# Identification of Rare Variants Involved in High Myopia Unraveled by Whole Genome Sequencing

Annechien E.G. Haarman, MD,[1,2] Caroline C.W. Klaver, MD, PhD,[1,2,3,4] Milly S. Tedja, MD,[1,2]
Susanne Roosing, PhD,[5,6] Galuh Astuti, MSc,[5] Christian Gilissen, PhD,[5] Lies H. Hoefsloot, MD, PhD,[7]
Marianne van Tienhoven, BSc,[7] Tom Brands, Ing,[7] Frank J. Magielsen, Ing,[7] Bert H.J.F.M.M. Eussen, Ing,[7]
Annelies de Klein, PhD,[7] Erwin Brosens, PhD,[7,*] Virginie J.M. Verhoeven, MD, PhD[1,7,*]

***Purpose:*** Myopia (nearsightedness) is a condition in which a refractive error (RE) affects vision. Although common variants explain part of the genetic predisposition (18%), most of the estimated 70% heritability is missing. Here, we investigate the contribution of rare genetic variation because this might explain more of the missing heritability in the more severe forms of myopia. In particular, high myopia can lead to blindness and has a tremendous impact on a patient and at the societal level. The exact molecular mechanisms behind this condition are not yet completely unraveled, but whole genome sequencing (WGS) studies have the potential to identify novel (rare) disease genes, explaining the high heritability.

***Design:*** Cross-sectional study performed in the Netherlands.

***Participants:*** We investigated 159 European patients with high myopia (RE > −10 diopters).

***Methods:*** We performed WGS using a stepwise filtering approach and burden analysis. The contribution of common variants was calculated as a genetic risk score (GRS).

***Main Outcome Measures:*** Rare variant burden, GRS.

***Results:*** In 25% (n = 40) of these patients, there was a high (> 75th percentile) contribution of common predisposing variants; that is, these participants had higher GRSs. In 7 of the remaining 119 patients (6%), deleterious variants in genes associated with known (ocular) disorders, such as retinal dystrophy disease (prominin 1 [*PROM1*]) or ocular development (ATP binding cassette subfamily B member 6 [*ABCB6*], TGFB induced factor homeobox 1 [*TGIF1*]), were identified. Furthermore, without using a gene panel, we identified a high burden of rare variants in 8 novel genes associated with myopia. The genes heparan sulfate 6-O-sulfo-transferase 1 (*HS6ST1*) (proportion in study population vs. the Genome Aggregation Database (GnomAD) 0.14 vs. 0.03, P = 4.22E-17), RNA binding motif protein 20 (*RBM20*) (0.15 vs. 0.06, P = 4.98E-05), and MAP7 domain containing 1 (*MAP7D1*) (0.19 vs. 0.06, P = 1.16E-10) were involved in the Wnt signaling cascade, melatonin degradation, and ocular development and showed most biologically plausible associations.

***Conclusions:*** We found different contributions of common and rare variants in low and high grade myopia. Using WGS, we identified some interesting candidate genes that could explain the high myopia phenotype in some patients.

***Financial Disclosure(s):*** The author(s) have no proprietary or commercial interest in any materials discussed in this article. *Ophthalmology Science 2023;3:100303 © 2023 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).*

*Supplemental material available at www.ophthalmologyscience.org.*

Nearsightedness (myopia) is a refractive error (RE) in which the optics of the eye fails to focus light rays entering the eye exactly on the retina, leading to blurred vision. This ocular condition is caused by an elongated length of the eye axis; the axial length (AL) of an emmetropic eye is on average 23.5 mm compared with at least 26 mm in high myopia (Fig 1). People with myopia have clear vision when they focus on objects nearby but blurred vision when looking at objects positioned further away. Myopia is often defined as an RE ≤ −0.5 diopters (D) and high myopia as an RE ≤ −6 D.[1] The prevalence of myopia and high myopia is increasing globally, which will put a significant burden on our health care system caused by the visual consequences.[2,3] The RE of myopia can be corrected by wearing glasses or contact lenses, but the visual impairment caused by complications (myopic macular degeneration, retinal detachment, glaucoma, and cataract) at older age is often irreversible.[4,5] Unraveling the etiology of myopia may help to find targets for prevention and therapy.

Myopia is the result of a combination of environmental influences and genetic factors. More than 500 common

genetic factors with a relatively small effect size have been identified through genome-wide association studies (GWASs) with large sample sizes (N = 160 420 to N = 542 934).[6,7] These genetic factors cumulatively explain around 18% of the heritability of the trait, whereas estimates of the total RE heritability in several other studies are around 70%.[7,8] Rare genetic factors with potentially large effect sizes could partly explain this missing heritability because previous GWAS studies were not designed nor had sufficient power to identify rare variants. Previous whole exome sequencing studies using both population-based cohorts, as well as high myopia families, identified some interesting new candidate genes harboring rare variants with roles in several pathways (see Tedja et al[8] for an overview of whole exome sequencing studies and RE).[9,10] Because these studies performed genetic techniques only covering the exome, application of whole genome sequencing (WGS) could identify new and relevant additional genes and could identify rare variants underlying previous GWAS signals.

Several pathways and processes have been linked to myopia, including retinal light transduction, extracellular matrix remodeling, cell cycle pathways, and cell growth.[6,7] The involved genes are expressed across all retinal cell types and probably function in a light-dependent retina-to-sclera signaling cascade causing axial elongation and myopia.[6,7] The role of these retinal cell types and the interaction between different processes in these retinal cells in myopia development are not completely understood.

Whole genome sequencing has the advantage that different types of genetic variation over the entire genome can be evaluated, including coding and noncoding single nucleotide variants (SNVs), copy number variants (CNVs), and structural variants (SVs). Rare variants (minor allele frequency < 1%) have a potentially large impact and could therefore explain a large part of the missing heritability of myopia. Until now, WGS has not been applied frequently to myopia and was only performed in selected families with high myopia as a follow-up study after linkage analysis to pinpoint a genetic locus.[11] In other complex genetic diseases, such as schizophrenia, diabetes, or retinitis pigmentosa, however, new association signals were identified or genetic diagnoses could be made with the use of WGS.[12–16] Given the complex genetic nature of myopia, in which the genetic contribution to the etiology is expected to be larger on the extreme end of the myopia spectrum, focusing on high myopia is a logical first step. Therefore, we studied a highly myopic population (RE > −10 D) of European ancestry to identify rare SNVs, CNVs, and SVs using WGS with high potential impact. We applied a stepwise approach focusing on putative pathogenic variants and subsequently performed a broader evaluation of genes of interest by evaluation of the topologically associating domain (TAD) regions.

## Methods

### Study Population

The Myopia Study (MYST) is a case−control study, conducted from 2010 to 2020 in Rotterdam, The Netherlands.[17] Participants were recruited via public media, eye care providers, and advertisements on the website for MYST and Erasmus Medical Center. People aged < 18 years and patients with a history of a confirmed syndromic cause of myopia were excluded.[17] For the current study, we included extreme high myopia patients (mean RE ≤ −10D) with available DNA and European ancestry (based on questionnaire data). We used genetic and ophthalmic data from the Rotterdam Study (RS) 1-3 for comparison of genetic risk score (GRS) with RE. The RS is a long running, prospective population-based study conducted in the city district Ommoord in Rotterdam, The Netherlands.[18] Only individuals with available genetic data and RE were included. Both studies adhered to the tenets of the Declaration of Helsinki and were approved by the local ethics committees of Erasmus Medical Center, Rotterdam, The Netherlands (MEC-02-1015 [RS] and MEC 2009-248 [MYST]). All participants provided informed consent.

### Ophthalmic Data

All participants from MYST and RS underwent extensive ophthalmic examinations, including RE measurements with Topcon RM-A2000 Auto-Refractor (Topcon Optical Company) and AL measurements with Lenstar LS900 (Laméris Ootech), or with the A-scan function of PacScan 300 AP (Sonomed Escalon) for participants with AL ≥ 30 mm. Measurements of both eyes were averaged; when these were missing in one eye (due to, e.g., cataract surgery without knowledge of prior RE), the measurement of the available eye was used. Refractive error was calculated as the sum of the full spherical value and half of the cylindrical value. In addition to ophthalmic examinations, questions about the family (e.g., parental myopia, RE of the parents) and ocular history were included in MYST.

### Genetic Analysis: WGS

Genomic DNA was isolated from ethylenediaminetetraacetic acid anticoagulated blood according to standard protocols. All samples were measured for sample purity and integrity using picogreen, ultraviolet 260/280, and agarose gel electrophoresis. Four-microgram double-stranded DNA was used as input for WGS by the Beijing Genomics Institute (BGI) (n = 156) on a BGISeq500 using a 2 × 100-bp paired-end reads, with a minimal median coverage per genome of 30-fold. Three samples, previously sent to Novogene for WGS, were also included. Novogene used 1-μg double-stranded DNA as the input for WGS on an Illumina HiSeq PE150 platform using a 2 × 150-bp paired-end reads, with a minimal median coverage per genome of 30-fold. The WGS data were mapped to the human genome (hg19) using Burrows-Wheeler Aligner. All data were based on percentage mapped reads, coverage, bases with > 20× coverage, error rate, insert size, and percentage duplicated mapped reads evaluated using Qualimap V.2.2.1 (BGI) and SAMtools and Picard (Novogene) (Table S1, available at www.ophthalmologyscience.org).[19] Calling of variants was performed using xAtlas V.0.1 (BGI) and GATK (Novogene).[20] Structural variants were detected based on paired-end and split-read evidence using Manta Structural Variant Caller V.1.1.0 (Illumina) with default parameters (WGS from BGI only).[21]

### GRS Calculation

For calculation of a GRS, we used the WGS data of MYST as described above and genotyping array data from RS.[18] Genetic risk score was calculated as the weighted number of risk alleles carried for 528 genetic variants associated with RE, which were identified in a GWAS by the Consortium for Refractive Error and Myopia, 23andMe, and the UK Biobank (N = 620 035) (Tedja MS,

Verhoeven VJM, Eriksson N, et al. Pathway-Specific Genetic Risk Scores Associated with High Myopia. Presented at: International Myopia Conference; September 12, 2019; Tokyo, Japan).[6,7] The regression coefficients for association with RE in D in the total sample were used as the weightings ($GRS = \sum_{i=1}^{k} \beta_i N_i$, where β indicates the weightings and N the number of risk alleles for each locus [indicated by i]).

## Variant Filtering and Prioritization

All variants were initially prioritized according to predefined criteria: we included variants with (1) coverage > 10 reads; (2) allele frequency < 1% in either 1000 Genomes (1KG), Exome Variant Server, Exome Aggregation Consortium, or the Genome Aggregation Database (GnomAD) V2.1.1 (ALL, including Finnish population)[22] and an in-house control cohort (600 unaffected parents from patients with unrelated phenotypes); (3) Combined Annotation-Dependent Depletion (CADD) score > 20; and (4) protein altering variant annotation including putative splice site and splice region alterations. We used Fabric genomics (V6.8, pipeline V6.6.6, https://app.fabricgenomics.com/) software to filter variants and used Plink1.9 to create a multisample variant call format (VCF) file (Fig S2, available at www.ophthalmologyscience.org). ANNOVAR (V.10-2019)[23,24] was used for annotation of all variants of this multisample VCF and the online available CADD score tool (V1.6) for assessing the CADD score.[25]

We applied a stepwise approach to identify new rare variants associated with myopia (Fig S2).

**1. Filtering Based on GRS.** We stratified the total study population (MYST) based on GRS and RE into 4 groups: (group 1) low RE (i.e., more extreme myopia, > −6D) and high (i.e., positive value or > 75th percentile) GRS; (group 2) low RE and low GRS; (group 3) high RE and high GRS; and (group 4) high RE and low GRS. Note that because myopia is on the negative side of the RE distribution, a more negative GRS is associated with a more negative RE, i.e., myopia, and vice versa. For illustrative purposes, we also plotted the GRS against RE of RS (Fig 3). Note that no WGS data were available for RS and that RS data were not used in methodology steps 2 to 5. A high GRS indicates the contribution of many common variants predisposing for myopia (Fig 3). Because GRSs are based on the sum of common haplotypes associated with myopia, we expected that individual patients with low RE and high GRS (group 2) would have a lower contribution of rare variants, and therefore, we excluded this group in the rare variant analyses (steps 2 to 4).[26]

**2. Exclusion of Patients with a Known Ocular Syndrome.** Although we excluded patients with clinically confirmed syndromic myopia, variable expressivity of key phenotypic features could potentially have masked syndromic patients. Therefore, we evaluated 348 exonic and splice site variants in known (Mendelian) ocular disease genes included in the Erasmus Medical Center whole exome sequencing gene panel (N = 477 genes; Table S2, available at www.ophthalmologyscience.org).[27] Variants of interest were validated using Sanger sequencing and classified according the American College of Medical Genetics and Genomics criteria using an online tool (https://franklin.genoox.com/clinical-db/home). We analyzed the clinical signs of the corresponding Online Mendelian Inheritance in Man phenotype in these individuals, using patient record information if available.[28,29] If an SNV in a gene associated with an autosomal recessive (AR) disorder was detected, we continued with screening the gene region for SVs using Manta output. Patients were only excluded for subsequent analyses when a pathogenic variant with a corresponding clinical phenotype was confirmed in that patient.
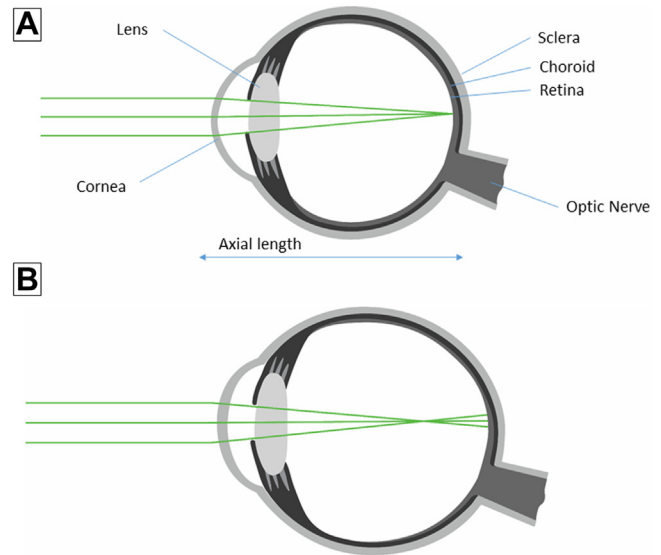


**Figure 1.** Schematic overview of (**A**) emmetropia and (**B**) myopia. In emmetropia, the optics of the eye focus light rays entering the eye exactly on the retina. In myopia, the ocular globe is elongated, and the light rays are focused in front of the retina. The different ocular components are indicated.

**3. Prioritization of Putative Pathogenic Variants on a Gene Level.** For individuals with a relative low burden of common variants (low RE [≤ −10D] and low GRS [< 75th percentile] [2.38 to −5.44]), we filtered all variants on a gene level, based on the frequency in the cohort (> 30 in our cohort) and CADD score (> 20). We did not apply a gene panel as in step 2. Highly variable genes (low missense Z-score) or genes with an extremely high variant frequency in our cohort (> 30) were excluded (Table S3, available at www.ophthalmologyscience.org). Of all genes with a variant present in 5 to 29 individuals, a burden gene-based test was performed to assess its association with myopia (see methods in the following paragraphs). We excluded genes reported to be highly tolerant for rare missense variants, that is, with a GnomAD (V2.1.1) missense Z-score < 1.[22] We compared the number of synonymous variants (with minor allele frequency < 1%) within a gene between the 2 datasets (119 individuals with high myopia from our study vs. GnomAD) to evaluate the presence of technical artifacts.

**4. Genome-Wide Evaluation of Genes of Interest.** Our study lacked the power to perform a rare variant burden test genome-wide because of its sample size. Therefore, we initially focused on rare putative deleterious variants impacting conserved amino acids (steps 1−3) and subsequently evaluated the TAD region of these gene regions (8 genes) for rare SVs and rare deleterious SNVs using a more lenient CADD cutoff (CADD score > 15). This CADD score focuses on the 5% most damaging variants predicted across the genome (http://cadd.gs.washington.edu/info).[25] Topologically associating domain regions are evolutionary conserved across species, contribute to the 3-dimensional nuclear organization, and are involved in regulation of gene expression.[30] We used TAD borders as described in brain cortical tissue as a proxy for ocular tissue.[31]

**5. Cohort-Wide Evaluation.** As a final step, we evaluated the rare variants in the most promising genes from step 3 in the total study population, including those with many common variants excluded in step 1. We analyzed the contribution of common variants (GRS) and rare variants. The rare variant status was
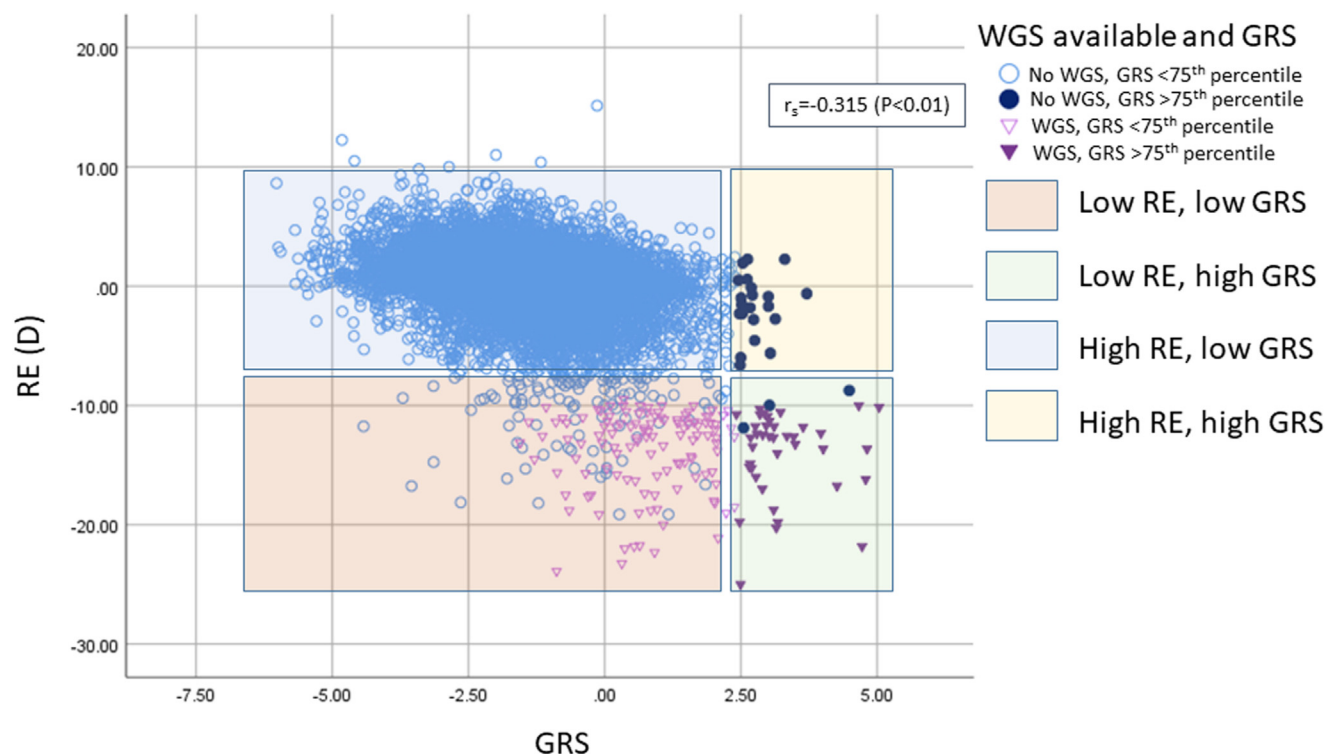
**Figure 3.** Genetic risk score in relation to spherical equivalent (diopters [D]) in our high myopia cohort and the Rotterdam Study (RS). Genetic risk score (GRS) was calculated using betas from 528 single nucleotide polymorphisms from Tedja et al. The different colors indicate 4 different groups: red, low refractive error (RE) and low GRS (< 75th percentile); green, low RE and high GRS (> 75th percentile); blue, high RE and low GRS; and yellow, high RE and high GRS. The 75th percentile of GRS was calculated in the high myopia cohort (MYST) and applied to the total group. The GRSs of the RS (included in the initial discovery genome-wide association study) were plotted for illustrative purposes. Spearman's correlation coefficient is also shown ($r_2$). WGS = whole genome sequencing.

defined as having > 1 variant in any of the candidate genes. The common variant status was defined as having a GRS below (−) or above (+) the 75th percentile, indicating less common variants and more common variants, respectively.

### Association Test and Statistical Analyses

An association burden test was performed as described previously.[32] In short, we compared the total frequency of variants in the genes of step 3 in our study population to the frequency in GnomAD (V2.1.1) as a control population (N = 141 456) using a chi-square test.[22] We used the gene borders defined by GnomAD to assess the genomic context of every gene. We applied a dominant model; that is, for each gene, the number of individuals in our study population who carried ≥ 1 qualifying variant in that gene were assessed. We used the sum of homozygote and heterozygote counts per gene from GnomAD as a control population estimate. As the input for these tests, we first created a gene-specific VCF file containing all unprioritized variants in every gene for 1 individual using Plink1.9.[33,34] Next, we created a multisample VCF file, including all individuals, and annotated this VCF file using ANNOVAR (V.10-2019) and CADD (V1.6).[23−25] As a next step, we excluded all variants within these genes with frequency > 1% in 1KG, Exome Variant Server, or Exome Aggregation Consortium and a BGI control dataset consisting of 702 unrelated individuals, excluded synonymous variants, and looked at all variants with a CADD score > 15 or a Genomic Evolutionary Rate Profiling score > 2.[35] For the burden test analyses, a $P$ value < $2.5 \times 10^{-6}$ (Bonferroni

correction for approximately 20,000 genes) was considered statistically significant.

We compared differences in means between groups using an independent sample $t$ test or analysis of variance for normally distributed data (GRS, age at inclusion, and age of stabilization) and the Mann−Whitney $U$ test or Kruskal−Wallis test for non-normally distributed data (RE, AL, age of stabilization, age of onset). Frequencies between groups were compared using the chi-square test or Fisher exact test. Correlation between RE and GRS was assessed by Spearman's correlation coefficient. The IBM SPSS Statistics version 25 (IBM Corp) was used for the statistical analyses. Evidence for expression in ocular tissue of all genes of interest was assessed using the recently published human ocular dataset from Cowan et al[36] and the online tool Spectacle, which includes data from 5 human datasets (all-retina-choroid-retinal pigment epithelium).[37−42] Because these are exploratory analyses, any expression higher than 0 was reported positive. Evidence for an ocular phenotype in a knockout mouse or zebrafish model was assessed using the Mouse Genome Informatics database,[43] the International Mouse Phenotyping Consortium databases,[44,45] and the Zebrafish Information Network.[46]

### Pathway Analysis

We performed 2 pathway analyses to investigate the potential functions and pathways of the associated genes. First, we conducted a Gene Ontology term enrichment analysis (Gene Ontology database released February 1, 2021) using the PANTHER over-representation test (Released February 24, 2021) and all *Homo*

*sapiens* genes as a reference.[47,48] We used the Gene Ontology biological process, molecular function, and cellular component as annotation datasets. Statistical evaluation using the Fisher exact and false discovery rate (FDR) correction for multiple testing (FDR < 0.05) was performed. Second, we used Ingenuity Pathway Analysis (QIAGEN Inc) for canonical pathway analysis and causal network analysis, using humans as the reference dataset.[49]

For both analyses, we analyzed the gene content of 2 groups of genes: genes with missense variants with a GnomAD Z-score > 1 and a frequency in our cohort ≥ 5 or startloss, stopgain, stoploss, and splicing variants with a frequency of ≥ 2 in our cohort (group 1) and genes with startloss, stopgain, stoploss, and splicing variants irrespective of their frequency in our cohort (group 2).

## Results

Our study population (N = 159) had a mean (standard deviation [SD]) age of 48.6 (12.9) years and was composed of 57 (36%) male participants (Table 4). Because this female predominance is not related to our outcome measures, it therefore will probably not lead to any bias. Mean (SD) RE was −13.8 D (3.4), and 30% of our population had RE −15 D or worse. Myopia onset was known in a subset of participants (n = 156 [98.1%]) and was on average (SD) 7.5 (3.5) years. Considering the family history, 43% had a myopic mother and 46% a myopic father with a mean (SD) RE of −5.2 D (3.8) and −5.0 D (4.6), respectively. The majority (58%) had both a myopic father and a myopic mother.

### 1. Myopia Explained by GRS

The mean GRS in our study population (MYST) was 1.44 (1.44) and the range was −1.55 to 5.40 compared with a mean of −0.80 (1.40) and a range of −5.44 to 5.40 in the individuals with any myopia in the total (MYST + RS) population. We compared the distribution of GRS in our study population to the population-based RS (n = 11 496). The mean RE of RS was 0.46 D (SD 2.59), and the mean GRS was −1.45 (SD 1.30). We pooled both populations together and observed a negative correlation between RE and GRS, that is, a higher GRS in more myopic participants (i.e., more negative RE) (Spearman's ρ = −0.315, P < 0.01). We could identify 4 groups in our pooled population: (1) low RE (i.e., more myopic, negative RE) and low GRS; (2) low RE and high GRS; (3) high RE (less myopic, more positive RE), low GRS; and (4) high RE and high GRS (Fig 3). We observed 40 (25%) participants of MYST in the second group. Because these individuals with many common variants are less likely to carry rare (high potential impact) variants, we excluded this group in subsequent analyses from the MYST population (40/159). The remaining individuals (n = 119) were included in the initial burden analysis.

When taking a closer look at the associated common loci in MYST, we identified a higher-than-expected frequency of the missense variant rs5442 in G protein subunit beta 3 (*GNB3*) in our cohort compared with other cohorts (27% vs. 14% vs. 10% vs. 9% vs. 5% in MYST and RS [included in the discovery cohort of this variant]),[6,7] 849 European controls, Genome of the Netherlands (GoNL) (V1.0.0),[50] and GnomAD (V2.1.1),[22] respectively. This gene was highly ranked in the GWAS study of Tedja et al,[6] based on the significant association signal of the exonic variant located in this gene, the expression in ocular tissue, and phenotype in both humans (congenital stationary night blindness) and animal studies (ocular globe enlargement). We screened the *GNB3* gene with a 10 000-kb flanking region for a potential causal variant present in the 26 carriers of this SNP and absent or with low frequency in the remaining myopes. We did not identify other potential causal variants explaining the previous GWAS signal. Moreover, we did not observe rare high impact variants located in the region of other common GRS loci associated with RE, occurring with high frequency in this study cohort.

### 2. Exclusion of Patients with a Known Myopia Syndrome

Patients with obvious phenotypical characteristics of a myopia syndrome were excluded from participation in our study. Nevertheless, we were interested in deleterious variants in known ocular disease genes. After excluding common variants and applying our prioritization protocol, we observed 348 SNVs in 192 known ocular disease genes (Fig S2). The majority of these SNVs (89%) were missense or silent variants, 10% were nonsense variants, and 1% were canonical splice site variants. We identified 11 nonsense and splice region variants with a minor allele frequency < 1% in genes associated with autosomal dominant (AD) disorders, 26 rare variants in genes associated with AR disorders, and 1 rare variant in a gene associated with a disorder with both AD and AR inheritance patterns. The prevalence of these ocular diseases is < 5 per 10 000. All individuals were heterozygous for these variants. Therefore, we screened the AR genes for deletions but did not identify any deletions in addition to the SNVs annotated to these genes (Table S5, available at www.ophthalmologyscience.org). We could validate 6 SNVs in 6 individuals associated with an AD disorder and 1 SNV associated with a disorder causing both AD and AR phenotypes (Table 6). These 7 variants were located in genes associated with (ocular) developmental disorders (*ABCB6, TGIF1*), retinal dystrophy (*PROM1*), corneal dystrophy (AGBL carboxypeptidase 1 [*AGBL1*]), cataract (crystallin beta A1 [*CRYBA1*], glucosaminyl (N-acetyl) transferase 2 [*GCNT2*]), and isolated high myopia (prolyl 4-hydroxylase subunit alpha 2 [*P4HA2*]). The GRS and RE of carriers of these variants were not significantly different compared with noncarriers (mean [SD] GRS −0.78 [0.97] vs. −1.05 [0.86], P = 0.487 and mean [SD] RE −14.5 D [3.4] vs. −13.7 D [3.4], P = 0.560 for carriers vs. noncarriers, respectively).

In addition to nonsense and splice site variants, we identified 309 missense variants (99% in heterozygous

Table 4. Baseline Characteristics of the Total Study Population (n = 159)

| Variable | Available for n (%) | Total Population (n = 159) | Low RE (< −10D) and High GRS (Above 25th Percentile) (n = 119) | Low RE (< −10D) and Low GRS (Below 25th Percentile) (n = 40) | P value |
|---|---|---|---|---|---|
| Age at inclusion, y | 159 (100) | 48.6 (12.9) | 48.1 (12.6) | 49.8 (13.9) | 0.496 |
| Range | | 21.0−85.0 | | | |
| Gender, % male | 159 (100) | 57 (35.8) | 76 (63.9) | 26 (65.0) | 0.897 |
| RE, D | 158 (99.4) | −13.8 (3.4) | −13.8 (3.4) | −14.0 (3.6) | 0.589 |
| Range | | −25.0 to −9.5 | | | |
| −10.0D to −15.0D (%) | 158 (99.4) | 111 (70.3) | 84 (70.6) | 27 (69.2) | 0.872 |
| ≤ −15.0D (%) | | 47 (29.7) | 35 (29.4) | 12 (30.8) | |
| AL, mm | 143 (89.9) | 28.6 (1.6) | 28.6 (1.7) | 28.6 (1.4) | 0.948 |
| Range | | 24.4−32.0 | | | |
| Age of onset, y | 156 (98.1) | 7.5 (3.5) | 7.4 (3.7) | 7.6 (2.4) | 0.161 |
| Range | | 2.0−27.0 | | | |
| Age of stabilization | 85 (53.5) | 26.2 (13.3) | 26.3 (14.6) | 25.8 (8.7) | 0.835 |
| Range | | 0−65 | | | |
| Myopia father | 90 (56.6) | 73 (45.9) | 50 (83.3) | 23 (76.7) | 0.446 |
| RE in case of myopia, D | 63 (37.7) | −5.0 (4.6) | −5.1 (4.3) | −4.8 (5.1) | 0.347 |
| Range | | −24.1 to −0.5 | | | |
| Myopia mother | 99 (62.3) | 69 (43.4) | 46 (64.8) | 23 (82.1) | 0.091 |
| RE in case of myopia, D | 54 (34.0) | −5.2 (3.8) | −5.1 (4.0) | −5.2 (3.6) | 0.660 |
| Range | | −16.00 to −0.25 | | | |
| Myopia | | | | | |
| Both parents, % | 71 (44.7) | 41 (57.7) | 24 (52.2) | 17 (68.0) | 0.489 |
| One parent, % | | 21 (29.6) | 15 (32.6) | 6 (24.0) | |
| None of parents, % | | 9 (12.7) | 7 (15.2) | 2 (8.0) | |
| GRS | 159 (100) | −1.44 (1.44) | −0.80 (0.96) | −3.3 (0.79) | 3.75E-32 |
| Range | | −5.40 to 1.55 | | | |

Age of onset of myopia is displayed, as well as age of RE stabilization. Continuous variables are presented as mean (SD) and categorical variables as n (%). Genetic risk score was calculated using 529 single nucleotide polymorphism (SNPs) from Tedja et al.[8] P value was calculated using the chi-square test or exact test (myopia parents) for categorical variables and the t test (GRS, age at inclusion, and age of stabilization) or Mann−Whitney U test (RE, AL, age of stabilization, and age of onset) for continuous variables.
AL = axial length; D = diopters; GRS = genetic risk score; RE = refractive error; SD = standard deviation.

form) occurring in 0.8% (1/119) to 1.7% (2/119). Two hundred twenty (72%) were predicted as deleterious by SIFT, PolyPhen, and MutationTaster, and 87 of these 220 variants were located in genes associated with AD disorders. Variants in ATP binding cassette subfamily B member 6 (*ABCB6*), ATP binding cassette subfamily A member 4 (*ABCA4*), cadherin related 23 (*CDH23*), crystallin gamma D (*CRYGD*), forkhead box E3 (*FOXE3*), myosin VIIA (*MYO7A*), and phosphodiesterase 6B (*PDE6B*) and > 4 different variants per gene were observed.

We observed 237 768 rare (< 1%) SVs in 4644 genes, of which 2239 SVs were annotated to a gene associated with a known ocular disorder. Most of these SVs were insertions (62%, length 1−10 bp) followed by translocations (33%) and deletions (4%, length 51−7656 bp). Structural variants were in centrosomal protein 290 (*CEP290*), chromodomain helicase DNA binding protein 7 (*CHD7*), eyes shut homolog (*EYS*), Fraser extracellular matrix complex subunit 1 (*FRAS1*), protocadherin related 15 (*PCDH15*), and WD repeat containing planar cell polarity effector (*WDPCP*), and we identified > 100 SVs. Based on these analyses, we did not exclude any individual for subsequent analysis.

## 3. Prioritization of Putative Pathogenic Variants Affecting Coding Sequences and the Splice Region

In this analysis step, we did not apply an ocular disease gene panel. Of the 8635 potentially deleterious (CADD > 20) variants, 595 were nonsense variants and 112 were splice site variants located in 638 unique genes. Nonsense variants or splice site variants in chromodomain helicase DNA binding protein 5 (*CHD5*), hemoglobin subunit zeta (*HBZ*), oncomodulin (*OCM*), parkin coregulated like (*PACRGL*), pericentriolar material 1 (*PCM1*), ribonuclease L (*RNASEL*), SPNS lysolipid transporter 3, sphingosine-1-phosphate (*SPNS3*), or WASH complex subunit 2A (*WASHC2A*) were the most frequent (0.025% to 4.2%). We ranked all 8635 variants, including missense variants, based on frequency and identified variants with a frequency > 5 (4.2%) in 32 genes (Table S7, available at www.ophthalmology science.org). Seventeen genes were excluded because of the high variability of these genes (a missense Z-score < 1), and therefore, we performed a burden test of the remaining 15 genes. The frequency of synonymous variants in dual specificity phosphatase 7 (*DUSP7*),

Table 6. Seven Missense and Splice Site Variants Located in Known Ocular Disease Genes Associated with AD Ocular Disorders According to OMIM

| Gene | Chr | Exon | Variant | HGVS | CADD | dbSNP | Classification | OMIM | Sex | RE (D) | GRS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ABCB6 | 2 | 4 | c.833- 1G>A: p.(?) | NC_000002.11: g.220080903C> T | 27 | rs1450661565 | Likely pathogenic | Dyschromatosis universalis hereditaria 3, 615402, AD; microphthalmia, isolated, with coloboma, 614497, AD; pseudohyperkalemia, familial, 2, due to red cell leak, 609153, AD | F | −12.44 | −1.39 |
| PROM1 | 4 | 23 | c.2470C>T: p.(Q824*) | NC_000004.11: g.15972698G> A | 36 | rs566891826 | VUS | Cone-rod dystrophy 12, 612657, AD, AR; retinitis pigmentosa 41, 612095, AR; Stargardt disease 4, 603786, AD; macular dystrophy, retinal, 2, 608051, AD | M | −11.88 | −0.73 |
| P4HA2 | 5 | 15 | c.1555C>T: p.(R519*) | NC_000005.9: g.131528750G> A | 41 | rs200583507 | VUS | Myopia 25, AD, 617238, AD | F | −11.19 | −1.33 |
| GCNT2 | 6 | 3 | c.14G>A:p. (W5*) | NC_000006.11: g.10529158G> A | 35 | rs185805779 | VUS | Cataract 13 with adult i phenotype, 116700, AR; adult i phenotype without cataract, 110800, AD | F | −19.13 | 0.10 |
| AGBL1 | 15 | 23 | c.3220C>T: p.(R1074*) | NC_000015.9: g.87217666C>T | 33 | rs185919705 | Pathogenic | Corneal dystrophy, Fuchs endothelial, 8, 615523, AD | F | −19.00 | −2.23 |
| TGIF1 | 18 | 1 | c.177C>A: p.(Y59*) | NC_000018.9:g.3452154C>A | 22.3 | rs121909070 | Pathogenic | Holoprosencephaly 4, 142946, AD | M | −15.25 | −0.04 |
| CRYBA1 | 17 | 5 | c.500+1G> C: p.(?) | NC_000017.10: g.27580801G>C | 25.7 | rs775038545 | Likely pathogenic | Cataract 10, multiple types, 600881, AD | F | −12.50 | −1.70 |

The variants and clinical details (sex, RE, GRS) of these patients are described. Variants were classified according to the ACMG guidelines using the online tool available through https://frank-lin.genoox.com/clinical-db/home. Genome build GRCh37 is used as reference. Associated clinical disorder from OMIM is described, as well as the phenotype MIM number and CADD score. Information of the family was available in 4 out of 7 patients. The patient with the ABCB6 variants had a myopic father (RE −7 D) and mother (RE −4 D). The parents of the patient with the P4HA2 variant were both mildly myopic (RE ∼ −2 D). The parents of the patient carrying the GCNT2 variant were not myopic. The patient with the CRYBA1 variant had a mother with myopia (RE −2 D).

ABCB6 = ATP binding cassette subfamily B member 6; ACMG = American College of Medical Genetics; AD = autosomal dominants; AGBL1 = AGBL carboxypeptidase 1; AR = autosomal recessive; CADD = Combined Annotation-Dependent Depletion; CRYBA1 = crystallin beta A1; D = diopters; dbSNP = The Single Nucleotide Polymorphism Database; F = female; GCNT2 = glucosaminyl (N-acetyl) transferase 2; GRCh37= Genome Reference Consortium Human Build 37; GRS = genetic risk score; HGVS = Human Genome Variation Society; M = male; MIM = Mendelian Inheritance in Man; OMIM = Online Mendelian Inheritance in Man; P4HA2 = prolyl 4-hydroxylase subunit alpha 2; RE = refractive error; TGIF1 = TGFB induced factor homeobox 1; VUS = variant of uknown significance.

dynein light chain LC8-type 1 (*DYNLL1*), intercellular adhesion molecule 5 (*ICAM5*), and insulin receptor substrate 2 (*IRS2*) was significantly higher in our cohort than in GnomAD (Table S8, available at www.ophthalmologyscience.org) (all $P < 10 \times 10^{-5}$), indicating potential technical artifacts. Of the remaining genes, ATP binding cassette subfamily D member 1 (*ABCD1*), RUNX family transcription factor 1 (*RUNX1*), and microtubule associated protein 6 (*MAP6*) showed the most significant association with high myopia (Fig 4 and Table S9, available at www.ophthalmologyscience.org). Furthermore, these genes were not enriched for any SVs. *RUNX1* was expressed in ocular pericytes, *MAP6* was expressed in cone cells, *HS6ST1* was expressed in both cones and rods, *RBM20* and *MAP7D1* were expressed in retinal ganglion cells, and *CHADL1* was expressed in retinal pigment epithelium cells (Fig 5 and Table S5).

## 4. Evaluation of the TAD Region of Loci of Interest

Lacking the sample size and control cohort for a formal burden analysis, we evaluated the TAD region of the 8 promising candidate genes identified in step 3. We did not find any deleterious noncoding variant, SV, or CNV influencing expression. We identified an intronic variant in *MAP7D1* that might affect expression of this gene. In addition to the 8 promising candidate genes, we observed rare variants in an additional 13 genes (Table S10, available at www.ophthalmologyscience.org). Eleven of these genes showed evidence for expression in ocular tissue, but only solute carrier family 6 member 8 (*SLC6A8*) was associated with a syndrome that can be characterized by an ocular phenotype. *SLC6A8* causes a syndrome characterized by hyperopia, which is RE with a positive value, that is, RE > 0.5 D (cerebral creatine deficiency syndrome Mendelian Inheritance in Man [MIM] 300352). None of these genes or variants were identified in previous GWASs.

## 5. Cohort-Wide Evaluation

First, we focused on the genes from step 3 (*ABCD1, PTGDR2, RPL19, RUNX1, STAT5A, DYNLL1, DUSP7, HS6ST1, RBM20, MAP7D1, MAP6, ICAM5, IRS2, DDX19B,* and *CHADL*) in the total population (N = 159). A higher proportion of the 119 individuals with low RE and high GRS had a pathogenic variant in the *ABCD1* gene than the 40 individuals with low RE and low GRS (*P* = 0.00058) (Table S11, available at www.ophthalmologyscience.org). When comparing the total population (159) to the GnomAD population, the association became stronger (i.e., more significant) for most loci (Table S11). We stratified individuals based on their burden of the most promising rare variant-enriched candidate genes (n = 8) and observed that individuals carrying more rare variants did not have a different contribution of common variants (i.e., lower GRS) (Fig S6). The RE did not differ between groups. We observed a trend toward a more myopic refraction in individuals with a combination of common and rare variants; however, this trend was not significant (Fig S6).

**Pathway Analyses.** We performed pathway analyses using both PantherDB and Ingenuity Pathway Analysis with 2 groups of input genes: (1) genes with missense variants with a GnomAD Z-score > 1 and a frequency in our cohort ≥ 5, or startloss, stopgain, stoploss, and splicing variants with a frequency of ≥ 2 in our cohort (72 genes); and (2) genes with startloss, stopgain, stoploss, and splicing variants irrespective of their frequency in our cohort (638 genes) (Tables S12−S14, available at www.ophthalmologyscience.org). Analysis of the 72 genes did not reveal any significant results in Panther DB, but analysis of the 638 genes identified cilium-dependent cell motility (*P* = 7.06E-06; FDR = 0.03), microtubule-based movement (*P* = 2.33E-05; FDR = 0.05), and organelle assembly (*P* = 1.13E-05; FDR= 0.03) as significant pathways (Table S12, available at www.ophthalmologyscience.org). The most significant molecular functions were ion binding (GO:0043167, FDR = 1.05E-05) and ATPase activity (GO:0016887, FDR = 1.13E-05); the most significant cellular components were cilium (GO:0005929, FDR = 1.49E-05) and microtubule cytoskeleton (GO:0015630, FDR = 2.61E-04). Using Ingenuity Pathway Analysis and 72 genes and 638 genes as input resulted in 28 and 13 significant pathways, respectively (Tables S13 and S14). These pathways were related to biotransformation, melatonin degradation, neuronal signaling, and collagen transformation. Among the top diseases and functions associated with these genes were cellular developments and cell morphology but also connective tissue disorders.

## Discussion

In this first WGS study including 159 individuals with high myopia, the genetic make-up was enriched with common variants in 25% of patients. In the remaining individuals of the high myopia study population, we screened the genome for rare exonic variants located in known ocular disease genes and identified variants in 7 genes associated with (ocular) developmental disorders (*ABCB6, TGIF1*), retinal dystrophy (*PROM1*), corneal dystrophy (*AGBL1*), cataract (*GCNT2, CRYBA1*), and isolated high myopia (*P4HA2).* Furthermore, we searched for highly deleterious genes enriched for SNVs, CNVs, and SVs in individuals with high myopia compared with controls and identified 8 novel candidate genes associated with high myopia. In the entire TAD region of the most promising candidate genes, we did not find any deleterious noncoding variant, SV, or CNV influencing expression except for an intronic variant in *MAP7D1*. Knockout mouse models of the genes *ABCD1, HS6ST1,* or *RBM20* were reported to have an ocular phenotype in mice (Tables S15 and S16, available at www.ophthalmologyscience.org). Given their expression in ocular tissue and ocular phenotype in knockout mouse models, *HS6ST1, RBM20,* and *MAP7D1* are the most promising genes with a potential role in the development of high myopia.

The WGS data allowed us to investigate the entire genomic region around the previously identified common
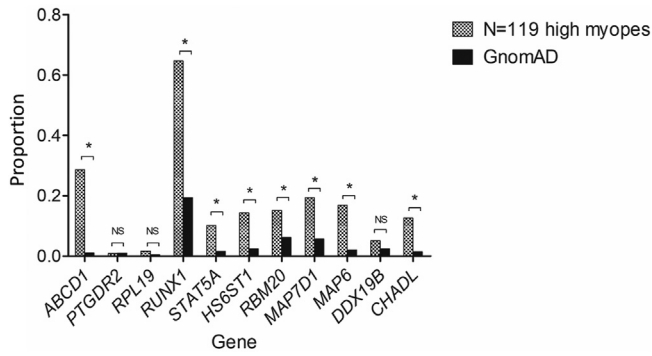
**Figure 4.** Burden test of top genes identified through whole genome sequencing in our study population (N = 119) with low refractive error and high genetic risk score. The frequency of variants in our study population (N = 119) and in the comparing control GnomAD study population (N = 141 456) are displayed. The burden test was performed as explained in detail in Guo et al.[32] The frequency of synonymous variants in *DUSP7, DYNLL1, ICAM5,* and *IRS2* was significantly higher in our cohort than in GnomAD (Table S8, all $P < 10 \times 10^{-5}$), indicating potential technical artifacts. *$P$ value of $< 10 \times 10^{-5}$. GnomAD = The Genome Aggregation Database; NS = not significant.

top hit in *GNB3*.[51] We did not find any variant with a similar frequency, and therefore, it is most likely that rs5442 is the most causal SNV. Furthermore, we identified potential deleterious variants in known ocular disease genes in 6% of our study population. This is lower than we observed in a cohort of myopic children and adults admitted to a tertiary care hospital (yield was 15%), who displayed a more extreme phenotype with respect to treatment response and myopia progression.[52] Other studies that evaluated whole exome sequencing with a gene panel in individuals with high myopia reported a yield between 10% and 80%.[52–56] The discrepancy between our 2 studies may be caused by differences in the patient selection (lower mean age in the previous study), the exclusion of syndromic disorders beforehand, and the study design.

Among the most promising 8 novel genes associated with myopia are *HS6ST1, RBM20,* and *MAP7D1*. The *HS6ST1* gene encodes the heparan sulfate 6-*O*-sulfotransferase 1 protein, which plays a critical role in neuronal development.[57] This heparin transferase regulates neural branching and is associated with AD hypogonadotropic hypogonadism (MIM 614880). This gene is expressed in both rods and cones, and knockout mouse models displayed abnormal eye development and aberrant retinal ganglion layer morphology and influenced retinal axon guidance at the optic chiasm.[58,59] RNA binding motif protein 20 (*RMB20*) encodes the RNA-binding motif protein 20, which regulates the splicing of several genes involved in ion homeostasis, cardiomyopathy (MIM 613172), and sarcomere biology.[60] Among these regulated genes is titin (*TTN*), which was associated with myopia previously in both GWASs as whole exome sequencing studies with larger sample size.[7,9] *MAP7D1* encodes the microtubule associated protein 7 domain-containing protein 1 and plays a role in Wnt5 signaling and the dynamics of microtubules, which are major components of the

cytoskeleton.[61] The Wnt signaling pathway was implicated in myopia development previously in both a human GWAS and in experimental myopia.[7,62] Disruption of this pathway might lead to abnormal eye development and result in myopia.[62,63]

Pathway analyses highlighted cilium-dependent cell motility and microtubule-based movement as underlying processes. Both are related to the aforementioned Wnt signaling cascade[64] but also have a role in intracellular transport.[65] Microtubules are one of the filaments of the cytoskeleton, which gives the cell its overall shape, supports the plasma membrane, aids in the correct positioning of organelles, provides tracks for the transport of vesicles, and (in many cell types) allows the cell to move.[66] Motor proteins, for example, muscle fibers, associate with the cytoskeleton and are responsible for movement within the cell but also muscle contraction, depending on the type of filament binding (actin or microtubules).[67] Muscle fibers are surrounded by connective tissue, which plays a crucial role in the function and integrity of the skeletal muscle cells.[67] Interestingly, collagen transformation (which is a major component of connective tissue) and connective tissue disorders were also identified as underlying pathways. These pathways are interesting given the occurrence of (high) myopia in several collagen-related syndromic disorders (MIM 108300, 604841, and 267750). Furthermore, collagen is the major component of the sclera,[68] and altered remodeling of the human sclera might play a role in (exaggerated) axial elongation, which is a hallmark of myopia and leads to associated complications such as staphyloma.[69]

In addition to cilium-dependent cell motility, microtubule-based movement, and collagen transformation, pathway analyses identified melatonin degradation and neuronal signaling as underlying pathways. Unfortunately, we could not provide any direction of effect because gene expression data were unavailable. Additionally, it remains to be studied if missense variation has a dominant-negative or gain-of-function effect. Regardless, the significant enriched pathways provide valuable clues for important underlying mechanisms. Melatonin is a key neuromodulator of circadian rhythm and is related to light exposure, which are both implicated in eye growth.[70] Levels of melatonin are altered in myopes,[51,71] but the exact mechanisms underlying the potential association between melatonin and myopia development have yet to be elucidated.[72]

In complex genetic disorders and traits, the individuals carrying less burden of common risk variants (a lower GRS in case of myopia) are more likely to carry rare variants with larger effect size.[27] When comparing the number of carriers of any of the top rare burden genes, however, between individuals with low versus high common risk burden, we did not observe a significant difference, except for *ABCD1*. This can be attributed to a lack of power to detect a difference or unmeasured environmental or epigenetic effects in these groups. As a next step, we evaluated these rare variants in the total population. Unfortunately, we could not determine the effect size of these variants but dichotomized the carriership of any of
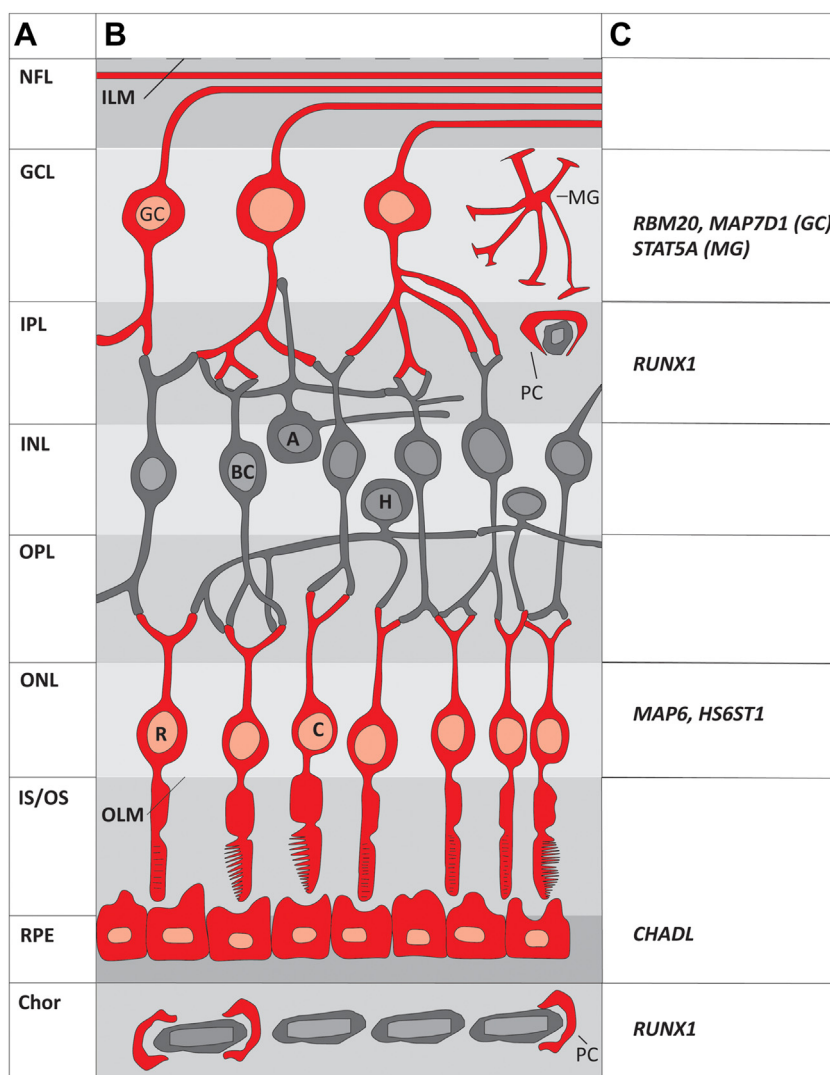
**Figure 5.** Expression of top genes identified through whole genome sequencing in our study population (N = 119) with low refractive error and low genetic risk score. Evidence for expression in ocular tissue of these genes was assessed using the recently published human ocular dataset from Cowan et al[31] and the online tool Spectacle, which includes data from 5 human datasets (all-retina-choroid-retinal pigment epithelium [RPE]) (32-37). The ocular cell type with highest expression is marked red. Panel A shows the different retinal layers. Panel B shows the presence of the top genes (in red), and Panel C shows the gene name. A = amacrine cell; BC = bipolar cells; C = cone photoreceptor cell; GC = ganglion cell; GCL = ganglion cell layer; H = horizontal cell; ILM = inner limiting membrane; INL = inner nuclear layer; IPL = inner plexiform layer; IS = inner segment; MG = microglia; NFL = optic nerve fiber layer; OLM = outer limiting membrane; ONL = outer nuclear layer; OPL = outer plexiform layer; OS = outer segment; PC = pericyte; R = rod photoreceptor cell.

the rare variants containing candidate genes instead and observed no different proportion of carriers in the individuals with less common variant risk (higher GRS). The RE in these groups was not significantly different. We also stratified groups based on the combination of common and rare variants to evaluate the effect on RE and observed a trend toward a more myopic RE in individuals carrying both many common variants, as well as rare variants. This difference was not significant, probably due to a lack of power. A trend could suggest that the identified genes harboring rare variants are not independent from common GRSs and that they act additively or even interactively. Future research, however,

is necessary to further explore this hypothesis. We must admit that the GRS filtering approach is not an ideal method to select individuals most susceptible for carrying rare variants. Multiple large studies focused on other diseases (e.g., breast cancer) have shown that rare variants and high GRSs can coexist and can contribute independently to disease risk.[73,74] Unfortunately, we did not have access to the raw data of a control cohort, which would indeed be preferred in a rare variant burden strategy. We used stringent filtering criteria based on GRS as a first step instead and later a more lenient filtering approach in which we evaluated the total group irrespective of GRS. Furthermore, we were able to

evaluate our main candidate genes in a reference database using similar sequencing technology and analysis procedures.

This is the first study that performed WGS on a large number of individuals with high myopia. This technique enabled us to review the molecular genetic anatomy of high myopia. Before we examined the TAD regions, we assessed the contribution of common variants identified through large scale GWASs by using GRS and used this as the first step of our filtering approach. The cutoff for high versus low GRS was chosen based on the distribution of GRSs in this population and might be affected by potential confounding factors such as cryptic relatedness or potential selection bias, leading to a distorted frequency of these variants in this population.

Other limitations of this study are the lack of an emmetropic control group with WGS, which could provide insights into the frequency of variants in a confirmed nonmyopic population useful for prioritization of variants and to evaluate the presence of differences in ancestry and population stratification. Alternatively, we compared the frequency of variants in our population to the populations included in GnomAD and used the number of homozygotes and heterozygotes as a proxy for carriership of any variant. Because some individuals might carry > 1 rare variant in a single gene, this approximation may be an overestimation in the control group, resulting in a more conservative test. Moreover, the RE status of the controls of GnomAD is unknown, but some individuals could be (highly) myopic because of the increasing prevalence of this condition in the general population. To correct for technical artifacts, ancestry, and population structure differences, we prioritized all variants using 704 genomes from unrelated controls run on the same BGI platform. Additionally, the current sample size was too small to detect rare variants with large effect genome-wide, and the 1% allele frequency cutoff might be too strict given the increasing prevalence of myopia.[75] Limitations in the processing and storage of large-scale WGS data and the high costs of including large numbers of patients and phenotypically evaluated controls to reach statistical power prevented the use of a more lenient filtering strategy, a formal genome-wide burden analysis, and a completely hypothesis-free methodology. We expect that variants in regulatory elements not evaluated in this study will explain a proportion of the high heritability of myopia.

To conclude, we identified rare variants in known ocular disease genes in some individuals of our study population. By screening of the entire genome filtered on highly deleterious variants, we identified 8 novel genes associated with myopia that are related to Wnt signaling, ocular development, and processing of melatonin. Whole genome sequencing has the potential to unravel more of the genetic architecture of high myopia, but more knowledge about the exact role and function of intergenic regions and sequencing of phenotypically evaluated control groups and larger groups of patients to increase power are necessary to make big steps.

## Acknowledgments

## Footnotes and Disclosures

[1] Erasmus MC, Department of Ophthalmology, Rotterdam, The Netherlands.

[2] Erasmus MC, Department of Epidemiology, Rotterdam, The Netherlands.

[3] Department of Ophthalmology, Radboud University Medical Center, Nijmegen, The Netherlands.

[4] Institute of Molecular and Clinical Ophthalmology, Basel, Switzerland.

[5] Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands.

[6] Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Nijmegen, The Netherlands.

[7] Erasmus MC, Department of Clinical Genetics, Rotterdam, The Netherlands.

*E.B. and V.J.M.V. contributed equally to this work.

# References

1. Flitcroft DI, He M, Jonas JB, et al. IMI - defining and classifying myopia: a proposed set of standards for clinical and epidemiologic studies. *Invest Ophthalmol Vis Sci*. 2019;60: M20−M30.
2. Fricke TR, Jong M, Naidoo KS, et al. Global prevalence of visual impairment associated with myopic macular degeneration and temporal trends from 2000 through 2050: systematic review, meta-analysis and modelling. *Br J Ophthalmol*. 2018;102:855−862.
3. Holden BA, Fricke TR, Wilson DA, et al. Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. *Ophthalmology*. 2016;123:1036−1042.
4. Haarman AEG, Enthoven CA, Tideman JWL, et al. The complications of myopia: a review and meta-analysis. *Invest Ophthalmol Vis Sci*. 2020;61:49.
5. Tideman JWL, Polling JR, Vingerling JR, et al. Axial length growth and the risk of developing myopia in European children. *Acta Ophthalmol*. 2018;96:301−309.
6. Tedja MS, Wojciechowski R, Hysi PG, et al. Genome-wide association meta-analysis highlights light-induced signaling as a driver for refractive error. *Nat Genet*. 2018;50:834−848.
7. Hysi PG, Choquet H, Khawaja AP, et al. Meta-analysis of 542,934 subjects of European ancestry identifies new genes and mechanisms predisposing to refractive error and myopia. *Nat Genet*. 2020;52:401−407.
8. Tedja MS, Haarman AEG, Meester-Smoor MA, et al. IMI - myopia genetics report. *Invest Ophthalmol Vis Sci*. 2019;60: M89−M105.
9. Guggenheim JA, Clark R, Cui J, et al. Whole exome sequence analysis in 51 624 participants identifies novel genes and variants associated with refractive error and myopia. *Hum Mol Genet*. 2022;31:1909−1919.
10. Liu Y, Zhang JJ, Piao SY, et al. Whole-exome sequencing in a cohort of high myopia patients in Northwest China. *Front Cell Dev Biol*. 2021;9:645501.
11. Musolf AM, Simpson CL, Alexander TA, et al. Genome-wide scans of myopia in Pennsylvania Amish families reveal significant linkage to 12q15, 8q21.3 and 5p15.33. *Hum Genet*. 2019;138:339−354.
12. Turro E, Astle WJ, Megy K, et al. Whole-genome sequencing of patients with rare diseases in a national health system. *Nature*. 2020;583:96−102.
13. de Bruijn SE, Fiorentino A, Ottaviani D, et al. Structural variants create new topological-associated domains and ectopic retinal enhancer-gene contact in dominant retinitis pigmentosa. *Am J Hum Genet*. 2020;107:802−814.
14. Chen J, Wu JS, Mize T, et al. A frameshift variant in the CHST9 gene identified by family-based whole genome sequencing is associated with schizophrenia in Chinese population. *Sci Rep*. 2019;9:12717.
15. Benonisdottir S, Oddsson A, Helgason A, et al. Epigenetic and genetic components of height regulation. *Nat Commun*. 2016;7:13490.
16. Steinthorsdottir V, Thorleifsson G, Sulem P, et al. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nat Genet*. 2014;46:294−298.
17. Tideman JW, Snabel MC, Tedja MS, et al. Association of axial length with risk of uncorrectable visual impairment for Europeans with myopia. *JAMA Ophthalmol*. 2016;134:1355−1363.
18. Ikram MA, Brusselle G, Ghanbari M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. *Eur J Epidemiol*. 2020;35:483−517.
19. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*. 2016;32:292−294.
20. Farek J, Hughes D, Salerno W, et al. xAtlas: scalable small variant calling across heterogeneous next-generation sequencing experiments. *GigaScience*. 2022;12:giac125.
21. Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016;32: 1220−1222.
22. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581:434−443.
23. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38:e164.
24. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc*. 2015;10:1556−1566.
25. Kircher M, Witten DM, Jain P, et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46:310−315.
26. Zhou D, Yu D, Scharf JM, et al. Contextualizing genetic risk score for disease screening and rare variant discovery. *Nat Commun*. 2021;12:4418.

27. Erasmus MC DCG. Whole Exome Sequencing Gene package Vision disorders, version 2-9. https://www.erasmusmc.nl/nl-nl/patientenzorg/laboratoriumspecialismen/klinische-genetica. In: Erasmus MC DCG, ed. 2019; v. 2021.

28. Amberger JS, Hamosh A. Searching Online Mendelian Inheritance in Man (OMIM): a knowledgebase of human genes and genetic phenotypes. *Curr Protoc Bioinformatics*. 2017;58: 1.2.1−1.2.12.

29. Online Mendelian Inheritance in Man, OMIM. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD).

30. McArthur E, Capra JA. Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. *Am J Hum Genet*. 2021;108:269−283.

31. Schmitt AD, Hu M, Jung I, et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep*. 2016;17:2042−2059.

32. Guo MH, Plummer L, Chan YM, et al. Burden testing of rare variants identified through exome sequencing via publicly available control data. *Am J Hum Genet*. 2018;103:522−534.

33. Chang SPaC. Package : PLINK [1.9]. 1.9 ed.

34. Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015;4:7.

35. Davydov EV, Goode DL, Sirota M, et al. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLOS Comput Biol*. 2010;6:e1001025.

36. Cowan CS, Renner M, De Gennaro M, et al. Cell types of the human retina and its organoids at single-cell resolution. *Cell*. 2020;182:1623−1640.e34.

37. Voigt AP, Whitmore SS, Lessing ND, et al. Spectacle: an interactive resource for ocular single-cell RNA sequencing data analysis. *Exp Eye Res*. 2020;200:108204.

38. Voigt AP, Whitmore SS, Flamme-Wiese MJ, et al. Molecular characterization of foveal versus peripheral human retina by single-cell RNA sequencing. *Exp Eye Res*. 2019;184: 234−242.

39. Voigt AP, Mulfaul K, Mullin NK, et al. Single-cell transcriptomics of the human retinal pigment epithelium and choroid in health and macular degeneration. *Proc Natl Acad Sci U S A*. 2019;116:24100−24107.

40. Voigt AP, Binkley E, Flamme-Wiese MJ, et al. Single-cell RNA sequencing in human retinal degeneration reveals distinct glial cell populations. *Cells*. 2020;9:438.

41. Voigt AP, Whitmore SS, Mulfaul K, et al. Bulk and single-cell gene expression analyses reveal aging human choriocapillaris has pro-inflammatory phenotype. *Microvasc Res*. 2020;131: 104031.

42. Voigt AP, Mullin NK, Whitmore SS, et al. Human photoreceptor cells from different macular subregions have distinct transcriptional profiles. *Hum Mol Genet*. 2021;30:1543−1558.

43. Bult CJ, Blake JA, Smith CL, et al. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res*. 2019;47:D801−D806.

44. Dickinson ME, Flenniken AM, Ji X, et al. High-throughput discovery of novel developmental phenotypes. *Nature*. 2016;537:508−514.

45. Groza T, Gomez FL, Mashhadi HH, et al. International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Res*. 2023;51:D1038−D1045.

46. Ruzicka L, Howe DG, Ramachandran S, et al. The Zebrafish Information Network: new support for non-coding genes,

47. richer Gene Ontology annotations and the Alliance of Genome Resources. *Nucleic Acids Res*. 2019;47:D867−D873.

47. Mi H, Ebert D, Muruganujan A, et al. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res*. 2021;49:D394−D403.

48. Mi H, Thomas P. PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. *Methods Mol Biol*. 2009;563:123−140.

49. Krämer A, Green J, Pollard Jr J, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*. 2014;30:523−530.

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

51. Chakraborty R, Micic G, Thorley L, et al. Myopia, or near-sightedness, is associated with delayed melatonin circadian timing and lower melatonin output in young adult humans. *Sleep*. 2021;44:zsaa208.

52. Haarman AEG, Thiadens AAHJ, van Tienhoven M, et al. Whole exome sequencing of known eye genes reveals genetic causes for high myopia. *Hum Mol Genet*. 2022;31:3290−3298.

53. Liu F, Wang J, Xing Y, Li T. Mutation screening of 17 candidate genes in a cohort of 67 probands with early-onset high myopia. *Ophthalmic Physiol Opt*. 2020;40:271−280.

54. Wan L, Deng B, Wu Z, Chen X. Exome sequencing study of 20 patients with high myopia. *PeerJ*. 2018;6:e5552.

55. Sun W, Huang L, Xu Y, et al. Exome sequencing on 298 probands with early-onset high myopia: approximately one-fourth show potential pathogenic mutations in RetNet genes. *Invest Ophthalmol Vis Sci*. 2015;56:8365−8372.

56. Zhou L, Xiao X, Li S, et al. Frequent mutations of RetNet genes in eoHM: further confirmation in 325 probands and comparison with late-onset high myopia based on exome sequencing. *Exp Eye Res*. 2018;171:76−91.

57. Tornberg J, Sykiotis GP, Keefe K, et al. Heparan sulfate 6-O-sulfotransferase 1, a gene involved in extracellular sugar modifications, is mutated in patients with idiopathic hypogonadotrophic hypogonadism. *Proc Natl Acad Sci U S A*. 2011;108:11524−11529.

58. Habuchi H, Nagai N, Sugaya N, et al. Mice deficient in heparan sulfate 6-O-sulfotransferase-1 exhibit defective heparan sulfate biosynthesis, abnormal placentation, and late embryonic lethality. *J Biol Chem*. 2007;282:15578−15588.

59. Pratt T, Conway CD, Tian NM, et al. Heparan sulphation patterns generated by specific heparan sulfotransferase enzymes direct distinct aspects of retinal axon guidance at the optic chiasm. *J Neurosci*. 2006;26:6911−6923.

60. Guo W, Schafer S, Greaser ML, et al. RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing. *Nat Med*. 2012;18:766−773.

61. Kikuchi K, Nakamura A, Arata M, et al. Map7/7D1 and Dvl form a feedback loop that facilitates microtubule remodeling and Wnt5a signaling. *EMBO Rep*. 2018;19:e45471.

62. Ma M, Zhang Z, Du E, et al. Wnt signaling in form deprivation myopia of the mice retina. *PLOS ONE*. 2014;9:e91086.

63. Liu Z, Xiu Y, Qiu F, et al. Canonical Wnt signaling drives myopia development and can be pharmacologically modulated. *Invest Ophthalmol Vis Sci*. 2021;62:21.

64. May-Simera HL, Cilia Kelley MW. Wnt signaling, and the cytoskeleton. *Cilia*. 2012;1:7.

65. Barlan K, Gelfand VI. Microtubule-based transport and the distribution, tethering, and organization of organelles. *Cold Spring Harb Perspect Biol*. 2017;9(5).

66. Cooper GM. *The Cell: A Molecular Approach*. 2nd ed. Sunderland, MA: Sinauer Associates; 2000.

67. Alberts B, Alexander J, Lewis J, et al. *Molecular Biology of the Cell*. 4th ed. New York, NY: Garland Science; 2002.

68. McBrien NA, Cornell LM, Gentle A. Structural and ultrastructural changes to the sclera in a mammalian model of high myopia. *Invest Ophthalmol Vis Sci*. 2001;42:2179−2187.

69. Yang Y, Li X, Yan N, et al. Myopia: a collagen disease? *Med Hypotheses*. 2009;73:485−487.

70. Chakraborty R, Ostrin LA, Nickla DL, et al. Circadian rhythms, refractive development, and myopia. *Ophthalmic Physiol Opt*. 2018;38:217−245.

71. Kearney S, O'Donoghue L, Pourshahidi LK, et al. Myopes have significantly higher serum melatonin concentrations than non-myopes. *Ophthalmic Physiol Opt*. 2017;37:557−567.

72. Ostrin LA. Ocular and systemic melatonin and the influence of light exposure. *Clin Exp Optom*. 2019;102:99−108.

73. Fahed AC, Wang M, Homburger JR, et al. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat Commun*. 2020;11:3635.

74. Mars N, Widén E, Kerminen S, et al. The role of polygenic risk and susceptibility genes in breast cancer over the course of life. *Nat Commun*. 2020;11:6383.

75. Lin DY. A simple and accurate method to determine genomewide significance for association tests in sequencing studies. *Genet Epidemiol*. 2019;43:365−372.