE from human embryonic stem cells in preclinical models of macular degeneration. *Stem Cells* **27**: 2126-2135.

Ma CJ, Lee W, Stong N, Zernant J, Chang S, Goldstein D, Nagasaki T, Allikmets R. 2019. 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* doi:10.1101/mcs.a003624.

Maugeri A, van Driel MA, van de Pol DJR, Klevering BJ, van Haren FJJ, Tijmes N, Bergen AAB, Rohrschneider K, Blankenagel A, Pinckers AJLG 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. *American Journal of Human Genetics* **64**: 1024-1035.

Mayer AK, Rohrschneider K, Strom TM, Glockle N, Kohl S, Wissinger B, Weisschuh N. 2016. Homozygosity mapping and whole-genome sequencing reveals a deep intronic PROM1 mutation causing cone-rod dystrophy by pseudoexon activation. *Eur J Hum Genet* **24**: 459-462.

Mesman RLS, Calleja F, Hendriks G, Morolli B, Misovic B, Devilee P, van Asperen CJ, Vrieling H, Vreeswijk MPG. 2019. The functional impact of variants of uncertain significance in BRCA2. *Genet Med* **21**: 293-302.

Murphy D, Cieply B, Carstens R, Ramamurthy V, Stoilov P. 2016. The Musashi 1 controls the splicing of photoreceptor-specific exons in the vertebrate retina. *PLoS genetics* **12**: e1006256.

Osoegawa K, de Jong PJ. 2004. BAC library construction. *Methods Mol Biol* **255**: 1-46.

Parfitt DA, Lane A, Ramsden CM, Carr AJ, Munro PM, Jovanovic K, Schwarz N, Kanuga N, Muthiah MN, Hull S et al. 2016. Identification and Correction of Mechanisms Underlying Inherited Blindness in Human iPSC-Derived Optic Cups. *Cell Stem Cell* **18**: 769-781.

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

Popp MW, Maquat LE. 2013. Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu Rev Genet* **47**: 139-165.

Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. *J Comput Biol* **4**: 311-323.

Richards AJ, McNinch A, Whittaker J, Treacy B, Oakhill K, Poulson A, Snead MP. 2012. Splicing analysis of unclassified variants in COL2A1 and COL11A1 identifies deep intronic pathogenic mutations. *Eur J Hum Genet* **20**: 552-558.

Rio Frio T, McGee TL, Wade NM, Iseli C, Beckmann JS, Berson EL, Rivolta C. 2009. A single-base substitution within an intronic repetitive element causes dominant retinitis pigmentosa with reduced penetrance. *Hum Mutat* **30**: 1340-1347.

Riveiro-Alvarez R, Xie YA, Lopez-Martinez MA, Gambin T, Perez-Carro R, Avila-Fernandez A, Lopez-Molina MI, Zernant J, Jhangiani S, Muzny D et al. 2015. New mutations in the RAB28 gene in 2 Spanish families with cone-rod dystrophy. *JAMA Ophthalmol* **133**: 133-139.

Ruan GX, Barry E, Yu D, Lukason M, Cheng SH, Scaria A. 2017. CRISPR/Cas9-Mediated Genome Editing as a Therapeutic Approach for Leber Congenital Amaurosis 10. *Mol Ther* **25**: 331-341.

Runhart EH, Valkenburg D, Cornelis SS, Khan M, Sangermano R, Albert S, Bax NM, Astuti GDN, Gilissen C, Pott JR et al. 2019. Late-Onset Stargardt Disease Due to Mild, Deep-Intronic ABCA4 Alleles. *Invest Ophthalmol Vis Sci* **60**: 4249-4256.

Sakharkar MK, Chow VT, Kangueane P. 2004. Distributions of exons and introns in the human genome. *In Silico Biol* **4**: 387-393.

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

Sangermano R, Garanto A, Khan M, Runhart EH, Bauwens M, Bax NM, van den Born LI, Khan MI, Cornelis SS, Verheij J et al. 2019. Deep-intronic ABCA4 variants explain missing

heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. *Genet Med* doi:10.1038/s41436-018-0414-9.

Schwarz JM, Cooper DN, Schuelke M, Seelow D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature Methods* **11**: 361-362.

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**: 861-872.

Thomas CE, Ehrhardt A, Kay MA. 2003. Progress and problems with the use of viral vectors for gene therapy. *Nat Rev Genet* **4**: 346-358.

van den Hurk JAJM, van de Pol DJR, Wissinger B, van Driel MA, Hoefsloot LH, de Wijs IJ, van den Born LI, Heckenlively JR, Brunner HG, Zrenner E et al. 2003. Novel types of mutation in the choroideremia (CHM) gene: a full-length L1 insertion and an intronic mutation activating a cryptic exon. *Human Genetics* **113**: 268-275.

Vaz-Drago R, Custodio N, Carmo-Fonseca M. 2017. Deep intronic mutations and human disease. *Hum Genet* **136**: 1093-1111.

Vazquez-Dominguez I, Garanto A, Collin RWJ. 2019. Molecular Therapies for Inherited Retinal Diseases-Current Standing, Opportunities and Challenges. *Genes (Basel)* **10**.

Vorechovsky I. 2010. Transposable elements in disease-associated cryptic exons. *Hum Genet* **127**: 135-154.

Webb TR, Parfitt DA, Gardner JC, Martinez A, Bevilacqua D, Davidson AE, Zito I, Thiselton DL, Ressa JH, Apergi M et al. 2012. Deep intronic mutation in OFD1, identified by targeted genomic next-generation sequencing, causes a severe form of X-linked retinitis pigmentosa (RP23). *Hum Mol Genet* **21**: 3647-3654.

Weisschuh N, Sturm M, Baumann B, Audo I, Ayuso C, Bocquet B, Branham K, Brooks BP, Catala-Mora J, Giorda R et al. 2019. Deep-intronic variants in CNGB3 cause achromatopsia by pseudoexon activation. *Hum Mutat* doi:10.1002/humu.23920.

Xie YA, Lee W, Cai C, Gambin T, Noupuu K, Sujirakul T, Ayuso C, Jhangiani S, Muzny D, Boerwinkle E et al. 2014. New syndrome with retinitis pigmentosa is caused by nonsense mutations in retinol dehydrogenase RDH11. *Hum Mol Genet* **23**: 5774-5780.

Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* **11**: 377-394.

Yu W, Wu Z. 2018. In Vivo Applications of CRISPR-Based Genome Editing in the Retina. *Front Cell Dev Biol* **6**: 53.

Zaneveld J, Siddiqui S, Li H, Wang X, Wang H, Wang K, Li H, Ren H, Lopez I, Dorfman A et al. 2015. Comprehensive analysis of patients with Stargardt macular dystrophy reveals new genotype-phenotype correlations and unexpected diagnostic revisions. *Genet Med* **17**: 262-270.

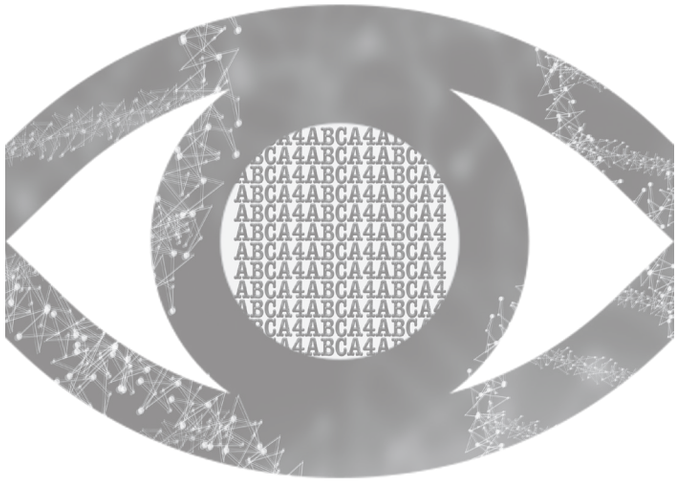
Zernant J, Lee W, Collison FT, Fishman GA, Sergeev YV, Schuerch K, Sparrow JR, Tsang SH, Allikmets R. 2017. Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration. *J Med Genet* **54**: 404-412.

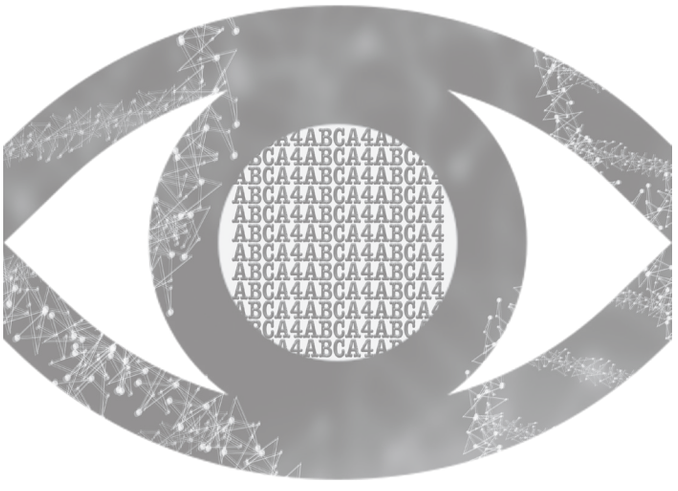
Zernant J, Lee W, Nagasaki T, Collison FT, Fishman GA, Bertelsen M, Rosenberg T, Gouras P, Tsang SH, Allikmets R. 2018. Extremely hypomorphic and severe deep intronic variants in the ABCA4 locus result in varying Stargardt disease phenotypes. *Cold Spring Harb Mol Case Stud* **4**.

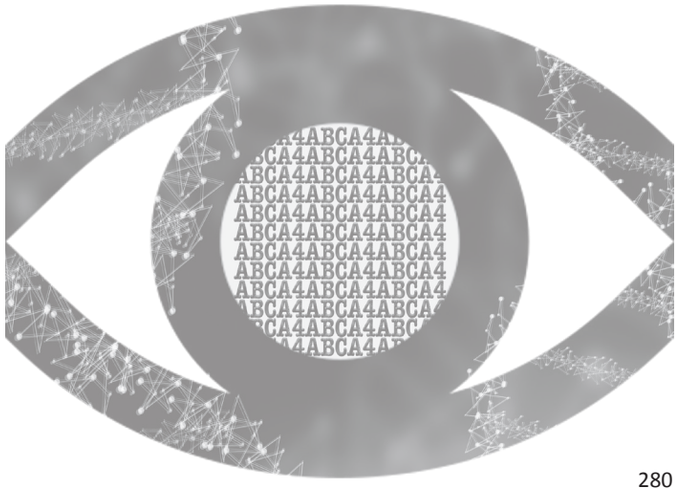
Zernant J, Xie YA, Ayuso C, Riveiro-Alvarez R, Lopez-Martinez MA, Simonelli F, Testa F, Gorin MB, Strom SP, Bertelsen M et al. 2014. Analysis of the ABCA4 genomic locus in Stargardt disease. *Hum Mol Genet* **23**: 6797-6806.

Chapter 10

Summary







ABCA4 is the most frequently mutated inherited retinal disease-associated gene, and mutations are mainly associated with Stargardt disease (STGD1). The severity of the macular phenotype is determined by the combination of mutations, leading to STGD1, autosomal recessive cone-rod dystrophy or autosomal recessive retinitis pigmentosa. Missing heritability until recently was one of the characteristics of *ABCA4*-associated STGD1. In ~25-35% of STGD1 probands, only one or no pathogenic *ABCA4* variants were found. In this study, we employed several genotyping methods beyond Sanger sequencing of coding elements, such as Haloplex- and smMIPs-based complete *ABCA4* gene sequencing, to detect SNVs as well as SVs. Special emphasis was given to the functional characterization of non-coding variants, in particular those that are predicted to lead to splicing defects. *In vitro* midigene splice assays were carried out in HEK293T cells. In some cases these assays were expanded with studies in patient-derived fibroblasts or more sophisticated photoreceptor precursor cells. This combination of extensive genotyping and functional studies allowed us to provide the genetic diagnosis for 98% of Dutch STGD1 cases. Finally, a proper understanding of the pathogenic mechanisms underlying novel *ABCA4* non-coding variants enabled us to successfully apply therapeutic approaches aiming to correct splicing defects, ultimately providing new avenues for treatment of patients affected by STGD1.

Chapter 1 provides a general introduction subdivided into four sections. **Chapter 1.1** describes background information on visual perception and the structure of the eye and retina, while **chapter 1.2** focuses on the genetics of inherited retinal diseases by describing various molecular diagnostic approaches that have gradually moved from coding only to non-coding sequencing methods. In addition, it emphasizes the importance of different databases that are necessary for the classification and expansion of the knowledge for newly identified variants by sharing information. **Chapter 1.3** describes the genetics of STGD1, in particular the spectrum of *ABCA4*-associated phenotypes, the function of the ABCA4 protein and the importance of proper clinical characterization due to the presence of phenocopies. Furthermore, it describes the role of mild variants in disease progression, i.e., c.5603A>T, mostly implicated in late onset STGD1. It recapitulates the *in vitro* and *in vivo* functional studies performed on coding and non-coding *ABCA4* variants, with a major emphasis on the complexity of pre-mRNA splicing. This includes the assessment of the functional consequences of splice variants by either employing *in vitro* splice assays or performing transcript analysis in patient-derived somatic non-retinal or photoreceptor precursor cells. Finally, this chapter provides the state-of-the-art of therapeutic strategies for STGD1 and describes, among others, antisense oligonucleotide-based splicing modulation which was extensively employed during

our studies. **Chapter 1.4** describes the aim and outline of this thesis.

Chapter 2 illustrates the design of a novel strategy in which a bacterial artificial chromosome clone containing the full genomic sequence of *ABCA4* was employed to generate 29 overlapping multi-exon wild-type midigenes, splice vectors of varying lengths (up to 11.7 kb) covering almost the entire gene. This study started with the observation that, when using small minigenes lacking the proper genomic context to investigate the splicing effect of the noncanonical splice site (NCSS) variant c.5714+5G>A, *in vitro* results did not correlate with the expected effects based on genotype-phenotype correlations. Since a significant number of misinterpreted variants are NCSS variants, we employed wild-type midigenes to perform site-directed mutagenesis and systematically analyzed the effect of all 44 reported and three novel NCSS variants on *ABCA4* pre-mRNA splicing. For 44 NCSS variants, we revealed a plethora of splicing defects, among which multi-exon skipping events, exon elongations and intron retentions, and quantitative analysis of these splicing defects allowed an improved classification of the severity of these variants. In conclusion, we created a toolbox for fast and robust *in vitro* analysis of splice variants in *ABCA4*, with potential applications for testing coding and non-coding variants in all human genes.

Chapter 3 describes the identification and characterization of one known and five novel non-coding variants associated with STGD1, and provides evidence for an effective therapeutic antisense oligonucleotide (AON) strategy. Targeted *ABCA4* locus sequencing was performed in 36 mono-allelic STGD1 patients and rare candidate intronic variants were selected based upon their *in silico* prediction affecting splicing or enrichment in STGD1 samples. We detected three known causal deep-intronic variants, and five novel non-coding variants in 24 probands and their family members. Functional characterization of six deep-intronic variants was performed by midigene-based splice assays and patient fibroblasts. Five variants (c.769-784C>T, c.859-506G>C, c.1937+435C>G, c.4539+1100A>G, c.4539+1106C>T) led to partial or complete pseudoexon (PE) inclusion, while the c.4253+43G>A variant mainly resulted in skipping of *ABCA4* exon 28. The most prevalent variants, c.769-784C>T and c.4253+43G>A, together found in 16 patients, showed a partial effect, and interestingly, were associated with a late age at disease onset. Finally, AONs targeting the five novel PEs and c.4253+43G>A were able to significantly correct the splicing defects *in vitro* and, for two variants, in patient-derived fibroblasts.

Chapter 4 shows the employment of a cost-effective sequencing method for *ABCA4* exons and selected regions carrying 14 known causal deep-intronic variants. Targeted

sequencing was performed in 411 STGD1 cases with one or no causal *ABCA4* variant by using smMIPs-based sequencing. In 155/411 (38%) probands, two pathogenic or likely pathogenic variants were identified. Among them, 62 alleles were carrying deep-intronic variants, of these c.4253+43G>A (n=28) and c.5196+1137G>A (n=13) were the most frequent. Next to the known deep-intronic variants, a novel deep-intronic variant c.4539+2065C>G (p.[Arg1514Lysfs*35,=]), was identified. Moreover, splice defects were established for two previously known (c.4849G>A and c.3608G>A) and two novel (c.4128G>C and c.4539G>A) NCSS variants employing *in vitro* splice assays. In conclusion, this study establishes the importance of the employment of a cost-effective mutation detection method in STGD1 cases.

Chapter 5 shows the genetic and clinical characteristics of *ABCA4* intron 36 variants. Four variants (c.5196+1013A>G, c.5196+1056A>G, c.5196+1137G>A and c.5196+1216C>A) were found to result in aberrant splicing of *ABCA4*, i.e., the inclusion of PEs, albeit to a different extent. Intriguingly, the splicing defect caused by the recurrent c.5196+1137G>A variant strongly increased upon differentiation of patient-derived induced pluripotent stem cells towards a retinal fate. All splicing defects could be rescued by the delivery of AONs that were designed to specifically target the corresponding PE insertion. These data illustrate the importance of intronic variants in *ABCA4* underlying STGD1, and expand the therapeutic possibilities to overcome splicing defects caused by these mutations.

Chapter 6 sheds light on the pathogenic role of previously published rare NCSS and deep-intronic *ABCA4* variants by employing midigene-based splice assays in HEK293T cells. Rare variants were selected based upon splice site predictions, i.e., the disruption of splicing of regular *ABCA4* exons or the erroneous inclusion of intronic sequences (i.e., PEs), as well as assessing the role of conserved motifs such as splicing enhancers and silencers, demonstrating the importance of both splice-site sequences and conserved motifs when selecting the variants.

Chapter 7 characterizes *ABCA4*-associated STGD1 disease as a model disease to investigate the missing causal variants in other monogenic hereditary conditions. Following up on the studies conducted in **chapter 4**, the entire 128-kb *ABCA4* locus was sequenced using a smMIPs-based semi-automated method. Analysis of 1,054 unsolved STGD and STGD-like probands issued from an international collaboration, resulted in bi-allelic variants in 448 probands. Of these, 117 alleles carried 27 deep-intronic variants. Among these, 13 deep-intronic variants were novel, and led to PE insertions (n=10) or exon elongations (n=3). Next to the SNVs, 11 unique SVs were identified in

16 alleles by employing a smMIPs read depth-based CNV analysis script. Moreover, uniparental isodisomy of chromosome 1 was identified in one proband. Integrated complete gene sequencing combined with transcript analysis identified pathogenic deep-intronic and SVs in 26% of bi-allelic cases not solved previously by sequencing of coding regions. This strategy serves as a model study that can be applied to other inherited diseases in which only one or a few genes are involved in the majority of cases.

Chapter 8 describes the role of both the splice acceptor and donor sites in proper recognition of exon-intron boundaries for correct splicing, which we coined 'splicing interdependency'. This study was based on the observation that some NCSS variants show a stronger splice defect than expected based on their predicted effects. This suggested that other sequence motifs influence the outcome. To test this hypothesis we investigated whether splicing defects due to human inherited disease-associated variants in NCSS sequences of the splice acceptor or donor could be rescued by strengthening the splice donor or acceptor site of the same exon in several genes, i.e., *ABCA4*, *DMD* and *TMC1.* In each gene, *in vitro* splice assays showed the complete or partial rescue of exon skipping caused by the aforementioned NCSS variants. This study clearly showed that splicing interdependency is an important principle for the classification of severity of NCSS variants associated with disease, and may open new opportunities for treatments.

Chapter 9 provides a general discussion of the main findings in this thesis and their implications towards improved diagnosis of STGD1. The application of different sequencing approaches as well as their advantages and disadvantages are explained. Noncoding variants were tested by midigenes splice assays using HEK293T cells and a comparison of the use of different cell lines was provided. Next to that, the need of an of accurate splicing prediction tool was conferred, by comparing various prediction algorithms, highlighting their importance in functional analysis. Finally, potential reasons for missing *ABCA4* variants were debated and it was hypothesized that missing alleles in STGD-like probands can be found in other maculopathy genes, due to presence of phenocopies.

Hoofdstuk 10

Samenvatting

ABCA4 is het meest gemuteerde gen betrokken bij erfelijke netvliesziekten, en mutaties zijn voornamelijk geassocieerd met de ziekte van Stargardt (STGD1). De ernst van het macula fenotype wordt bepaald door de combinatie van mutaties die leiden tot STGD1, autosomaal recessieve kegel-staaf dystrofie of autosomaal recessieve retinitis pigmentosa. Het ontbreken van *ABCA4* mutaties tot voor kort was een van de kenmerken van STGD1. In ~25-35% STGD1-probanden kon slechts één van de twee, of zelfs geen enkele van de verwachte pathogene *ABCA4*-varianten gevonden worden. In deze studie hebben we verschillende genotyperings-methoden gebruikt die verder gaan dan Sanger-sequencing van de eiwitcoderende elementen, zoals Haloplex- en smMIPs-gebaseerde complete *ABCA4*-sequencing, om zowel afzonderlijke nucleotide varianten als ook structurele varianten te detecteren. Speciale nadruk werd gelegd op de functionele karakterisering van niet-coderende varianten, in het bijzonder die waarvan wordt voorspeld dat ze leiden tot splicing defecten, door middel van een *in vitro* midigen assay. Waar mogelijk werden deze *in vitro* assays gekoppeld aan studies in van patiënten afgeleide fibroblasten of fotoreceptor precursorcellen. Deze combinatie van uitgebreide genotypering en functionele onderzoeken stelde ons in staat om de genetische diagnose te stellen voor 98% van de Nederlandse STGD1-patiënten. Tenslotte stelt een beter begrip van de pathogene mechanismen die ten grondslag liggen aan nieuwe *ABCA4* varianten ons in staat om therapeutische strategieën te ontwikkelen die gericht zijn op het herstellen van correcte splicing en uiteindelijk nieuwe mogelijkheden bieden voor de behandeling van patiënten met STGD1.

Hoofdstuk 1 geeft een algemene inleiding onderverdeeld in vier secties. **Hoofdstuk 1.1** geeft achtergrondinformatie over de visuele perceptie en de structuur van het oog en het netvlies, terwijl **hoofdstuk 1.2** zich richt op de genetica van erfelijke netvliesaandoeningen. Hier beschrijf ik verschillende moleculaire diagnostische benaderingen, die geleidelijk overgaan van het analyseren van alleen de coderende stukken DNA naar ook de niet-coderende sequenties. Bovendien wordt het belang van verschillende databanken benadrukt die nodig zijn voor de classificatie en uitbreiding van het kennisspectrum van nieuw geïdentificeerde varianten in *ABCA4*, middels het delen van deze informatie. **Hoofdstuk 1.3** beschrijft de genetica van STGD1, in het bijzonder het spectrum van *ABCA4*-geassocieerde fenotypes, de functie van het ABCA4 eiwit en het belang van goede klinische karakterisering i.v.m. de aanwezigheid van fenokopieën. Verder beschrijft het de rol van milde varianten, specifiek c.5603A>T, in het veroorzaken van STGD1 op latere leeftijd. Het recapituleert de *in vitro* en *in vivo* functionele studies die zijn uitgevoerd op coderende en niet-coderende *ABCA4*-varianten, met grote nadruk op de complexiteit van pre-mRNA-splicing. Dit omvat de

beoordeling van de functionele gevolgen van splice-varianten door ofwel *in vitro* splice-assays te gebruiken, of transcript analyse uit te voeren in van de patiënt afgeleide somatische niet-retinale of fotoreceptor-voorlopercellen. Tenslotte bespreekt dit hoofdstuk de huidige therapeutische strategieën voor STGD1 en beschrijft het, onder andere, antisense oligonucleotide-gebaseerde splicing modulatie die uitgebreid werd gebruikt tijdens onze studies. **Hoofdstuk 1.4** beschrijft het doel en de opzet van dit proefschrift.

Hoofdstuk 2 beschrijft het ontwerp van een nieuwe strategie waarbij een bacteriële kunstmatige chromosoom kloon met de volledige genomische sequentie van *ABCA4* werd gebruikt om 29 overlappende multi-exon wild-type midigenen te genereren, d.w.z. splice vectoren met een verschillende lengte (tot 11.7 kb) die bijna het hele gen omvatten. Deze studie begon met de waarneming dat, bij gebruik van kleine minigenen die de juiste genomische context missen om het splice effect van de 'noncanonical splice site' (NCSS) variant c.5714+5G>A te onderzoeken, de *in vitro* resultaten niet correleerden met het verwachte effect gebaseerd of fenotype-genotypen correlaties. Aangezien een aanzienlijk aantal verkeerd geïnterpreteerde varianten NCSS-varianten zijn, hebben we wild-type midigenen gebruikt om mutagenese uit te voeren en systematisch het effect van alle 44 gerapporteerde en drie nieuwe NCSS-varianten op *ABCA4* pre-mRNA splicing geanalyseerd. Voor 44 NCSS-varianten hebben we diverse splicing defecten gevonden, waaronder multi-exon-skipping-gebeurtenissen, exon-verlengingen en intron-retenties. Kwantitatieve analyses van deze splicing defecten maakten een verbeterde classificatie van de ernst van deze varianten mogelijk. Concluderend kunnen we stellen dat we een gereedschapskist gecreëerd hebben voor snelle en robuuste *in vitro* analyse van splice-varianten in *ABCA4*. Deze methode kan in principe toegepast worden voor het testen van coderende en niet-coderende varianten in alle menselijke genen.

Hoofdstuk 3 beschrijft de identificatie en karakterisering van zes nieuwe niet-coderende varianten geassocieerd met STGD1, en biedt bewijs voor een effectieve therapeutische antisense oligonucleotide (AON) strategie. Gerichte *ABCA4*-locussequencing werd uitgevoerd bij 36 mono-allelische STGD1-patiënten en zeldzame kandidaat intronische varianten werden geselecteerd op basis van hun *in silico* splicing voorspelling of verrijking in STGD1-monsters. We ontdekten drie bekende causale diep-intronische varianten, en vijf nieuwe niet-coderende varianten in 24 probanden en hun familieleden. Functionele karakterisering van zes varianten werd uitgevoerd door middel van op midigenen gebaseerde splicing assays en fibroblasten van patiënten. Vijf nieuwe varianten (c.769-784C>T, c.859-506G>C, c.1937+435C>G,

c.4539+1100A>G en c.4539+1106C>T) leidden tot gedeeltelijke of volledige pseudoexon (PE) inclusie, terwijl de c.4253+43G>A variant resulteerde in het overslaan van *ABCA4* exon 28 in het mRNA. De meest voorkomende varianten, c.769-784C>T en c.4253+43G>A, samen gevonden bij 16 patiënten, toonden een gedeeltelijk effect en werden interessant genoeg geassocieerd met een aanvang van de ziekte op latere leeftijd. Tenslotte waren AONs gericht op vijf nieuwe PEs en c.4253+43G>A in staat de splicing defecten *in vitro* of in van patiënten afgeleide fibroblasten, significant te corrigeren.

Hoofdstuk 4 beschrijft de ontwikkeling van een kosteneffectieve sequentiemethode voor *ABCA4* exonen en geselecteerde regio's met 14 bekende causale diep-intronische varianten. Gerichte sequencing werd uitgevoerd in 411 STGD1 patiënten met één of geen *ABCA4* allel m.b.v. smMIPs-gebaseerde sequencing. In 155/411 (38%) probanden werden twee pathogene of waarschijnlijk pathogene varianten geïdentificeerd. Zij droegen 62 allelen met diep-intronische varianten, waarvan c.4253+43G>A (n = 28) en c.5196+1137G>A (n = 13) de meest voorkomende waren. Naast de bekende diep-intronische varianten, werd ook een nieuwe diep-intronische variant, c.4539+2065C>G (p.[Arg1514Lysfs*35,=]), geïdentificeerd. Bovendien werden splicing defecten vastgesteld voor twee eerder beschreven (c.4849G>A en c.3608G>A) en twee nieuwe (c.4128G>C en c.4539G>A) NCSS-varianten, middels *in vitro* splice assays. Dit onderzoek onderstreept het belang van de inzet van een kosteneffectieve mutatie detectiemethode bij STGD1.

Hoofdstuk 5 beschrijft de genetische en klinische kenmerken van *ABCA4* intron 36-varianten. In totaal bleken vier varianten (c.5196+1013A>G, c.5196+1056A>G, c.5196+1137G>A, c.5196+1216C>A) te resulteren in een afwijkende splicing van *ABCA4*, d.w.z. de introductie van pseudoexon (PEs). Intrigerend is dat het splicing defect veroorzaakt door de veel voorkomende variant c.5196+1137G>A duidelijker tot uiting kwam na differentiatie van patiënt-afgeleide geïnduceerde pluripotente stamcellen naar een netvliesachtig celtype. Alle splice defecten konden worden hersteld door de toediening van AONs die speciaal waren ontworpen om de diverse PE inserties te blokkeren. Deze gegevens illustreren het belang van intronische varianten als oorzaak bij STGD1, en breiden de therapeutische mogelijkheden uit om splice defecten veroorzaakt door deze mutaties te corrigeren.

Hoofdstuk 6 beschrijft de pathogene rol van eerder beschreven zeldzame NCSS en diep-intronische *ABCA4*-varianten door middel van op midigenen gebaseerde splice assays in HEK293T-cellen. Zeldzame varianten werden geselecteerd op basis van de *in*

silico voorspelling van splice sites, d.w.z. hun effect op normale splicing van *ABCA4* exonen als ook de mogelijkheid dat intronische sequenties in het mRNA gespliced zouden kunnen worden. Tevens werd het effect van varianten op zogenaamde splice enhancers en silencers beoordeeld, die eveneens een rol kunnen spelen bij (pseudo)exon definitie.

Hoofdstuk 7 beschrijft *ABCA4*-geassocieerde STGD1-ziekte als een modelziekte om de ontbrekende oorzakelijke varianten in andere monogene erfelijke aandoeningen te onderzoeken. In navolging van de studies in **hoofdstuk 4** werd nu het volledige 128 kb *ABCA4*-locus gesequenced met behulp van een smMIP-gebaseerde half-automatische methode. Analyse van 1.054 onopgeloste STGD- en STGD-achtige probanden verkregen door een internationale samenwerking, resulteerde in bi-allelische varianten in 448 probanden en één oorzakelijke variant in 174 probanden. Deze droegen 117 allelen met 27 verschillende diep-intronische varianten. Hieronder bevonden zich 13 nieuwe diep-intronische varianten, die leidde tot pseudoexon (PE) inserties (n=10) of exon-verlengingen (n=3). Naast deze SNVs, werden 11 unieke SVs geïdentificeerd in 16 allelen, door gebruik te maken van een CNV-analyse script dat gebruik maakt van het aantal smMIPs reads. Bovendien werd uniparentale isodisomie van chromosoom 1 geïdentificeerd in één proband. Geïntegreerde complete gensequencing gecombineerd met transcriptanalyse, identificeerde pathogene diep-intronische en SVs in 26% van de bi-allelische patiënten die niet eerder waren opgelost door sequencing van enkel de eiwitcoderende gebieden. Deze strategie dient als een modelstudie die kan worden toegepast op andere erfelijke ziekten waarbij in de meeste gevallen slechts één of enkele genen betrokken zijn.

Hoofdstuk 8 beschrijft de rol van zowel de splice-acceptor als de donor-splice site bij de juiste herkenning van exon-intron grenzen voor correcte splicing, de zogenaamde ‘splicing-afhankelijkheid’. Deze studie was gebaseerd op de waarneming dat sommige NCSS-varianten een sterker splicing defect vertonen dan werd verwacht op basis van hun voorspelde effecten. Dit suggereerde dat andere sequentiemotieven de uitkomst beïnvloeden. Om deze hypothese te testen, hebben we onderzocht of splicing defecten als gevolg van ziekte gerelateerde varianten in sequenties van splice-acceptorplaatsen in *ABCA4*, *DMD* en *TMC1*, konden worden teniet gedaan door de respectievelijke splice-donorsite van hetzelfde exon in elk gen te versterken. *In vitro* splicing assays toonden volledig of gedeeltelijke herstel van exon-skipping veroorzaakt door de bovengenoemde NCSS-varianten. Uit deze studie kan de conclusie worden getrokken dat een goede beoordeling van het effect van een NCSS variant alleen mogelijk is door

ook de sterkte van de andere splice site van hetzelfde exon te beoordelen. Dit kan nieuwe mogelijkheden voor behandelingen bieden.

Hoofdstuk 9 geeft een algemeen overzicht van alle belangrijke bevindingen in dit proefschrift en hun implicaties voor een betere diagnose van STGD1. De toepassing van verschillende sequencing-benaderingen en hun voor- en nadelen werden besproken. Niet-coderende varianten werden getest door midigenes splice-assays in HEK293T-cellen, en een vergelijking van het gebruik van verschillende cellijnen werd gemaakt. Na het vergelijken van verschillende voorspellingsalgoritmen werd het duidelijk dat er een behoefte is aan een meer nauwkeurig hulpmiddel voor het voorspellen van splicing defecten. Tenslotte werden mogelijke redenen voor het niet kunnen opsporen van bepaalde *ABCA4*-varianten besproken en werd de hypothese geopperd dat ontbrekende allelen in STGD-probanden kunnen worden gevonden in andere maculopathie-geassocieerde genen, vanwege de aanwezigheid van fenokopieën.

Data management

Type of data	Subject to privacy (Yes/No)	Way of anonymization	Storage
DNA of patients used in this study	Yes	A DNA number was assigned to each individual by the cell culture facility of the Department of Human Genetics.	All DNA samples were stored at the cell culture facility of the Department of Human Genetics. Contact person for the DNA samples Saskia van der Velde-Visser, saskia.vandervelde-visser@radboudumc.nl
Patient-derived cell lines	Yes	All patients received an untraceable number, identity of the patients was only known to the concerning physician.	Cell lines were frozen in liquid nitrogen and stored in an assigned -80^{o}C freezer at the Department of Human Genetics. After registration at cell culture facility each cell line received a unique DNA number. The DNA number used for experiments can be found in the lab journal of Mubeen Khan and in Labguru account under the project name_ABCA4. https://radboudumc.labguru.com/knowledge/projects/1965 Contact person to find the cell lines in the assigned freezer: Saskia van der Velde-Visser, saskia.vandervelde-visser@radboudumc.nl
ABCA4 targeted sequencing by Haloplex	No	Samples in the dataset were already anonymized.	All the relevant data files can be found at H-drive location: H:\GR Theme groups\05 PI Group Frans Cremers\02 Other shared folders\04-ABCA4\6_Ghent haloplex
ABCA4 targeted sequencing smMIPs	No	Samples in the dataset were already anonymized.	All the relevant data files can be found at T-drive location: T:\PIgroup-Frans-Cremers\ABCA4 smMIPs
Genotyping data for SNPs	No	Samples in the dataset were already anonymized.	Results per SNP can be found at: H:\GR Theme groups\05 PI Group Frans Cremers\06 Manuscripts\ABCA4 intron 36\3_Dutch V1 patients_data

Plasmids	No	N.a.	Plasmids were stored in an assigned -80 °C freezer at the Department of Human Genetics with traceable numbers. The raw sequencing data for all wild-type splicing constructs (WT_BA) mutant splicing BA constructs (BA_c.), was uploaded in NCBI BioProject (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA417900. All plasmids and their details are listed in the Gene Therapy database file at pGT location: H:\GR Theme groups\10 PI Group Rob Collin\05 Primer_Vector_AAV_AON Database
All files for the publications presented in this study	No	N.a.	All files can be found at the H-drive of the Human Genetics department, at the location H:\GR Theme groups\05 PI Group Frans Cremers\06 Manuscripts

N.a., not applicable.

Acknowledgements

First of all, I am very thankful to The Netherlands for hosting me during these years. Nijmegen felt like home to me and I have always felt very welcome. My stay of four and a half years at the Radboudumc and the successful completion of my PhD studies would not have been possible without the help and support from many people, friends, family and colleagues, to whom I here have the opportunity to show my gratitude.

To my supervisors. **Prof. Dr. Frans P.M. Cremers**, thank you for your continuous support, motivation, enthusiasm, and immense knowledge. Your valuable supervision helped me to grow both at the professional and personal level. I could not have imagined having a better advisor and mentor for my PhD study. **Dr. Rob W.J. Collin**, thank you for the time you dedicated to my supervision as well as to my thesis. I have always admired your creativity and encouraging attitude. **Dr. Alejandro Garanto**, thank you very much for all the support and encouraging words I received from you throughout my PhD. I always value your willingness to help others and at the same time great commitment to the research. It has been great working with you.

To the members of the Blindness Genetics workgroup, present and past: Claire-Marie, Duaa, Ellen, Femke, Janine, Ketan, Marco, Marijke, Riccardo, Stéphanie, Silvia, Susanne, Suzanne, Tomas, Victor, Zeinab and Zelia.

Claire-Marie, we started working together as colleagues but ended up being sisters. It was a great pleasure working with you in a team. Working with you has been a pleasure, and I have learned a lot from you, both at the professional and personal level. Thanks a lot for all your sisterly advice and kind words. Hopefully, this journey will keep on going ☺ ☺.

Duaa, you joined our group as an intern and later we developed a strong friendship, which I am sure will grow deeper with time. I admire your passion for research and wish you a lot of success in the future.

Femke, you have been the perfect right-hand support in the lab, always there to help me with a smile. Apart from work I also look forward to our coffee breaks. I have always admired your openness and knowledge.

Stéphanie, I have always felt very blessed to know you. When in 2016 I first joined the group you were the only person in the group with whom I could always have small chats. I tried my first **''**Kaas boterham**''** in a farewell lunch for you. You are always an inspiration to me. I wish you lots of good luck for the future.

Silvia, it has always been fun to talk to you. Thanks for all your encouraging words.

Susanne (former group member ☺, now you have your own group), thanks a lot for always being there for me. For all your nice and encouraging words, whenever I needed to talk, I could always count on you.

Riccardo, OMG! What an amazing journey we had together. From never talking for the first six months (when I joined the group in 2016), to developing a strong bond from very closely working colleagues to becoming sister and brother. We always shared a great and unstoppable enthusiasm for research which kept us going. I am so thankful for this opportunity which allowed me to meet you and learn a lot from you.

Zeinab, thanks for all the support and the nice words I received from you, as well as thanks for all that I have learned from you. Over the years, I have seen you putting lots of effort and commitment in your studies and I am sure that all these sacrifices (long travelling) will soon materialize into a successful outcome.

Suzanne and **Janine**, it was a pleasure to meet you, and I wish you all the best in your careers. A special thanks to **Ellen** and **Marijke**, for all your help and small talks (whenever I was missing home). **Marco**, **Ketan**, **Tomas**, **Victor** and **Zelia**, it was a pleasure to know you all. I wish you all the best for your future endeavors.

All members of workgroup of Rob Collin and Alex Garanto, especially, **Anita, Nuria**, **Irene** and **Manon**. Thanks for all the support and encouraging words I received from you, good luck to you all for the future.

The **group leaders of Radboudumc**, in particular **Prof. dr. Han Brunner**, **Prof. dr. Barbara Franke**, **Dr. Christian Gilissen** and **Dr. Alexander Hoischen** for your help and generous words.

Moreover, a big thanks to all the people in the **Genome Diagnostics division**, especially to **Duaa Elmelik**, **Martine van Zweeden** and **Rianne Miller**, who always helped me to fix all problems for smMIPs sequencing, to make my PhD a huge success.

Marc Pieterse, thank you soo much for all your help for the data analysis and sorting/explaining the smMIPs pipeline for my PhD project.

The most friendly/helping **secretariat team** ever known to me (Ineke, Doménique, Miranda, and Dennis). Thanks you so much for all your help. **Doménique**, it was always nice to have a chat with you, wishing you all the very best!

To the **cell culture, pre/post-lab, and GMO friends and colleagues**, in particular Saskia van der Velde-Visser, Marlie Jacobs-Camps, Anita Roelofs, Mariam Aslanyan, Zeineb Bakey, Bjorn Bakker, Sarah de Jong, Mireia Coll Tané, Sarah Foriel, Dinu Anthony, Ideke Lamers, Ilse Eidhof, Marloes Steehouwer, Jaap Oostrik, Lilian Vreede, Remco Makkinje and Angelien Heister. Thank you for your interest, valuable technical and moral support.

To my students, **Eline** and **Jan**, it was a great experience to supervise you, I wish you all the best in your future.

To the international people who worked with me: **Manar Salameh** (from Palestine), **Marta del Pozo** (from Madrid), **Laura Whelan** (from Dublin), and **Claire-Marie** (from Lille) and all the ones I came to know, in particular Babette, Daniel, Esmee, Jeroen, Julio, Priscilla, Renée, Renske, Roos, Rosanne and Ting. I wish you all the best for the next steps. I am sure you will do great.

To all the wonderful collaborators, **Ala AlTalbishi**, Jerusalem; **Carmen Ayuso**, Madrid; **Sandro Banfi**, Naples; **Tamar Ben-Yosef**, Haifa; **Claire-Marie Dhaenens**, Lille; **Ana Fakin**, Ljubljana; **Jane Farrar**, Dublin; **Juliana Maria Ferraz Sallum**, Sao Paulo; **Kaoru Fujinami**, Tokyo; **Micheal Gorin**, Los Angeles; **Smaragda Kamakari,** Athens; **Petra Liskova**, Prague; **Ian MacDonald**, Edmonton; **Monika Oldak**, Warsaw; **Osvaldo Podhajcer**, Buenos Aires; **John de Roach**, Perth; **Lisa Roberts**, Cape Town; **Dror Sharon**, Jerusalem; **Anna Tracewska**, Warsaw; **Andrea Vincent**, Auckland; **Bernhard Weber**, Regensburg; **Carel Hoyng** and **Esmee Runhart**, Nijmegen; **Camiel Boon**, Amsterdam & Leiden and **Ingeborgh van den Born**, Rotterdam, without your valuable input, this would not be achievable, thanks to all of you.

To my wonderful office mates former (Babette, Daniel, Gido, Iris, Julio, Muriël, Xiuming,) and present (Priscilla, Sarita, Johana and Zeinab). I want to deeply thank all of you for your friendship and for the wonderful time we had together. I wish you all good luck for future.

To my friends in Nijmegen **Sidra Kashif**, **Dolores** and **Zeynep Polat**. Thank you all for your support and encouragement during all these years.

Most importantly, none of this could have happened without my family, which always has been kind and supportive to me. It would be an understatement to say that, as a family, we have experienced many ups and downs in the past four years. Every time I was ready to give up, you did not let me and I am forever grateful. To my beloved parents, **Habib Nawaz Khan Mughal** and **Mussarat Sultana**, who were always there for me. Especially, to my mother who offered her encouragement through skype calls and messages every day – despite sometimes my own limited devotion to the correspondence. My brother, **Ahmed Nawaz Khan** for all his support, encouragement and endless trust. My sisters **Api Talat**, **Baji Tahira**, **Maria** and **Misbah**. All my nieces and nephews (**Anum**, **Ayesha**, **Abdul Sami**, **Hifza**, **Husnain**, **Abdullah**, **Abdul-Rehman**, **Irfan**, **Hassan**, **Momina** and **Usman**). This dissertation stands as a testament to your unconditional love and encouragement.

List of publications

Identification and rescue of splice defects caused by two neighboring deep-intronic *ABCA4*
mutations underlying Stargardt disease
Albert S*, Garanto A*, Sangermano R, **Khan M**, Bax NM, Hoyng CB, Zernant J, Lee W, Allikmets R, Collin RWJ, Cremers FPM.
Am J Hum Genet. 2018; **102:** 517-527

***ABCA4* midigenes reveal the full splice spectrum of all reported non-canonical splice site variants in Stargardt disease**
Sangermano R*, **Khan M***, Cornelis SS, Richelle V, Albert S, Garanto A, Elmelik D, Qamar R, Lugtenberg D, van den Born LI, Collin RWJ, Cremers FPM.
Genome Res. 2018; **28:** 100-110

Identification and Analysis of Genes Associated with Inherited Retinal Diseases
Khan M, Fadaie Z, Cornelis SS, Cremers FPM, Roosing S.
Methods Mol Biol. 2019; **1834**: 3-27

Deep-intronic *ABCA4* variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides
Sangermano R*, Garanto A*, **Khan M***, Runhart EH, Bauwens M, Bax NM, van den Born LI, Khan MI, Cornelis SS, Verheij JBGM, Pott JWR, Thiadens AAHJ, Klaver CCW, Puech B, Meunier I, Naessens S, Arno G, Ana Fakin, Carss K, Raymond L, Webster AR, Dhaenens CM, Stöhr H, Grassmann F, Weber BHF, Hoyng CB, De Baere E, Albert S, Collin RWJ[#], Cremers FPM[#].
Genet Med. 2019; **21**: 1751-1760

***ABCA4*-associated disease as a model for missing heritability in autosomal recessive disorders: novel noncoding splice, cis-regulatory, structural, and recurrent hypomorphic variants**
Bauwens M, Garanto A[$], Sangermano R[$], Naessens S, Weisschuh N, De Zaeytijd J, **Khan M**, Sadler F, Balikova I, Van Cauwenbergh C, Rosseel T, Bauwens J, De Leeneer K, De Jaegere S, Van Laethem T, De Vries M, Carss K, Arno G, Fakin A, Webster AR, de Ravel de l'Argentière TJL, Sznajer Y, Vuylsteke M, Kohl S, Wissinger B, Cherry T, Collin RWJ, Cremers FPM, Leroy BP, De Baere E.
Genet Med. 2019; **21**: 1761-1771

Cost-effective molecular inversion probe-based *ABCA4* sequencing reveals deep-intronic variants in Stargardt disease
Khan M, Cornelis SS, Khan MI, Elmelik D, Manders E, Bakker S, Derks R, Neveling K, van de Vorst M, Gilissen C, Meunier I, Defoort S, Puech B, Devos A, Schulz HL, Stöhr H, Grassmann F, Weber BHF, Dhaenens C-M, Cremers FPM.
Hum Mutat. 2019; **40**: 1749-1759

Identification of splice defects due to non-canonical splice site or deep-intronic variants in *ABCA4*
Fadaie Z, **Khan M**, Del Pozo-Valero M, Cornelis SS, Ayuso C, [*ABCA4* Studygroup: Allikmets R, Bauwens M, Ghofrani M, Gorin MB, Keramatipur M, Simonelli F, Tayebi N, Vincent A, Weisschuh N], Cremers FPM, Roosing S.
Hum Mutat. 2019; **40**: 2365-2376

Late-onset Stargardt disease due to mild deep-intronic *ABCA4* alleles
Runhart EH, Valkenburg D, Cornelis SS, **Khan M**, Sangermano R, Albert S, Bax NM, Astuti G, Gilissen C, Pott JW, Verheij JBGM, Ellen A.W. Blokland EAW, Cremers FPM[#], van den Born LI[#], Hoyng CB[#].
Invest Ophthalmol Vis Sci. 2019; **60**: 4249-4256

Resolving the dark matter of *ABCA4* for 1,054 Stargardt disease probands through integrated genomics and transcriptomics
Khan M, Cornelis SS, Del Pozo-Valero M, Whelan L, Runhart EH, Mishra K, Bults F, AlSwaiti Y, AlTalbishi A, De Baere E, Banfi S, Banin E, Bauwens M, Ben-Yosef T, Boon CJF, L. van den Born LI, Defoort S, Devos A, Dockery A, Dudakova L, Fakin A, Farrar JG, Sallum JMF, Fujinami K, Gilissen C, Glavač D, Gorin MB, Greenberg J, Hayashi T, Hettinga Y, Hoischen A, Hoyng CB, Hufendiek K, Jägle H, Kamakari S, Karali M, Kellner U, Klaver CCW, Kousal B, Lamey T, MacDonald IM, Matynia A, McLaren T, Mena MD, Meunier I, Miller R, Newman H, Ntozini B, Oldak M, Pieterse M, Podhajcer OL, Puech B, Ramesar R, Rüther K, Salameh M, Salles MV, Sharon D, Simonelli F, Spital G, Steehouwer M, Szaflik JP, Thompson JA, Thuillier C, Tracewska A, Zweeden van M, Vincent AL, Zanlonghi X, Liskova P, Stöhr H, De Roach J, Ayuso C, Roberts L, Weber BHF, Claire-Marie Dhaenens C-M[#], Cremers FPM[#].
Genet Med. in press and BioRxiv. 2019; **doi:** https://doi.org/10.1101/817767

***ABCA4*-associated Stargardt disease**
Khan M and Cremers FPM.
Klin Monatsbl Augenheilkd. 2020; **237**: 267-274

In or Out? New insights on exon recognition through splice-site interdependency

Khan M*, Cornelis SS*, Sangermano R, Post IJM, Janssen Groesbeek A, Amsu J, Gilissen C, Garanto A**, Collin RWJ**, Cremers FPM

Int J Mol Sci. 2020**; 21**: 2300

*Shared first authors
[$]Shared second authors
**Equal contribution
[#]Shared senior authors

Curriculum vitae

Mubeen Khan was born on August 8, 1988 in Islamabad, Pakistan. After finishing her high school in 2006, she started her bachelor studies with a major in biology at Govt. Post Graduate College, Rawalpindi. During the years 2008-2013, she obtained both her MSc in Biochemistry and M.Phil (equivalent to MSc) in Advanced Biochemistry and Human genetics at the PMAS, UAAR, Rawalpindi, Pakistan. During her masters internship, she worked on the project aimed to find the ''Association of *ADAM33* SNPs with Asthma in a local Pakistani population'' at the Institute of Biomedical and Genetic Engineering (IB&GE), Islamabad, under the supervision of Dr. Atika Mansoor.

In January 2016, she joined the group of Prof. Dr. Frans P.M. Cremers in the Department of Human Genetics, at the Radboud University Medical Center, Nijmegen, The Netherlands, as a PhD candidate. Her project was mainly focused on unraveling the missing heritability in *ABCA4*-associated Stargardt disease, an autosomal recessive retinal disorder characterized by a significant fraction of mono-allelic (i.e. with incomplete molecular diagnosis) patients. To identify the missing pathogenic variants, an integrated genomics and transcriptomics approach was employed. She performed targeted *ABCA4* locus sequencing for the identification of genetic variants and functional characterization at the transcript level was carried out by employing midigene-based splice assays in HEK293T cells or by RNA analysis in patient-derived photoreceptor progenitor cells. Finally, an antisense oligonucleotides based splicing modulation approach was administered to rescue the splicing defects.

Mubeen Khan presented her work in several national and international congresses, among which, The Association for Research in Vision and Ophthalmology (ARVO) meeting, 2017 and 2018 and the Retinal Degeneration Meeting, 2018.