Babeş-Bolyai University Cluj-Napoca
Faculty of Biology and Geology
Doctoral School of Integrative Biology

# Whole exome sequencing reveals known and novel genetic defects in Israeli families with inherited retinopathies

**Summary of the Doctoral Thesis**

**Scientific adviser**

Dr. rer. nat. Annette Damert, MA.

**PhD student**

Lázár Csilla-Hanga

Cluj-Napoca

2016

# Contents

**Key words**

Retinal disease
Human genetics
Photoreceptor degeneration
Vision loss
Whole exome sequencing
*GUCY2D*
*ALMS1*
*DYSF*
*CEP78*

# Introduction

Vision disorders compromise the quality of life for millions of individuals worldwide. In humans photoreceptor degeneration is the primary cause of a group of inherited disorders referred to as retinal degenerative diseases (RDDs). To date, mutations in over 200 genes have been identified and over 250 genetic loci have been associated with RDDs (RetNet, https://sph.uth.edu/retnet/) (Figure 1) that have a wide range of clinical manifestations.

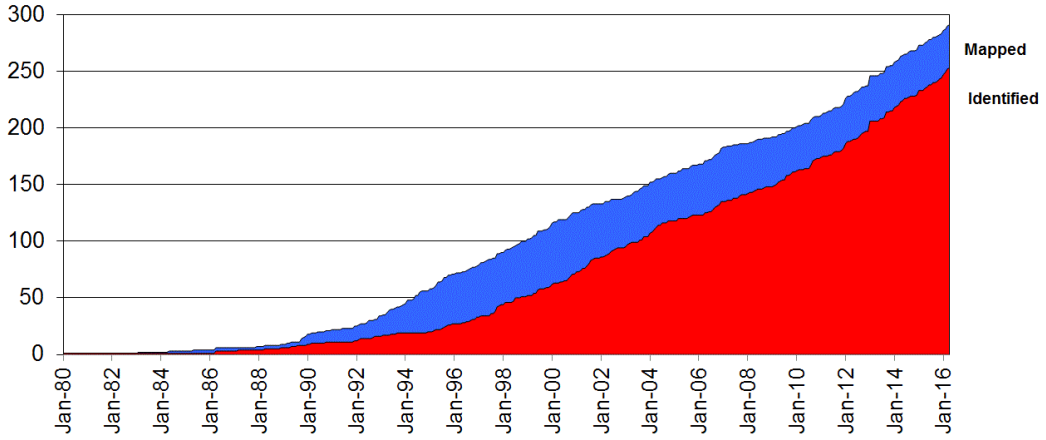



Figure 1: **Diagram of mapped and identified retinal disease genes.** The number of identified genes (red) and mapped genetic loci (blue) is plotted over a time course between 1980-2016 (RetNet, May 11, 2016).

Retinal dystrophies may affect one or both types of photoreceptors as well as additional tissues of the eye and they may be part of a more complex systemic disease affecting several tissues and organs in the affected individuals.

RDDs are clinically and genetically highly heterogeneousthat affect different aspects of normal photoreceptor function. Although mutations in a large number of genes were found to cause retinal disease, they explain only a fraction of the observed cases and there are many gaps in what is known about the underlying molecular mechanisms, limiting the available treatment options. Over the past

decades techniques such as linkage analysis, homozygosity mapping and targeted sequencing were successfully applied for disease gene identification. More recently next generation sequencing (NGS) techniques have offered an ever increasing opportunity for rapid and improved disease gene identification, discovery of novel genotype-phenotype associations and identification of genetic factors that affect phenotype severity and disease progression. The present thesis describes the successful application of molecular genetic, NGS and bioinformatic methods in disease gene discovery for RDD with different phenotypic manifestations. Genetic characterization, including identification of novel disease genes or novel genetic variants in known genes, is an important prerequisite for therapeutic intervention and may serve as a basis for development of novel therapeutic methods.

## Retinal degenerative disease

RDDs can be broadly divided into monogenic and multifactorial disorders. Monogenic subtypes are caused by mutations in a single gene that follow a Mendelian pattern of inheritance and can be autosomal dominant (AD), autosomal recessive (AR) or X-linked recessive (XL). Retinitis pigmentosa (RP), congenital stationary night blindness (CSNB), cone and cone-rod dystrophies (CD and CRD), Leber congenital amarousis (LCA) and macular degeneration (MD) represent a few of the monogenic subtypes (Berger et al., 2010). The complex multifactorial entities result from the interplay of multiple genetic and non-genetic (epigenetic and/or environmental) factors and include age-related macular degeneration, diabetic retinopathy and glaucoma (CookeBailey et al., 2013). The present work focuses on retinal dystrophies caused by mutations in a single gene.

Monogenic retinal diseases affect approximately 1 in 2000 individuals (Berger et al., 2010) and can manifest in mild retinal *DYSF*unction, severe retinal degeneration or legal blindness. For the majority of RDDs there are no available treatment options and their molecular diagnosis is made difficult by extreme genetic and phenotypic heterogeneity. Because of this a reliable genetic diagnosis is possible for only about half of the affected individuals (Berger et al., 2010).

RDDs can be divided into several groups based on rod versus cone involvement, a generalized degeneration of photoreceptors and their supportive tissues as

well as the involvement of other ocular tissues. The primary loss of rod photoreceptors is often followed by cone *DYSF*unction in RP, however a primary cone defect may or may not be accompanied by rod involvement. Both cones and rods are affected in CRD but only cone photoreceptors degenerate in CD. MD involves a more generalized *DYSF*unction of photoreceptors and retinal supportive tissues in the macular region resulting in central vision loss. Monogenic retinal dystrophies can be further classified into non-syndromic (where only ocular tissues are affected e.g. RP, CD, CRD, LCA etc.) and syndromic forms of disease (where the ocular phenotype is associated with the pathology of multiple other tissues). Examples for syndromic disorders include Usher syndrome (a combination of RP and hearing impairment), Bardet-Biedl syndrome (where the primary features include RP, polydactyly, obesity, renal and genital abnormalities, developmental delay, learning disability) and Alström syndrome (a progressive cone-rod photoreceptor dystrophy accompanied by various systemic manifestations like sensorineural hearing loss, obesity, short stature in adulthood, type II diabetis, cardiomyopathy, progressive pulmonary, hepatic and kidney *DYSF*unction) (Berger et al., 2010).

It is often hard to make reliable genotype-phenotype correlations because of the genetic and clinical overlap between different forms of disease (Figure 2) (Berger et al., 2010, Siemiatkowska et al., 2014).

Multiple clinical features may be caused by mutations in the same gene (e.g. *ABCA4* and *RPE65* gene mutations classically cause MD and LCA respectively but are often associated with RP like phenotype). Mutation in the same gene that follow different inheritance pattern (e.g. AD mutations in the *GUCY2D* gene cause CRD, while AR mutations cause LCA), or those that affect different functional regions of the encoded protein may also result in phenotypically distinct forms of disease (e.g. mutations at the 3' end of the alternative exon ORF15 of the *RPGR* gene tend to result in CRD over RP) (Berger et al., 2010). In addition mutation in genes associated with syndromic disease may also cause non-syndromic retinal disease (e.g. mutations in the *CEP290* gene are associated with Bardet-Biedl syndrome, Senior-Loken syndrome and Joubert syndrome but also cause LCA). Thus it is important that a thorough clinical characterization of the phenotype together with additional findings such as syndromic features is

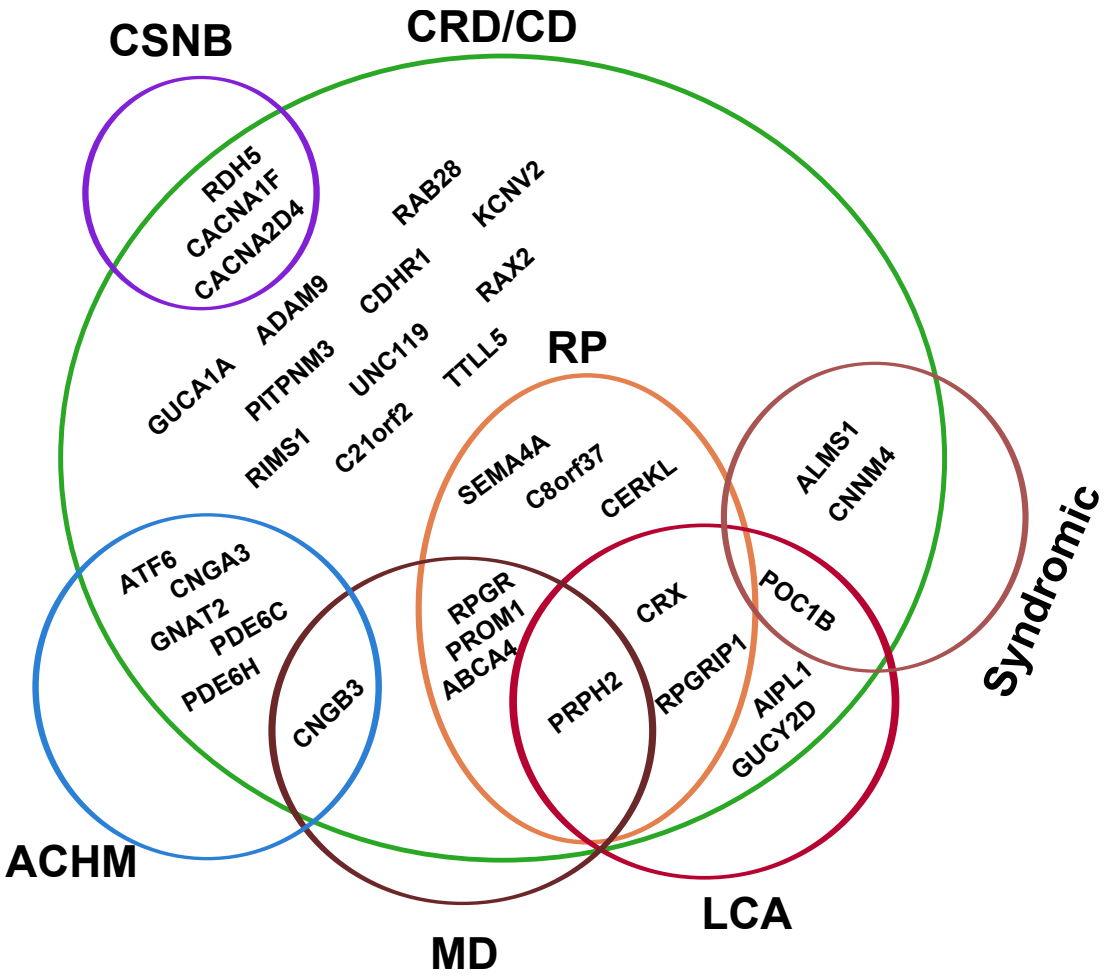



Figure 2: **Genetic overlap of cone and cone-rod dystrophy genes with other forms of retinal diseases.** Gene symbols that are part of overlapping areas show that mutations in the same gene can lead to different phenotypic manifestations. Different forms of retinal degeneration are indicated by colored circles. CSNB, congenital stationary night blindness; CRD, cone-rod dystrophy; CD, cone dystrophy; RP, retinitis pigmentosa; ACHM, achromatopsia; MD, macular degeneration; LCA, Leber congenital amaurosis and syndromic forms of retinal disease (modified after Berger et al., 2010).

available for optimal molecular characterization.

## Molecular biology techniques used for disease cause identification

Over the past few decades several strategies emerged that helped improve the chance to identify causative mutations in retinal disease (Neveling et al., 2013, Ratnapriya et al., 2013, Siemiatkowska et al., 2014).

Knowledge of the family tree is an important first step in disease gene identification, because it often provides significant clues regarding a potential selection of genes that should be considered first for molecular genetic testing. A family pedigree that spans over several generations together with detailed case history offers information about family relationships and the potential to predict the inheritance pattern followed by a trait (phenotypic character). However, sometimes an individual diagnosed with a particular disorder has no other family members with the condition e.g. as a result of *de novo* mutations (a mutation that is present for the first time in the family). A trait may also exhibit variable penetrance, in which case not all individuals who carry a particular mutation will exhibit similar clinical symptoms. Also due to environmental and additional genetic factors present in other individuals there could be milder signs of disease manifestation or a later age of onset.

Initial screening strategies used for the discovery of disease causing mutations often involve targeted analysis of mutation hotspots (regions where mutations are observed with higher frequency), specific founder mutations (mutations that occur as a result of the loss of genetic variation when a new population is established by a small number of individuals) or a selection of mutations frequently identified in similar cases, followed by sequencing of other exons of a gene (Berger et al., 2010, Siemiatkowska et al., 2014). Sanger sequencing is the technique of choice for targeted analysis of specific mutations that occur in a small subset of genes or targeted screening of genes with a limited number of exons. Its advantages include accuracy, speed and relatively low cost, however it is too labor intensive to perform large scale screenings using this method. If no known mutations are identified using targeted Sanger sequencing, samples may be analyzed using large

scale screening strategies such as single nucleotide polymorphism (SNP) arrays or NGS strategies.

NGS strategies allow a comprehensive analysis of parts or the whole genome and enable effective identification of mutations and novel disease genes even in small families with limited number of affected members or unrelated individuals (Metzker et al., 2010, Ng et al., 2010). Targeted analysis of all or a specific subset of genes previously involved in retinal disease may be performed using a targeted NGS approach. Targeted sequencing is cost effective, it allows identification of mutations in known causal genes and it provides flexibility regarding the clinical diagnosis, which can be difficult considering the phenotypic overlap among different forms of retinal disease. This approach is also useful in isolated cases, where textitde novo mutations occur in only one member of a family (Siemiatkowska et al., 2014). If a targeted approach does not reveal a potential causal variant, exome sequencing is usually considered. Whole exome sequencing (WES) is the targeted sequencing of all protein-coding regions (exons) that constitute ∼1-2% of the human genome (Ng et al., 2009) and harbor ∼85% of the rare mutations that affect protein function (Cooper 1995). The power of exome sequencing for identification of novel disease genes and potential disease causing genetic variants in rare Mendelian disorders is widely recognized (Ng et al., 2010, Bamshad et al., 2011, Gilissen et al., 2011). Exome sequencing is well suited for the molecular diagnosis of small families affected by genetically and phenotypically heterogeneous disorders such as RDDs. The success rate of WES is increased substantially when used in combination with positional information obtained from linkage analysis or HM with several examples in RDD (Estrada-Cuzcano et al., 2012, Neveling et al., 2013, Roosing et al., 2013, Kohl et al., 2015). However, the analysis of WES data can be challenging due to the high number of variants identified that increase the chance of incidental findings (Neveling et al., 2013). Rigorous data analysis and careful variant filtering are crucial for accurate identification of disease associated genetic variants. Limitations of exome sequencing include the inability to analyze conserved genomic or regulatory regions that lie outside the coding region and may cause a decrease or imbalance in transcript levels. WES does not detect non-coding or deep intronic variants, long range genomic alterations and structural variants (insertions, deletions, duplications, copy-number variants, inversions and

translocations) although small insertions and deletions present within the coding region may be successfully detected on occasion (Ng et al., 2009, Bamshad et al., 2011, Ratnapriya et al., 2013, Siemiatkowska et al., 2014). At the same time it allows analysis of 20 times as many samples as whole genome sequencing (WGS) and is therefore optimal for identification of rare variants with medical implication (Ng et al., 2009). WGS may be used for identification of variants beyond coding regions in cases where other methods are unsuccessful (Audo et al., 2012, Nishiguchi et al., 2013). As the costs of WGS decreases, it becomes available for a broader group of laboratories, however technological limitations associated with the analysis and handling of large datasets still exist (Ratnapriya et al., 2013, Siemiatkowska et al., 2014). It is difficult to identify a potential disease causing variant from the ∼3 million variants detected in each individual genome, in addition the effects of regulatory and intronic variants is not certain and additional studies are needed to understand their functional implication (Siemiatkowska et al., 2014).

## Studied population group

The Israeli and Palestinian populations include a large number of semi-isolated subpopulations where the frequency of intracommunity marriages is relatively high (Cohen et al., 2004). These ethnic groups include Jews of different origins, Arab Muslim, Arab Christian, Bedouin, and Druze subpopulations that differ in their religion and geographic/cultural origin (Vardi-Saliternic et al., 2002, Zlotogora2010 et al., Lazar et al., 2015a). As a consequence there is a widespread presence of consanguineous marriages leading to a unique genetic make-up (Vardi-Saliternic et al., 2002, Beryozkin et al., 2014). Consanguinity rates are highest among Bedouins where around 70% of marriages are between spouses related as second cousins or closer (Zlotogora et al., 2010). Among Arab-Muslims and Druze more than 25% of marriages occur between spouses related as first cousins with an additional 20% where spouses are related in other ways (Beryozkin et al., 2014). In comparison less than 1% of marriages are consanguineous in the North American population (Beryozkin et al., 2014).

The relatively high rate of consanguinity results in increased prevalence of

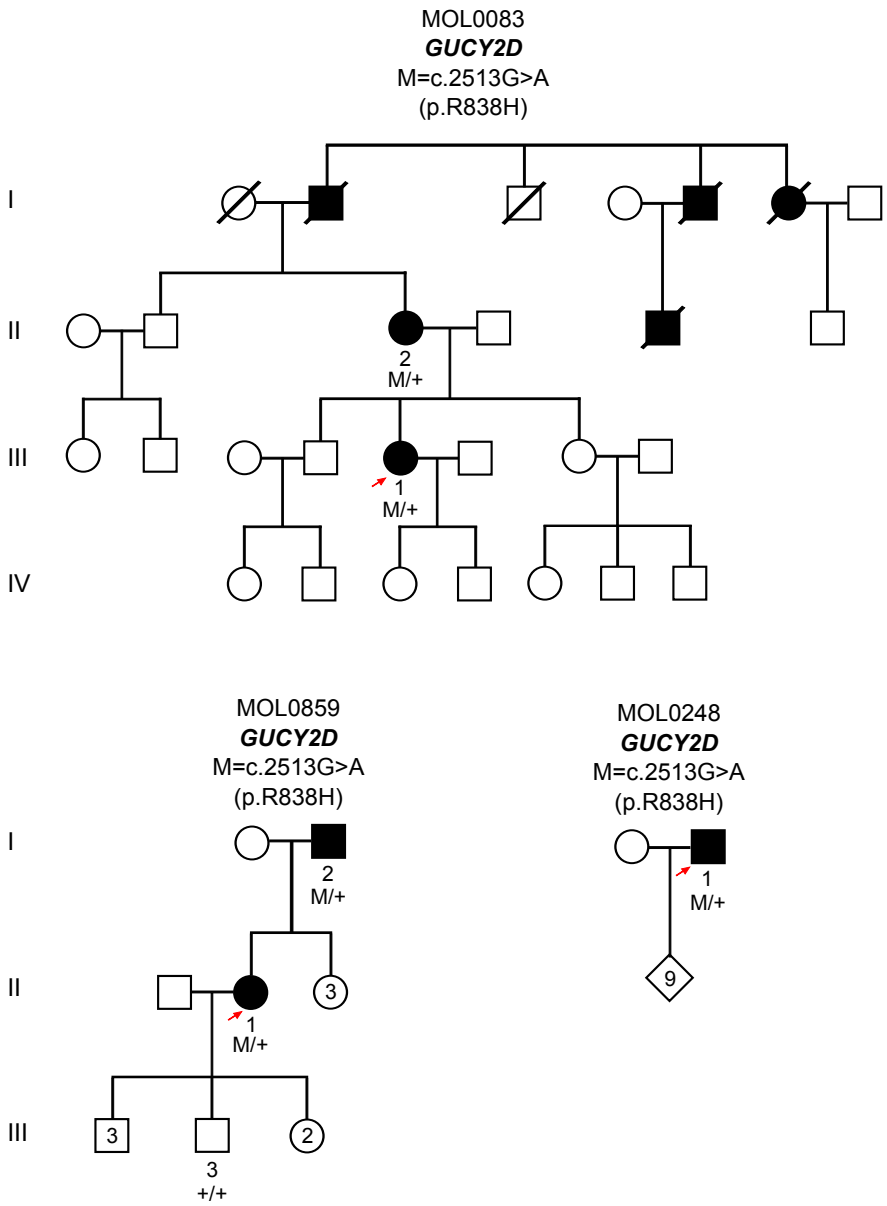
hereditary diseases, particularly those with AR pattern of transmission (Zlotogora et al., 2010). This is also true for retinal dystrophies where homozygous disease causing mutations are often identified (Zelinger et al., 2010, Cohen et al., 2012, Beryozkin et al., 2015). For example, 49% of RP cases in a cohort of 183 families were found to be AR due to consanguinity and limited mixing among different subpopulations (Sharon et al., 2015). In contrast AD inheritance of RP in the Israeli and Palestinian populations appears to be infrequent with only 8% versus 30% to 40% in other populations (Hartong et al., 2006, Sharon et al., 2015). The high prevalence of AR disease increases the potential for discovery of novel genes and novel mutations in known genes.

Thus, modes of inheritance and genetic causes of disease in Israel differ from those described in the European and North American populations, making the Israeli population a rich source for novel genetic discovery.

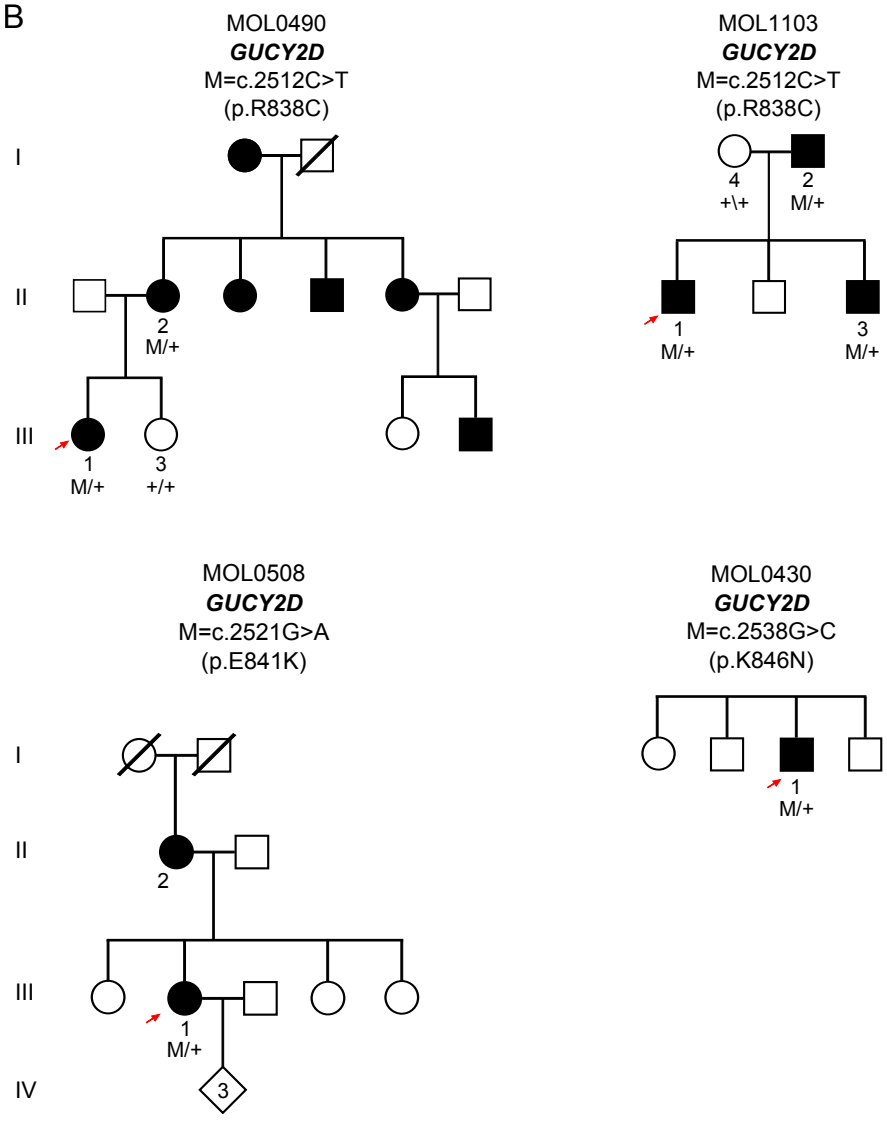
# Results

Whole genome SNP analysis followed by WES was performed on an initial set of 27 individuals from 8 unrelated families with cone dominated retinal phenotypes (Figure 3).

A


MOL0083
GUCY2D
M=c.2513G>A
(p.R838H)
I
II
III
IV
2
M/+
1
M/+
MOL0859
GUCY2D
M=c.2513G>A
(p.R838H)
I
II
III
2
M/+
1
M/+
3
3
2
3
+/+
MOL0248
GUCY2D
M=c.2513G>A
(p.R838H)
1
M/+
9

B


MOL0490
GUCY2D
M=c.2512C>T
(p.R838C)
I
II
III
2
M/+
1
M/+
3
+/+
MOL1103
GUCY2D
M=c.2512C>T
(p.R838C)
4
+\+
2
M/+
1
M/+
3
M/+
MOL0508
GUCY2D
M=c.2521G>A
(p.E841K)
I
II
III
IV
2
1
M/+
3
MOL0430
GUCY2D
M=c.2538G>C
(p.K846N)
1
M/+

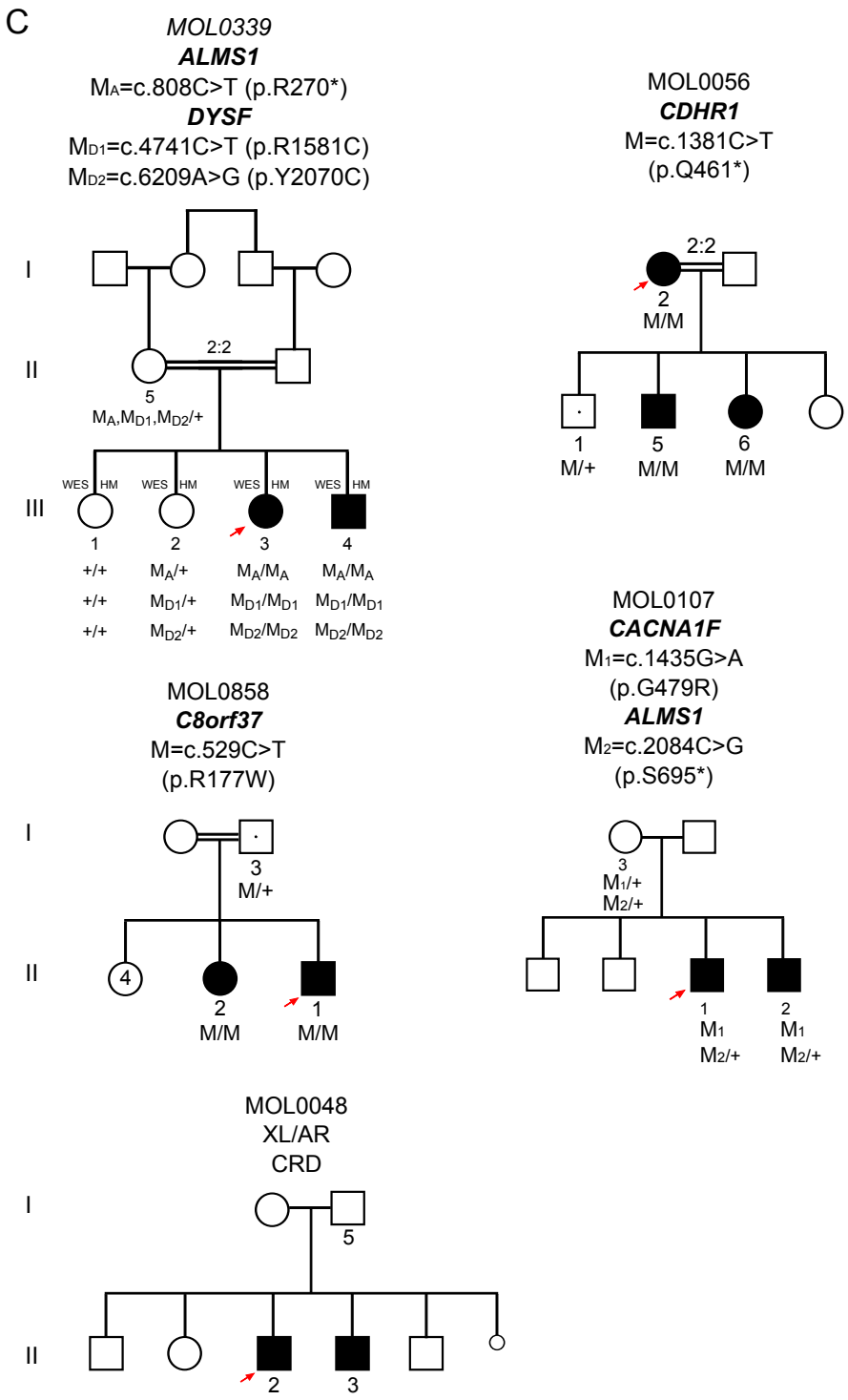



Figure 3: **Pedigree structure of the studied families.** Family identifiers, gene

names and identified coding changes are shown. Segregation in the families is indicated as: M/M, homozygous for mutation; M/+, heterozygous; +/+ homozygous for wild type allele. Filled symbols represent affected individuals, open symbols represent unaffected individuals, circles and squares represent female and male individuals respectively, diamond shapes represent individuals of unknown gender, a small circle signifies still birth. Numbers inside symbols represent the number of similar gender siblings. Red arrows indicate index cases. Consanguineous marriages are represented by double lines, first cousin relationships are marked by 2:2. (A) Families showing dominant inheritance pattern with p.R838H *GUCY2D* mutations. (B) Families with dominant inheritance pattern of p.R838C, p.E841K and p.K846N *GUCY2D* mutations. (C) Families showing recessive inheritance pattern with *ALMS1*, *CDHR1*, *C8orf37*, *CACNA1F* mutations and MOL0048 unsolved family (modified after Lazar et al., 2015ab ).

Families MOL0490, MOL0859 and MOL1103 showed an apparent AD inheritance pattern, families MOL0056, MOL0858 and MOL0339 were consanguineous with AR inheritance and disease segregation was suggestive of XL or AR inheritance in families MOL0107 and MOL0048. As part of an ongoing effort to characterize retinal disease in the Israeli population, an additional set of 20 (Table 1) families presenting different forms of retinal degeneration were subsequently studied.

Table 1: **List of additional families presenting different forms of retinal degeneration.**

| | Family Identifier | Presumed Inheritance | Phenotype | Homozygosity mapping |
|---|---|---|---|---|
| 1 | MOL0061 | XL/AR | CRD | NO |
| 2 | MOL0360 | AR | CRD | YES |
| 3 | MOL0364 | AR | CRD | YES |
| 4 | MOL0454 | AR/XL | CRD | NO |
| 5 | MOL0474 | AR | CRD+Deafness | YES |
| 6 | MOL0679 | XL/AR | CRD/Usher syndrome | YES |
| 7 | MOL1124 | AR | CD/CRD/ACHM/LCA | NO |
| 8 | MOL1182 | AR | CRD/CD | NO |
| 9 | MOL1190 | XL | ACHM/CRD | NO |
| 10 | MOL0499 | AR | Maculopathy | NO |
| 11 | MOL0563 | AR | Maculopathy | YES |
| 12 | MOL0584 | AD/AR | Maculopathy | NO |
| 13 | MOL1126 | AR/AD | Maculopathy | NO |
| 14 | MOL1152 | AR | Maculopathy | NO |
| 15 | MOL1154 | AD | Maculopathy | NO |
| 16 | MOL0039 | XL/AR | RP | NO |
| 17 | MOL0331 | XL/AR | RP | NO |
| 18 | MOL0622 | AD/XL | RP | NO |
| 19 | MOL0625 | AD/XL | RP | NO |
| 20 | MOL0864 | AR | RP | NO |

CRD, cone rod dystrophy; CD, cone dystrophy; ACHM, achromatopsia; LCA, Leber congenital amaurosis, RP, retinitis pigmentosa; XL, X-linked; AR, autosomal recessive; AD, autosomal dominant.

Molecular genetic analysis that included screening of families for founder mutations prevalent among families of the same ethnic origin did not reveal causal mutations (Lazar et al., 2015a). Whole genome SNP analysis revealed six homozygous genomic regions in family MOL0858 on chromosomes 2, 8, 11, 16 and 21 and a single homozygous region in family MOL0339 on chromosome 2 (Table 2). Whole genome SNP analysis was also performed on members of 5 of the 20 additional studied families: MOL0360, MOL0364, MOL0474, MOL0679 and MOL0563 (Table 2).

Table 2: **Segregating homozygous regions in the studied families.**

| **Family Identifier** | **Phenotype** | **Segregating homozygous regions** |
|---|---|---|
| MOL0339 | CRD | chr2: 49M-84.3M |
| MOL0360 | CRD | chr1: 92M-109M, chr2: 180M-204M, chr3: 53M-68M and 95M-109M, chr4: 37M-58M, chr18: 8M-55M |
| MOL0364 | CRD | chr2: 224M-238M, chr3: 153M-175M, ch21: 19M-39M |
| MOL0474 | CRD+Deafness | chr7: 70M-78M, chr16: 17M-64M |
| MOL0563 | Macupolathy/ACHM | chr8: 73M-95M |
| MOL0679 | CRD/Usher syndrome | chr9: 20M-85M, chr19: 15M-49M |
| MOL0858 | CRD | chr2: 8M-15M and 224M-228M, chr8: 76M-103M, chr11: 110M-118M, chr16: 53M-59M, chr21: 23M-27M |

CRD, cone rod dystrophy; ACHM, achromatopsia.

Analysis of WES data led to the identification of five previously reported mutations and four novel potential disease causing variants (Table 3). Families MOL0490, MOL0859 and MOL1103 were found to harbor *GUCY2D* gene mutations. In families MOL0056, MOL0107 and MOL0858 we identified known mutations in *CDHR1*, *ALMS1* and *C8orf37* genes respectively. A novel *ALMS1* gene variant was identified in family MOL0339 and a second novel potential disease associated variant in *CACNA1F* gene was identified in family MOL0107. All variants were validated by Sanger sequencing and segregated with the disease

phenotype. Extensive analysis of the variants identified from the exome data did not reveal a potential causative mutation that showed segregation with disease in family MOL0048 (Lazar et al., 2015a).

Table 3: **Genetic variants identified in Israeli families with primary cone involvement.**

| Family Identifier | Clinical Diagnosis | Presumed Mode of Inheritance | Origin / Ethnic group | Causative Gene (exon) | Location of Nucleotide Change (Protein)** | Reference |
|---|---|---|---|---|---|---|
| MOL0490 | CD/CRD | AD | Turkish Jewish | *GUCY2D* (13) | c.2512C>T (p.R838C) heterozygous | Kelsell et al.,1998 |
| MOL0859 | CD | AD | Ashkenazi Jewish | *GUCY2D* (13) | c.2513G>A (p.R838H) heterozygous | Weigell-Weber et al.,2000 |
| MOL1103 | CD/CRD | AD | Arab Muslim | *GUCY2D* (13) | c.2512C>T (p.R838C) heterozygous | Kelsell et al.,1998 |
| MOL0083 | CD | AD | Ashkenazi Jewish | *GUCY2D* (13) | c.2513G>A (p.R838H) heterozygous | Weigell-Weber et al.,2000 |
| MOL0248 | CRD | Isolate | Ashkenazi Jewish | *GUCY2D* (13) | c.2513G>A (p.R838H) heterozygous | Weigell-Weber et al.,2000 |
| MOL0430 | CRD | Isolate | North African Jewish | *GUCY2D* (13) | c.2538G>C (p.K846N) heterozygous | Novel† |
| MOL0508 | Maculopathy + CD | AD with reduced penetrance | Ashkenazi Jewish | *GUCY2D* (13) | c.2521G>A (p.E841K) heterozygous | Novel† |
| MOL0056 | RD | AR | Arab Muslim | *CDHR1* (13) | c.1381C>T (p.Q461*) homozygous | Duncan et al.,2012 |
| MOL0107 | CRD | XL/AR | North African / Ashkenazi Jewish | *CACNA1F* (11) | c.1435G>A (p.G479R) hemizygous | Novel |
| MOL0107 | CRD | XL/AR | North African / Ashkenazi Jewish | *ALMS1 (8)* | c.2084C>G (p.S695*) heterozygous | Liang et al., 2013 |
| MOL0339 | CRD | AR | Arab-Muslim | *ALMS1* (5) | c.808C>T (p.R270*) homozygous | Novel†† |
| MOL0339 | LGMD2B | AR | Arab-Muslim | *DYSF (43)* | c.4741C>T (p.R1581C) | Novel†† |
| MOL0339 | LGMD2B | AR | Arab-Muslim | *DYSF (55)* | c.6209A>G (p.Y2070C) | Novel†† |
| MOL0858 | CRD | AR | Arab Muslim | *C8orf37* (6) | c.529C>T (p.R177W) homozygous | Estrada-Cuzcano et al.,2012 |
| MOL0048 | CRD | XL/AR | Arab-Muslim | - | - | - |

*Modified from [Lazar et al.,2015a], CD, cone dystrophy; CRD, cone-rod dystrophy; RD, retinal

dystrophy; LGMD2B, limb-girdle muscular dystrophy type 2B; AD, autosomal dominant; AR, autosomal recessive; XL, X-linked.

**The nucleotide position of each genetic variant is based on the following GenBank cDNAs (accession number): *GUCY2D* (NM_000180), *CDHR1* (NM_001171971), *CACNA1F* (NM_001256789), *ALMS1* (NM_015120), *DYSF* (NM_003494.3), *C8orf37* (NM_177965); Nucleotide numbers reflect cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence. The initiation codon is codon 1.
Published by †[Lazar et al.,2015a], ††[Lazar et al.,2015b]

## Autosomal dominant families with *GUCY2D* gene variants

Two previously reported *GUCY2D* missense mutations were identified in three families that follow dominant inheritance patter (Table 3) (Figure 3) (Lazar et al., 2015a). Families MOL0490 and MOL1103 harbored a c.2512C>T (p.R838C) (Kelsell et al., 1998) mutation and family MOL0859 a c.2513G>A (p.R838H) (Weigell-Weber et al., 2000) mutation in exon 13 of the *GUCY2D* gene. Mutations in the *GUCY2D* gene were not previously described in patients with retinal disease from the Israeli population. To search for additional families with *GUCY2D* mutation, we performed targeted Sanger sequencing of exons 13 and 14 in 106 additional Israeli index cases selected from over 1300 families with hereditary retinal disease (Lazar et al., 2015a). This resulted in the identification of four additional families with *GUCY2D* mutation (Table 3) (Figure 3). Family MOL0083 characterized with AD cone dominated disease and an isolate case in family MOL0248 were found to harbor the previously described c.2512C>T (p.R838H) heterozygous sequence change. Two novel potential disease causing variants were identified in the other two families: a c.2538G>C (p.K846N) variant in an isolate case from family MOL0430 and a c.2521G>A (p.E841K) variant in an AD cone dominated family MOL0508 (Lazar et al., 2015a). The affected amino acid residues are highly conserved across species and are part of the dimerization domain of the protein (Figure 4) (Lazar et al., 2015a).

Clinical assessment of individuals affected with retinal degeneration from the seven families revealed phenotypic characteristics that are similar to those reported earlier for AD *GUCY2D*-associated disease (Gregory-Evans et al., 2000,

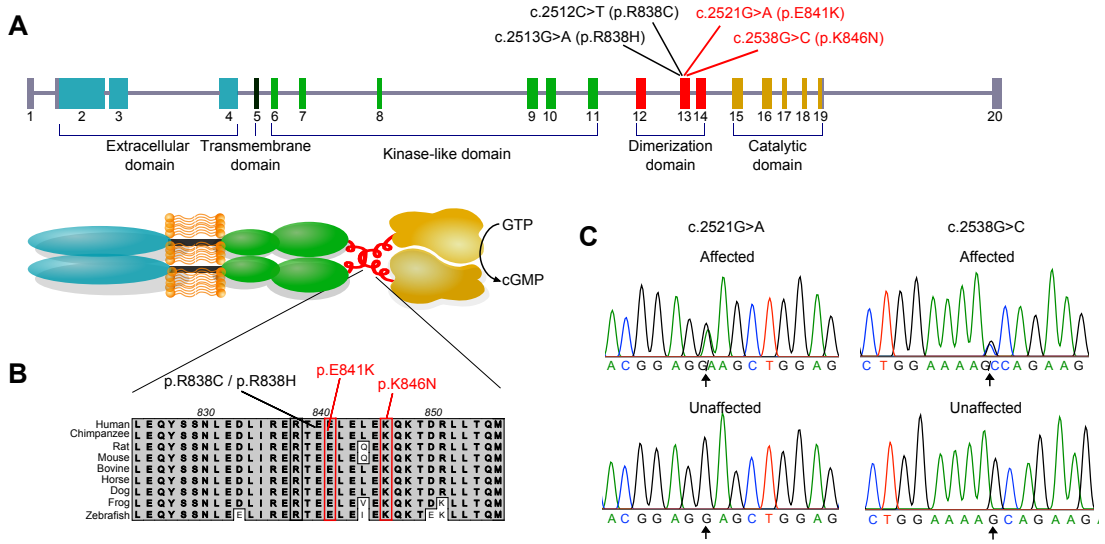



Figure 4: **Schematic representation of the *GUCY2D* gene and protein structure, interspecies conservation of affected amino acid residues and sequencing chromatograms of the two novel variants discovered.**
(A) The *GUCY2D* gene is composed of 20 exons that encode 5 protein domains: an extracellular domain, a transmembrane domain, a kinase-like domain, a dimerization domain, and a catalytic domain. Exons are color coded according to the respective encoded protein domains. All four genetic variants identified (red, novel; black, previously reported) are located in exon 13 of the gene and are predicted to affect the dimerization properties of the resulting protein monomers. (B) Interspecies comparison was generated using GenBank protein sequences (accession number) from human (NP_000171.1), chimpanzee (XP_003315414.1), rat (NP_077356.1), mouse (NP_032218.2), bovine (NP_776973.2), horse (XP_005597817.1), dog (NP_001003207.1), frog (XP_002942678.2), and zebrafish (NP_001103165.1). Amino acid residues affected by novel (red frames) and known (black frame) genetic variants are highly conserved. (C) Sequencing chromatograms showing the novel c.2521G>A and c.2538G>C heterozygous changes (top), and the wild type allele (bottom) in affected and unaffected individuals (arrows) [Lazar et al., 2015a].

Kitiratschky et al., 2008, Lazar et al., 2015a). Fundus examination revealed degeneration that was largely confined to the macular area, the retinal pigment epithelium (RPE) showed "salt and pepper" changes that occur when tiny flecks of dark pigment are mixed with areas of whitish depigmentation. In more severe cases marked circumscribed atrophy was evident in the foveal and parafoveal area (Figure 5 A, B) (Lazar et al., 2015a).

FAF imaging revealed atrophic changes that ranged from small hypofluorescent spots in young patients presenting milder phenotype and later progressed to large, well delimited hypofluorescent areas that were surrounded by a hyperfluorescent ring in advanced cases. The presence of cone dominated disease was confirmed by visual field testing that revealed marked macular involvement. OCT imaging showed a similar degree of macular RPE and photoreceptor involvement. Mild thinning of the photoreceptor layer was evident in the early phases with hypodense "cavitations" of the photoreceptor inner segment-outer segment complex in the foveal area (Figure 5 A) (Lazar et al., 2015a). In severe cases there was a complete loss of the photoreceptor layer and atrophy with choroidal backscatter (Figure 5 B) (Lazar et al., 2015a). Except for changes associated with high myopia, the peripheral retina was largely within normal limits. Perimetry and microperimetry testing showed central partial and/or absolute scotomas with preserved peripheral fields (Figure 5 A, B) (Lazar et al., 2015a).

## Autosomal recessive family MOL0339 with *ALMS1* and *DYSF* gene variants

Family MOL0339 is a two-generation consanguineous family with two siblings presenting an early-onset severe autosomal recessive CRD and muscular dystrophy phenotype (Figure 3) (Lazar et al., 2015b). The parents of the two affected siblings are first cousins therefore disease causing mutations are likely to be part of large homozygous regions shared by the affected individuals. The combination of cone dominated ocular phenotype and muscular dystrophy has not been reported previously, thus we assumed that they could be part of a novel syndrome caused by a single gene mutation or result from distinct mutations in two different genes. Whole genome SNP analysis followed by WES was performed to identify

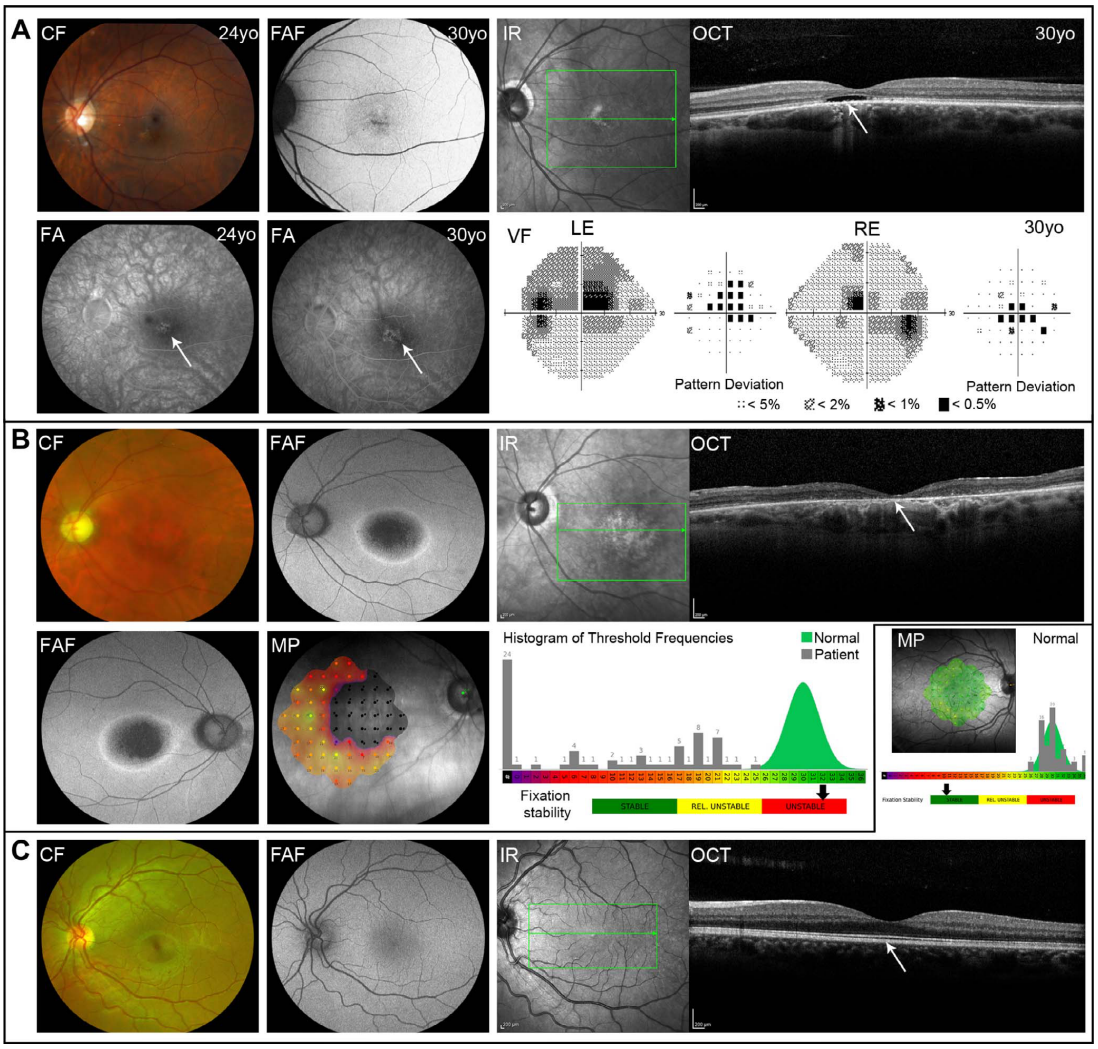



Figure 5: **Ocular symptoms in *GUCY2D*-associated retinal degeneration.**
(A) Retinal imaging and function in patient MOL0430 I:1, with relatively mild manifestation of the disease. The color fundus (CF) photo and fluorescein angiography (FA) performed at age 24 reveal localized macular RPE changes with staining, which increased by age 30. Fundus autofluorescence (FAF) study at age 30 shows small areas of atrophy in the foveal area and hypodense "foveal cavitation" is present on the optical coherence tomography OCT (arrow). Static perimetry testing documents central scotomas in both eyes. LE, left eye; RE, right eye. (B) Ocular pathology in patient MOL0248 I:1, manifesting severe disease at the age of 59. Multicolor, FAF and infrared (IR) imaging of the left eye show central macular atrophy, with loss of the photoreceptor layer in the foveal area on OCT (arrow). Symmetry between the left and right eyes is exemplified by the FAF images (note dark hypofluorescent area and surrounding hypofluorescent ring). Microperimetry examination of the right eye reveals an absolute central scotoma and reduced sensitivity beyond it, with poor fixation. (C) Microperimetry and imaging in a 32-year-old unaffected individuals are shown for comparison. Clinical assessment and imaging was performed by collaborating researches from the Department of Ophthalmology Hadassah Medical Center. [Lazar et al., 2015a].

the cause of the two phenotypes. Homozygosity mapping was performed on two affected (MOL0339 III:3; MOL0339 III:4) and two unaffected siblings (MOL0339 III:1, MOL0339 III:2) and data analysis revealed a single homozygous region on chromosome 2, encompassing a 35.3 Mb region (Lazar et al., 2015b). None of the 263 genes present in this region were previously associated with nonsyndromic AR CRD.

WES analysis revealed three novel genetic variants in two independent genes *ALMS1* and *DYSF* (Table 3). Both *ALMS1* and *DYFS* are part of the 35.3 Mb homozygous region identified by HM. Mutations in *ALMS1* cause Alström syndrome (Marshall et al., 2015) and mutations in *DYSF* cause limb-girdle muscular dystrophy (LGMD) or Miyoshi muscular dystrophy (Liu et al., 1998). One novel nonsense variant c.808C>T (p.R270*) was identified in exon 5 of ALMS1 gene and two novel potentially disease causing missense variants c.4741C>T (p.R1581C) and c.6209A>G (p.Y2070C) in exons 43 and 55 of the *DYSF* gene (Lazar et al., 2015b). Sanger sequencing was used to validate all variants and to confirm co-segregation of identified variants with the observed phenotype. In an attempt to eliminate the possible presence of disease associated variants in genes previously associated with photoreceptor degeneration accompanied by some form of muscular involvement we performed visual examination of sequencing reads from WES data. No rare coding variants or larger genetic aberrations were found in any of the following genes: *FLVCR1, DTHD1, OPA3, ELOVL4, ABHD12, PRPS1, MT-ATP6* (Lazar et al., 2015b) (RetNet, https://sph.uth.edu/retnet/). A set of 87 index cases with inherited retinal degeneration were screened for the c.808C>T change in *ALMS1* to examine if the exon 5 variant is a founder mutation in the Arab-Muslim population (Lazar et al., 2015b). No additional individual was identified to carry this mutation or other potential disease causing variants in exon 5.

## Autosomal recessive families MOL0056 and MOL0858 with *CDHR1* and *C8orf37* gene variants

Previously published genetic variants were identified in two of the studied autosomal recessive families MOL0056 and MOL0858 (Figure 3) (Table 3). WES

followed by data analysis revealed a known nonsense c.1381C>T (p.Q461*) (Duncan et al., 2012) mutation in exon 13 of the *CDHR1* gene in family MOL0056 and a previously published missense c.529C>T (p.R177W) (Estrada-Cuzcano et al., 2012) mutation in exon 6 of the *C8orf37* gene in family MOL0858 (Lazar et al., 2015a). Family MOL0056 was clinically diagnosed with early onset retinal degeneration while clinical evaluation of family MOL0858 revealed an advanced CRD phenotype with marked color vision impairment. Whole genome SNP genotyping in preceded WES analysis in family MOL0858 and revealed six large homozygous segregating regions: chr2, 8M-15M and 224M-228M; chr8, 76M-103M; chr11, 110M-118M; chr16, 53M-59M and chr21, 23M-27M (Table 2) (Lazar et al., 2015a). The reported c.529C>T (p.R177W) *C8orf37* gene variant is part of a 27Mb homozygous region identified between 76 and 103 Mb on chromosome 8.

## Autosomal recessive family MOL0107 with *CACNA1F* and *ALMS1* variants

Affected siblings from family MOL0107 were diagnosed with early onset cone-rod dystrophy. WES followed by data analysis revealed a novel potentially disease causing variant c.1435G>A (p.G479R) in exon 11 of *CACNA1F* gene (Figure 3) (Table 3). In addition all 3 individuals were heterozygous for a known c.2084C>G (p.S695*) (Liang et al., 2013) nonsense mutation in exon 8 of *ALMS1* gene (Table 3). The presence of both variants was validated by Sanger sequencing. Clinical diagnosis of the two affected individuals revealed an early onset, advanced CRD phenotype.

## Autosomal recessive family MOL0048

Family MOL0048 was clinically diagnosed with cone dominated retinal degeneration, both affected male siblings presented mild to moderate cone *DYSF*unction (Figure 3) (Lazar et al., 2015a). There was no report of consanguinity within the family, thus we presumed that retinal disease was inherited either in XL or AR pattern. Candidate gene screening and exome sequencing analysis performed on two affected brothers (MOL0048 II:2 and MOL0048 II:3) and their unaffected

father (MOL0048 I:5) did not reveal a potential disease associated genetic defect. Mutations in the ORF15 (open reading frame 15) exon of the *RPGR* gene are known to account for a large fraction of XL CRD cases. We ruled out the involvement of this gene in family MOL0048 by examining the *RPGR* gene variants identified by exome sequencing in the affected brothers and found that they inherited different maternal alleles in this region (Lazar et al., 2015a).

## Additional families presenting retinal degenerative disease

WES was performed on a total of 28 individuals from a set of 20 families (Table 1) including four families diagnosed with CRD, five families with a combination of multiple phenotypes, six families presenting maculopathy and five families with RP. Analysis and filtering of variants revealed by WES led to the identification of seven novel potential disease causing genetic variants and four known disease causing mutations in 9 of the studied families (Table 4).

Table 4: **List of genetic variants identified in the additional set of retinal degeneration families.**

| Family Identifier | Phenotype | Presumed Mode of Inheritance | Origin / Ethnic group | Gene (exon) | Location of Nucleotide Change (Protein)* | Sanger validation (segregation) | Reference |
|---|---|---|---|---|---|---|---|
| MOL0061 | CRD | XL/AR | North African Jewish | *CACNA1F (35)* | c.4051C>T (p.R1351*) hemizygous | Yes (Yes) | Novel |
| MOL0360 | CRD | AR | Bedouin | *CNGB3 (6)* | c.782A>G (p.D261G) homozygous | Yes (Yes) | Novel |
| MOL0364 | CRD | AR | Arab Muslim | - | - | - | - |
| MOL0454 | CRD | AR/XL | Druze | - | - | - | - |
| MOL0474 | CRD+Deafness | AR | Arab Muslim | *CNGA3 (7)* | c.931G>T (p.G311C) heterozygous | Yes (Yes) | Novel |
| MOL0679 | CRD/Usher syndrome | XL/AR | Oriental Jewish | *CEP78 (7)* | c.893-1G>A (p.D298Vfs*17) homozygous | Yes (Yes) | Novel** |
| MOL0773 | CRD/Usher syndrome | AR | Oriental Jewish | *CEP78 (7)* | c.893-1G>A (p.D298Vfs*17) homozygous | Yes (Yes) | Novel** |
| MOL1310 | CRD/Usher syndrome | AR/XL | Oriental Jewish | *CEP78 (4)* | c.534delT (p.K179Rfs*10) homozygous | Yes | Novel** |
| TB279 | CRD/Usher syndrome | AR/XL | Oriental Jewish | *CEP78 (4)* | c.534delT (p.K179Rfs*10) homozygous | Yes | Novel** |
| TB365 | CRD/Usher syndrome | AR/XL | Oriental Jewish | *CEP78 (4,7)* | c.534delT (p.K179Rfs*10) c.893-1G>A (p.D298Vfs*17) compoundheterozygous | Yes (Yes) | Novel** |
| MOL1124 | CD/CRD/ ACHM/LCA | AR | Arab Muslim | *RPGRIP1 (15)* | c.2249A>G (p.Y750C) homozygous | Yes (Yes) | Novel |
| MOL1182 | CRD/CD | AR | Moroccan/ Ashkenazi Jewish | - | - | - | - |
| MOL1190 | ACHM/CRD | XL/AR | Ashkenazi Jewish | *CNGB3 (6,10)* | c.819_826del (p.P273fs) c.1148delC (p.T383fs) compoundheterozygous | Yes (Yes) | Kohl2000, Sundin2000 |
| MOL0499 | RD+ Maculopathy | AR | Arab Christian | *CYP4V2 (11)* | c.1508G>A (p.G503E) homozygous | - | Novel |
| MOL0499 | RD+ Maculopathy | AR | Arab Christian | *CNGB3 (4)* | c.C467T (p.S156F) heterozygous | - | Sundin2000 |
| MOL0563 | Maculopathy/ ACHM | AR | Arab Muslim | *CNGB3 (10)* | c.1148delC (p.T383fs) homozygous | Yes (Yes) | Kohl2000, Sundin2000 |

| Family Identifier | Phenotype | Presumed Mode of Inheritance | Origin / Ethnic group | Gene (exon) | Location of Nucleotide Change (Protein)* | Sanger validation (segregation) | Reference |
|---|---|---|---|---|---|---|---|
| MOL0584 | Maculopathy | AD/AR | Syrian/Turkish Jewish | *ABCA4 (35,38)* | c.4919G>A (p.R1640Q) c.5318C>T (p.A1773V) compoundheterozygous | Yes (Yes) | Rozet1998, Chacon-Camacho2013 |
| MOL1126 | Maculopathy | AR/AD | Arab Muslim | - | - | - | - |
| MOL1152 | Maculopathy | AR | Arab Muslim | - | - | - | - |
| MOL1154 | Maculopathy | AD | Georgian Jewish | - | - | - | - |
| MOL0039 | RP | XL/AR | Turkish/Spanish Jewish | - | - | - | - |
| MOL0331 | RP | XL/AR | French Jewish | - | - | - | - |
| MOL0622 | RP | AD/XL | Yemeni/Iraqi Jewish | - | - | - | - |
| MOL0625 | RP | AD/XL | Bukharan/Egyptian Jewish | - | - | - | - |
| MOL0864 | RP | AR | Arab Muslim | - | - | - | - |

CRD, cone-rod dystrophy; CD, cone dystrophy; RD, retinal dystrophy; ACHM, achromatopsia; LCA, Leber congenital amarousis; RP, retinitis pigmentosa; AR, autosomal recessive;

AD, autosomal dominant; XL, X-linked.
*The nucleotide position of each genetic variant is based on the following GenBank cDNAs (accession number): *CACNA1F* (NM_001256789), *CNGB3* (NM_019098), *CNGA3* (NM_001079878), *CEP78* (NM_032171), *RPGRIP1* (NM_020366), *CYP4V2* (NM_207352), *ABCA4* (NM_000350). Nucleotide numbers reflect cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence. The initiation codon is codon 1.
**Published by [Namburi et al., 2016].

Preliminary data analysis did not reveal potential disease causing variants in the remaining 11 families.

Known disease causing mutations were identified in the *CNGB3* gene in families MOL1190 (a combination of the c.1148delC (p.T383fs) and the c.819_826del (p.P273fs) (Kohl et al., 2000, Sundin et al., 2000) changes) and MOL0563 (a homozygous c.1148delC (p.T383fs) variant) and in the *ABCA4* gene in family MOL0584 (a combination of the c.4919G>A (p.R1640Q) (Rozet1998) and the c.5318C>T (p.A1773V) (Chacon-Camacho et al., 2013) variants).

Novel genetic variants were identified in the *CACNA1F* gene in family MOL0061 (a hemizygous c.4051C>T (p.R1351*) nonsense variant), in the *CNGB3* gene in family MOL0360 (a homozygous c.782A>G (p.D261G) variant), in the *CNGA3* (cyclic nucleotide gated channel alpha 3) gene in family MOL0474 (a heterozygous c.931G>T (p.G311C) variant previously seen in a family with similar cone-dominate retinal phenotype), in the *RPGRIP1* gene in family MOL1124 (a homozygous c.2249A>G (p.Y750C) variant), in the *CYP4V2* gene in family MOL0499 (homozygous variant c.1508G>A (p.G503E) in combination with a known heterozygous variant c.C467T (p.S156F) (Sundin et al., 2000) in the *CNGB3* gene) and in the *CEP78* gene in family MOL0679 (homozygous splice-site variant (c.893-1G>A) (Namburi et al., 2016)). Mutations in the *CEP78* gene have not been previously reported to cause retinal disease and members of the MOL0679 family were diagnosed with a unique combination of phenotypes that included CRD and sensorineural hearing loss (Namburi et al., 2016). Targeted screening of additional families presenting a phenotype similar to that observed in family MOL0679 led to the identification of four additional families (MOL0773 (homozygous for the c.893-1G>A varinat), MOL1310 and TB279 (a homozygous c.534delT deletion), TB365 (compound heterozygous for the c.893-1G>A and

the c.534delT (p.Lys179Argfs*10) variants) that carry *CEP78* variants (Table 4) (Namburi et al., 2016). Validation and co-segregation analysis using Sanger sequencing was performed for the majority of identified variants (Table 4).

# Discussion

The genetic and phenotypic heterogeneity of RDDs greatly limit our ability to successfully identify the genetic cause of disease by systematical screening of all currently known genes. Tools such as mutation detection arrays and homozygosity mapping offer limited possibilities as they are mostly efficient only in specific populations and family structures. WES is an efficient tool for high throughput analysis that offers a relatively rapid and simple process for the analysis of a large set of candidate genes in large patient cohorts. Over the past several years WES was successfully used for identification of novel retinal disease genes as well as novel mutations in know genes (Estrada-Cuzcano et al., 2012, Khateb et al., 2012, Roosing et al., 2013, Sergouniotis et al., 2014, Beryozkin et al., 2015, Kohl et al., 2015, Lazar et al., 2015a, Lazar et al., 2015b) because it allows simultaneous assessment of a large number of nucleotide changes affecting coding genetic regions.

As part of an effort to determine phenotype and genotype in Israeli patients presenting RDD and discover correlations between specific ethnic origins and founder mutations, we have identified 11 novel and 10 known genetic variants in 11 genes associated with retinal disease. WES led to the identification of a potential genetic cause of disease in 16 out of 28 studied families. Disease causing mutations were identified in 8 additional families using a targeted screening approach. Our studies led to the identification of a novel disease gene *CEP78* identified in families presenting a unique combination of phenotypes including cone dominated retinal degeneration and sensorineural hearing loss (Namburi et al., 2016). In addition our studies revealed *GUCY2D* gene mutations as a major cause of cone dystrophy in the Israeli population (Lazar et al., 2015a), implicate *ALMS1* as a disease-related gene for nonsyndromic CRD (Lazar et al., 2015b) in contrast to previous findings that linked *ALMS1* mutations to Alström syndrome (Marshall et al., 2015) and identify *DYSF* gene mutations as a cause of co-occurring phenotypes

in a family with CRD and mild muscular dystrophy (Lazar et al., 2015b).

The presented results further support the use of WES for identification of the genetic cause of disease in specific families as well as identification of novel and known genetic variants as a source of information for further investigation by targeted screening of studied population groups. WES is an efficient tool for novel disease gene discovery even in small families with isolated cases. In addition WES allows identification and evaluation of modifier variants that can potentially influence disease progression and result in phenotype variability among different members of the same family.

The limitations of WES are illustrated by the inability to identify potential disease associated variants in 12 of the studied families. Additional data analysis may reveal potential disease associated genetic variants in novel disease genes in some of these families. However, due to the limitations of exome sequencing in detecting non-coding or deep intronic variants, long range genomic alterations and other structural variants, as well as the possibility of insufficient coverage for specific exons, the cause of disease in some of the studied families may remain unresolved. WGS will soon provide an opportunity for a more in-depth analysis of coding and non-coding genomic regions alike and offer a source of additional findings in families where potential disease associated genetic variants could not be previously identified. With the discovery of novel genes and genetic defects there is an ever increasing need for understanding molecular disease mechanisms for development of novel treatment strategies (Veleri et al., 2015). Model organisms and especially mouse models that mimic human disease genetically and produce similar phenotypic characteristics are the most widely used tool for gaining additional knowledge about disease mechanisms and treatment development (Veleri et al., 2015). In addition, experimental systems such as stem cells and induced pluripotent stem cell technology will aid the study of pathogenic mechanisms as well as initial steps of therapy development. The next important step beyond genetic discovery will be to decipher the functional properties of proteins affected in retinal degenerative diseases.