Exome Sequence Analysis of 14 Families With High Myopia

Bethany A. Kloss,1 Stuart W. Tompson,1 Kristina N. Whisenhunt,1 Krystina L. Quow,2 Samuel J. Huang,1 Derek M. Pavelec,3 Thomas Rosenberg,4,5 and Terri L. Young1

1Department of Ophthalmology and Visual Sciences, University of Wisconsin, Madison, Wisconsin, United States
2Center for Human Genetics, Duke University Medical Center, Durham, North Carolina, United States
3Biotechnology Center, University of Wisconsin, Madison, Wisconsin, United States
4The National Eye Clinic, Rigshospitalet, Kennedy Center, Glostrup, Denmark
5Institute of Clinical Medicine, University of Copenhagen, Denmark

Correspondence: Terri L. Young, Department of Ophthalmology and Visual Sciences, University of Wisconsin, 2828 Marshall Court, Suite 200, Madison, WI 53705, USA; tyoung6@wisc.edu.
Submitted: October 6, 2016
Accepted: February 21, 2017

PURPOSE. To identify causal gene mutations in 14 families with autosomal dominant (AD) high myopia using exome sequencing.

METHODS. Select individuals from 14 large Caucasian families with high myopia were exome sequenced. Gene variants were filtered to identify potential pathogenic changes. Sanger sequencing was used to confirm variants in original DNA, and to test for disease cosegregation in additional family members. Candidate genes and chromosomal loci previously associated with myopic refractive error and its endophenotypes were comprehensively screened.

RESULTS. In 14 high myopia families, we identified 73 rare and 31 novel gene variants as candidates for pathogenicity. In seven of these families, two of the novel and eight of the rare variants were within known myopia loci. A total of 104 heterozygous nonsynonymous rare variants in 104 genes were identified in 10 out of 14 probands. Each variant cosegregated with affection status. No rare variants were identified in genes known to cause myopia or in genes closest to published genome-wide association study association signals for refractive error or its endophenotypes.

CONCLUSIONS. Whole exome sequencing was performed to determine gene variants implicated in the pathogenesis of AD high myopia. This study provides new genes for consideration in the pathogenesis of high myopia, and may aid in the development of genetic profiling of those at greatest risk for attendant ocular morbidities of this disorder.

Keywords: high myopia, candidate genes, variants, exome sequencing, GWAS

Myopic refractive error, also known as short-sightedness, is the most common eye disorder worldwide. The extreme form, high myopia, is defined as a refractive error of −6.00 diopters (D) or higher. High myopia affects 2% to 5% of Americans, Western Europeans, and Australians aged older than 40 years. In East Asian countries, the prevalence of high myopia has been reported as low as 1.8% to 5% in Chinese adults aged older than 30 years, while higher rates are seen in Singaporean children aged 7 to 9 years (14.6%–17.9%), and university students in Taiwan (up to 38%).

High myopia is characterized by axial elongation and thinning of the sclera, with predisposition to additional ocular morbidities such as retinal detachment, glaucoma, and premature cataracts. Recently, exome sequencing of large families with high myopia has identified mutations in the genes LEPRE1/P3H2 and SCO2. The disease caused by mutations in LEPRE1/P3H2 followed an autosomal recessive inheritance pattern. Genome-wide association studies (GWAS) have identified over 60 candidate genes associated with refractive error. Familial linkage studies have also mapped 23 myopia (MYP) loci associated with high myopia. Significant linkage was first reported in eight families with autosomal dominant (AD) high myopia on chromosome 18p11.3. Since then, additional loci have been mapped on several chromosones (Table 1). Known myopia genes, myopia-associated genes identified by GWAS, and MYP chromosomal regions have been extensively reviewed.

Recently, Holden et al. estimated an increase in the prevalence of high myopia (greater than −5.00 D) from 163 million people (2.7% global population) in 2000 to 938 million people (9.8% global population) by 2050. This estimation utilized published data from 145 studies since 1995, covering 2.1 million participants. Vitale et al. also noted an 8-fold increase in high myopia (greater than −7.9 D) in the United States over a 50-year period from 1971 to 1972 to 1999 to 2004. Increased prevalence of myopia was observed regardless of age, race, sex, severity of the phenotype, or years of formal education. While the prevalence of high myopia may be increasing worldwide, the condition is still relatively uncommon with many unanswered questions. Increased availability of formal education and decreased exposure to rural or outdoor environments, compared to earlier decades, may partially account for an earlier onset of a myopic phenotype when combined with genetic susceptibility. Genetic analysis of high myopia within families such as those presented in this study may provide further insight into the intersecting contributions of biology with environment.

In this study, we sought to identify causal gene mutations in our high myopia cohort by examining known myopia genes, susceptibility loci identified through published GWAS data and chromosomal regions within the reported MYP loci coordi-
nates. Selected members from 14 nonsyndromic high myopia families underwent exome sequencing. As each family pedigree appeared to demonstrate AD inheritance transmission, it was hypothesized that each family would carry a single heterozygous private/rare mutation that caused the disease. We also employed a filtering strategy to identify additional rare and conserved variants, thereby expanding the list of candidate genes associated with high myopia. Determining causative genetic and therefore biologic factors for high myopia development will inform more effective strategies to reduce or minimize the likelihood of developing blindness due to associated ocular risks, and improve quality of life for those affected.

**Materials and Methods**

**Human Subjects**

We recruited 14 families with high myopia when DNA could be obtained from at least three living affected relatives (Supplemental Figs. S1–S14). Informed consent was obtained from all participants, with approval by the Institutional Review Board at the University of Wisconsin (Madison, WI, USA) according to the tenets of the Declaration of Helsinki. Spherical equivalent (SE) refractive error measurements were conducted and DNA was extracted from blood and/or saliva from participating family members. In our study, 8 of the 14 families were from Denmark, 4 families were from the United States, 1 family was of English/Canadian descent, and 1 family was of Hawaiian/Chinese descent. The families ranged in size from three individuals with high myopia up to the largest family (family 56) with 14 affected individuals. Ethnicity, SE refractive error ranges, and affection status information for each family is available in Table 2.

The affected high myopia phenotype was determined to be those with SE refractive error greater than −6.00 D. Unaffected persons were defined by having SE refractive error less than −2.00 D. Those refractive errors reported by hearsay are indicated in gray in the pedigree figures. They represent primarily deceased and older generations who were not genotyped in this study: family 8, II:1, II:3, family 56, II:1, II:3, III:13, family 104, I:1, family 126 I:1, II:2, IV:7, and family 130, I:1.

Refractive errors were self-reported for the following patients: family 104, individual I:1 (high myopia in both eyes, cataract removed without intraocular lens placement in the left eye to mitigate myopia); and family 111, individuals I:2, II:1, II:2 (refractive errors obtained prior to LASIK corneal surgery to mitigate their myopia). These individuals were genotyped in this study.

To the best of our knowledge, none of the subjects used in this study presented with systemic disorders associated with myopia, including Marfan syndrome, Ehlers-Danlos syndrome, Stickler syndrome, or Wagner syndrome. Affected subjects had no other known ocular disorders or abnormalities, apart from family 101 where some affected family members developed glaucoma later in life. Individuals with SE refractive errors between −2.00 D and −6.00 D were excluded from the study, as they could not be considered neither affected nor unaffected. We sought to perform an extreme phenotype genotyping screening approach for this quantitative trait.

**Exome Sequencing**

Exome sequencing of pedigrees selected for this study were performed at three facilities. Samples sequenced by the Center for Human Genome Diversity (Duke University, Durham, NC, USA) used a commercial capture kit (SeqCap EZ Exome Capture Kit v2 and v5 and 2 × 100 bp, paired-end; Roche NimbleGen, Inc., Madison, WI, USA) reads on a sequencing platform (HiSeq2000; Illumina, Inc., San Diego, CA, USA). Samples sequenced by the Hudson Alpha Institute for Biotechnology (Huntsville, AL, USA) used a commercial capture kit (SeqCap EZ Exome Capture Kit v2 and v5 and 2 × 100 bp, paired-end; Roche NimbleGen, Inc., Madison, WI, USA) reads on a sequencing platform (HiSeq2000; Illumina, Inc., San Diego, CA, USA). Samples sequenced by the University of Wisconsin Biotechnology Center DNA Sequencing Facility (Madison, WI, USA) used a commercial capture kit (SureSelect Exome v5 and 2 × 100 bp, paired-end; Agilent Technologies, Inc., Santa Clara, CA, USA).
reads on a sequencing platform (Illumina, Inc.). University of Wisconsin Biotechnology Center analysis of DNA sequencing data was performed as follows: Adapter sequences and low-quality sequencing reads were trimmed using Skewer. We used BWA-maximal exact matches (MEM) to align all reads to the reference genome, GRCh37/hg19 assembly. Duplicate alignments were then removed using Picard (http://picard.sourceforge.net, in the public domain). Finally, variants were called using GATK HaplotypeCaller version 3.3 and annotated with SNPeff.

**Variant Analysis**

Single nucleotide polymorphisms (SNPs) and insertion/deletion (indel) changes were filtered and analyzed using SNP and Variation Suite Software v8.3 (Golden Helix, Bozeman, Montana). Variant filtering included the following steps (Fig.):

1. Excluded variants outside of exonic and splice-site regions (within 2 bp; RefSeq Genes 105v2, NCBI);
2. Excluded synonymous variants;

![Figure](https://example.com/figure.png)

**Number of Variants**

- Total variants: 389,326
- Coding variants (within 2bp of splice sites): 164,815
- Nonsynonymous variants: 74,585
- Global minor allele frequency (MAF) less than 0.001: 14,095
- Evolutionarily conserved: 8857
- Heterozygous; present in all affected individuals within family: 926
- Not present in unaffected individuals, confirmed in BAM file data, MAF < 0.001 in any ethnicity: 104
- Novel variants + rare variants in MYP loci: 39
- Sanger sequence confirmed: co-segregation where possible: 39
- Insertions/deletions: 2625
- Other rare variants: 65
3. Excluded variants with global minor allele frequencies (MAF) greater than 0.001 in the Exome Aggregation Consortium (ExAC) database (ExAC Variant Frequencies 0.3 BROAD)47.
4. With the exception of indels, excluded variants with negative conservation level across vertebrate species (PhyloP 100-way vertebrate score less than zero)48.
5. Excluded variants that were not heterozygous in all affected individuals within each family.
6. Excluded variants present in unaffected individuals within the corresponding family, confirmed variant present in raw sequence alignment data, and excluded variants observed in the ExAC database with a MAF greater than 0.001 in any ethnicity.

Additionally, a filtering track was created for use in the Golden Helix software to retrieve all gene variants contained within the 25 MYP chromosomal regions (discussed in OMIM 160700).49 Confirmation of variants that were novel or located within MYP loci was performed by Sanger sequencing. Cosyregen analysis was also performed when additional family members were available for screening. Primers used for screening SCO2 were as follows: 5’-GCTTGTTTCCAGGAGCAT-3’ and 5’-TGCGCTCAAGACAGGACACT-3’.

## Results

After filtering of the exome data, a total of 104 heterozygous variants in 104 genes were identified in 10 families with nonsyndromic high myopia. Variants meeting our filtering criteria were not identified in four families. Each variant identified was located within the coding region or splice site of the gene, reported at a frequency of less than 1 in 1000 alleles in any population and present in all relevant individuals with high myopia in each family that were available for screening. None of the variants were identified in unaffected family members (SE refractive error of −2.00 D or less severe) or individuals that married into the family. Ten variants were identified within myopia-associated (MYP) chromosomal regions (Table 3). We found 31 of the 104 variants identified were novel changes unreported in genetic population databases (Table 4). Two variants were both novel and located within MYP loci. In our cohort, 73 variants identified were uncommon (less than 1 in 1000 alleles in any population; Supplementary Table S2). Confirmation of the 31 novel variants and 8 rare variants located within MYP loci was achieved by Sanger sequencing.

### Discussion

In this study, exome sequencing was utilized to identify gene variants in individuals affected with nonsyndromic high myopia. The association of high myopia and specific genes, chromosomal loci, and SNPs has been previously investigated and reported in several studies. For this analysis, we sought to examine these associations within our own high myopia cohort. We did not observe more than one variant in the same gene in more than one family and hypothesized that each family harbors its own private pathogenic mutation resulting in high myopia. Since high myopia is a rare disease (2%–5% in the global population), we hypothesized that a causal variant that segregates with high myopia should also be rare. We also examined variants that were novel or observed in less than 1 in 1000 alleles in ~61,000 individuals represented in the ExAC database. We identified potentially pathogenic heterozygous
variants in 10 families with high myopia that followed the autosomal dominant inheritance pattern. A premature stop codon variant (p.Gln53*) in SCO2 was previously reported in a family with autosomal dominant high myopia.10 The gene SCO2 encodes a protein involved in mitochondrial cytochrome c oxidase activity and maps to chromosome 22q13.33, which is located within the MYP6 locus (OMIM 608908).49 A link between cytochrome c oxidase deficiency and refractive error has not been demonstrated. Tran-Viet et al.10 also reported three additional variants in 10 families with high myopia follow the autosomal dominant inheritance pattern. Affected individuals in each family showed dominantly inherited disease, and affected individuals in one family were identified through GWAS.12,13 We examined the gene nearest to each associated locus (Table S1). However, no potentially pathogenic variants were identified.

To date, more than 60 common SNPs implicating loci for myopia/refractive error have been identified through GWAS.12,13 We examined the gene nearest to each associated SNP for variants in our high myopia cohort (Supplementary Table S1). However, no potentially pathogenic variants were identified within these genes in our exome data. While over 6000 significant SNP-trait associations have been reported through GWAS,51 these studies were designed to detect associations between common variants in the population and common diseases. Several ophthalmic GWAS have uncovered loci that may help identify molecular pathways associated with ocular diseases such as age-related macular degeneration and glaucoma.52 However, for the study of less common diseases such as high myopia, it is possible that many loci still remain unidentified.53,54 Furthermore, approximately 80% of trait-associated SNPs reported are located in intergenic or noncoding regions,55 raising additional questions about our current understanding of gene regulation in disease pheno-
types. Another consideration is that the age at which GWAS-identified variants may contribute to myopia in the global population is not known. Fan et al. examined the age-of-onset correlation between refractive error and GWAS-identified variants in 5200 children examined longitudinally from ages 7 to 15 years, and found many significant SNPs and their effects during very early life, childhood, or adulthood. Some SNPs showed progressively stronger effects during later childhood while others appeared to have no progressive effect on refractive error. This study also examined the effects of gene-environment interactions involving near work or time spent outdoors, but these associations were rare or absent for most of the GWAS-identified SNPs.

To investigate the variation in our cohort within loci linked to refractive error, the 23 MYP loci were examined. In our exome data, 10 variants were identified within these coordinates, two of which are novel (Table 3). Three variants, in genes PTPRO, TMPO, and TMT2C, were identified at the MYP5 locus, which was mapped to chromosome 12q21-q23 in a large German/Italian family segregating autosomal dominant high myopia. Variants in PTPRO and TMT2 were identified in family 115, and the variant in TMPO was identified in family 111. PTPRO encodes a protein-tyrosine phosphatase implicated in autosomal recessive deafness. A heterozygous missense change in Thymopoeitin (TMPO) was identified in two brothers with a severe form of cardiomyopathy. The variant identified in TMPO in this study was novel. In 2015, a meta-analysis of genome-wide association performed by the International Glaucoma Genetics Consortium identified a SNP near the gene transmembrane and tetratricopeptide repeat containing 2 (TMT2C) associated with optic disc morphology and glaucoma, which may indicate a possible role for TMT2C in the eye.

In family 104, one variant in the gene ETNPPL (alanine-glyoxylate aminotransferase 2-like 1 [AGXT2LI]) was identified in MYP11, which was mapped to chromosome 4 in a large Chinese family with autosomal dominant high myopia. Two variants were identified at MYP14 (chromosome 1p26) which was mapped using 49 multigenerational Ashkenazi Jewish families with high myopia. These two variants are nonsynonymous changes in the genes AGRN and PLOD1 identified in families 130 and 122, respectively. Agrin (AGRN) plays a role in synapse formation and brain development. A homozygous mutation in AGRN was reported in a Swiss brother and sister with congenital myasthenia syndrome (CMS8) which affects skeletal muscle and neuromuscular junctions, while compound heterozygosity for two mutations in AGRN was identified in another individual with CMS8. Homozygous mutations in the collagen-modifying gene PLOD1 have been implicated in Ehlers-Danlos syndrome, a heritable connective tissue disorder that includes several ocular manifestations including myopia (OMIM 135443). Common SNPs in PLOD1 were also studied in a cohort of 600 unrelated Chinese Han individuals (300 with high myopia), but were determined unlikely to result in genetic susceptibility for high myopia.

Originally, MYP4 was mapped to chromosome 7q36 after linkage analysis was performed in 21 French and 2 Algerian families with autosomal dominant high myopia. However, a second study using the same families along with nine newly collected families failed to confirm linkage to 7q36, but rather to 7p15 and the MYP4 locus was renamed MYP17. A study performed by Ciner et al. in the same year also identified linkage to chromosome 7p15 after performing quantitative trait locus linkage analyses in 493 African American individuals and in 90 African American families. In our study, nonsynonymous variants were identified at the MYP17/4 locus in genes DNAH11 and HOXA2. Mutations in the dynactin gene DNAH11 have been reported in Kartagener syndrome, a type of ciliary dyskinesia in which patients suffer from chronic respiratory infections, recurrent bronchiitis, and pneumonia. Homeobox A2 (HOXA2) mutations have been implicated in autosomal dominant and recessive forms of microtia and hearing impairment, sometimes accompanying cleft palate. In this study, the variant identified in HOXA2 was novel.

Two variants in the genes essential meiotic structure-specific endonuclease 1 (EME1) and transducer of ERBB2 1 (TOB1) were identified at the MYP5 locus, which was mapped to chromosome 17q21-q22 in a multigenerational English/Canadian family with autosomal dominant high myopia. However, no causal mutation was reported in that publication. The same family was utilized for this study (family 68), and notably the proband of family 68 had the most severe SE refractive error (~50.00/-50.00 D) of our cohort. Exome data from three affected and one unaffected individuals were filtered for potentially pathogenic variants. Two variants were identified; one variant identified in the gene Ftsj Homolog 3 (FTSJ3) encodes a missense change, while a variant identified in the gene Solute Carrier Family 7 Member 13 (SLC7A13) results in a premature stop codon in the third of four exons (Supplementary Table S2). Neither of these variants were located within the MYP5 locus previously linked to this family or within any other MYP loci.

A second large family in our cohort has also undergone prior linkage analysis. Family 56, a Caucasian family from the United States, is a large autosomal dominant high myopia family in which nearly half of the individuals are highly myopic. A novel locus at chromosome 2q37.1 (MYP12) was determined, but causal mutations were not identified. To identify candidate causative variants in this family, we performed exome sequencing on 11 individuals, including 8 that were affected. However, no variants were identified that were present in all 8 affected individuals sequenced. It is interesting to note that within both families 56 and 68, no variants that met our filtering criteria were identified within the chromosomal regions where linkage studies previously mapped myopia-associated loci. While genetic linkage analysis is a powerful tool for mapping disease loci in Mendelian disorders in families, there are limitations. Since proper linkage requires a large family with multiple generations demonstrating the phenotype, most of our families were too small to evaluate using this method. Genetic linkage studies are less helpful for multifactorial or complex traits, where multiple genes are important in disease causation. However, given its severity in young children and infants, and the clearly Mendelian inheritance pattern, there are likely to be strong genetic determinants in high myopia. Nevertheless, myopia is a highly variable disease, with refractive measurements varying greatly between affected individuals in the same family as well as between eyes of the same individual. Numerous studies have examined the correlation between environmental factors and myopia, including levels of education, increased near work, rural versus urban settings, diet, and outdoor activities. A combination of genetic susceptibility and environmental triggers may play a role together in disease manifestation and progression.

Notably for this study, potentially pathogenic variants cosegregating with high myopia were only identified in 10 of 14 families in our cohort. No variants meeting our filtering criteria were identified in families 8, 19, 56, 126. In addition to the added complications given the variability in myopia severity, there are also limitations using exome sequencing as a method to detect causal variants. It was previously estimated that approximately 85% of disease-causing mutations are found in the coding regions of the genome, but our understanding...
of the noncoding region’s importance has recently rapidly evolved. The protein-coding exome comprises approximately 2% of the human genome, and as functional consequences can occur in response to sequence variation outside of coding regions, whole-genome sequencing may resolve the underlying molecular etiology. Furthermore, exome sequencing is unable to reliably detect large insertions, deletions, or rearrangements, and errors in calling small indel variants are common. Exome capture kits offered by companies such as Agilent or Roche NimbleGen are dependent on our knowledge of the expressed regions of the human genome, and therefore exclude variants in regions of the genome that are not currently known to be transcribed. Exome coverage is also incomplete, with less than 80% of the exome captured at more than \( \times 20 \) coverage with the Agilent platform (SureSelect) that was utilized for the most recent exome sequencing performed in this study,\(^7\)\(^3\)\(^4\) though some groups have reported up to 95% coverage of RefSeq coding exons in their exome data.\(^7\)\(^2\)\(^5\) Another potential source of error is the possibility of healthy control populations having unreported high myopia. Many do not consider high myopia a disease and may underestimate their roles as controls in genetic studies. It is therefore difficult to be certain that genomic databases are free of highly myopic individuals when making conclusions in our studies.

While whole genome sequencing remains cost prohibitive for many research groups, it offers advantages over exome sequencing when the pathogenic mutation is suspected to be located in a noncoding region. Variations in upstream elements such as promoters, enhancers, and silencers may be assessed for potential regulatory impact. Noncoding regions that show high conservation among different species as well as predicted transcription factor binding sites can be analyzed for variations. Sequence capture kits are also expected to offer expanded targets in the future which may help to identify variations in noncoding DNA. Advancements in complex prediction algorithms may also aid in the assessment of potentially pathogenic gene variants.

In conclusion, we have identified 104 new candidate variants for high myopia. These variants will require future study in additional patient subjects with high myopia, functional analyses, and/or animal modeling in order to determine their role in this phenotype.

Acknowledgments

The authors thank all patients, subjects, and their families for their participation in this study.

Supported by the National Institutes of Health (NIH), National Eye Institute Grants R01 EY14685 and R01EY018246; NIH National Eye Institute P30 EY016665 CORE Grant for vision research; an unrestricted grant from Research to Prevent Blindness, Inc., to the Department of Ophthalmology and Vision Sciences at the University of Wisconsin; and a Centennial Scholars Grant from the University of Wisconsin School of Medicine and Public Health.

Disclosure: B.A. Kloss, None; S.W. Tompson, None; K.N. Whisenhunt, None; K.L. Quow, None; S.J. Huang, None; D.M. Pavelc, None; T. Rosenberg, None; T.L. Young, None.

References


45. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a mapReduce framework for analyzing next-genera


63. Yip SP, Leung KH, Fung WY, Ng PW, Sham PC, Yap MK. A DNA pooling-based case-control study of myopia candidate genes.


