

Somatic Variants in the Human Lens Epithelium: A Preliminary Assessment

Rosana Mesa,¹ Manoj Tyagi,^{*,2} George Harocopos,^{1,3} David Vollman,¹ and Steven Bassnett¹

¹Department of Ophthalmology and Visual Sciences, Washington University School of Medicine, St. Louis, Missouri, United States

²Genome Technology Access Center, Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, United States

³Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri, United States

Correspondence: Steven Bassnett, Department of Ophthalmology and Visual Sciences, Washington University School of Medicine, 660 S. Euclid Ave, Box 8096, St. Louis, MO 63110, USA; bassnett@vision.wustl.edu.

Current affiliation: *National Cancer Institute, Bethesda, Maryland, United States.

Submitted: April 11, 2016

Accepted: June 9, 2016

Citation: Mesa R, Tyagi M, Harocopos G, Vollman D, Bassnett S. Somatic variants in the human lens epithelium: a preliminary assessment. *Invest Ophthalmol Vis Sci.* 2016;57:4063-4075. DOI:10.1167/iovs.16-19726

PURPOSE. We hypothesize that somatic mutations accumulate in cells of the human lens and may contribute to the development of cortical or posterior sub-capsular cataracts. Here, we used a Next-generation sequencing (NGS) strategy to screen for low-allelic frequency variants in DNA extracted from human lens epithelial samples.

METHODS. Next-Generation sequencing of 151 cancer-related genes (WUCaMP2 panel) was performed on DNA extracted from post-mortem or surgical specimens obtained from 24 individuals. Usually, pairwise comparisons were made between two or more ocular samples from the same individual, allowing putative somatic variants detected in lens samples to be differentiated from germline variants.

RESULTS. Use of a targeted hybridization approach enabled high sequence coverage (>1000-fold) of the WUCaMP2 genes. In addition to high-frequency variants (corresponding to homozygous or heterozygous SNPs and Indels), somatic variants with allelic frequencies of 1-4% were detected in the lens epithelial samples. The presence of one such variant, a T > C point substitution at position 32907082 in *BRCA2*, was verified subsequently using droplet digital PCR.

CONCLUSIONS. Low-allelic fraction variants are present in the human lens epithelium, at frequencies consistent with the presence of millimeter-sized clones.

Keywords: somatic mutation, cataract risk factors, genomics

Clouding of the ocular lens, or cataract, is the leading cause of blindness globally, affecting tens of millions of people.¹ Cataracts are classified according to where in the lens opacities are located. Nuclear cataracts, for example, involve the center of the lens, whereas cortical cataracts are located in the periphery. Posterior subcapsular cataracts (PSCs), affect a layer of cells immediately beneath the posterior lens pole. Although loss of tissue transparency is the common endpoint, each type of cataract affects a particular region of the lens, is associated with a unique spectrum of risk factors, and probably reflects a distinct etiology.²

Work from many laboratories supports the notion that the progressive aggregation of aged and extensively modified lens proteins is the proximate cause of nuclear cataracts (reviewed in Ref. 3). The etiology of cortical cataracts and PSCs is less well established, even though, collectively, these subtypes can account for most cataracts in some populations.⁴ Several observations support the idea that the accumulation of somatic mutations in the nucleated cells of the lens could play a role in the development of cortical cataracts and PSCs, at least under certain conditions. The strongest evidence comes from longitudinal studies of human cohorts exposed to ionizing radiation. The lens is among the most radiation-sensitive tissues in the body, second perhaps only to the germ cells.⁵ In astronauts,⁶ Chernobyl cleanup workers,⁷ radiologic technicians,⁸ and patients exposed to radiotherapy,⁹ studies have consistently identified an association between radiation expo-

sure and latent development of cataracts. It should be noted, however, that it is notoriously difficult to collect accurate dosimetry data in long-term retrospective studies,¹⁰ and this caveat should be borne in mind when considering the conclusions of such studies. One cohort that has been carefully followed for many decades consists of survivors from the atomic bomb attacks on Hiroshima and Nagasaki. Careful analysis of cataract in atomic bomb survivors has shown that both the prevalence and incidence of cataracts (generally cortical cataracts and PSCs) that required surgical removal are increased at doses of 0.5 Gy or lower.^{11,12} Those who were young at the time of exposure appear to have been at particular risk. These and other findings have led the International Commission on Radiological Protection to recently recommend lowering the exposure threshold for vision-impairing cataract from 5 Gy to 0.5 Gy.⁵

A second line of evidence comes from epidemiological studies that have examined the link between cortical cataract formation and lifetime UV-B exposure. Ultraviolet-B radiation is a potent environmental mutagen of particular relevance to the eye. Due to their relatively high energy, UV-B photons may be absorbed directly by double bonds in pyrimidine bases (thymine and cytosine) promoting the formation of cyclobutane pyrimidine dimers (CPDs) and 6-4 photoproducts in the DNA; two helix-distorting photo-lesions. Photo-lesions are usually repaired quickly through the nucleotide excision repair pathway, but unrepaired adducts can cause mispairing during



DNA replication, leading to the formation of classic UV signature mutations C > T and CC > TT.¹³ In animal models, UV-B (at levels equivalent to ambient exposure) can penetrate the cornea, leading to the production of CPDs and 6-4 photoproducts in the DNA of anterior lens epithelial cells.¹⁴ In humans, most epidemiological studies examining risk factors for cataract have concluded that exposure to UV radiation (and, in particular, UV-B) is associated with an increased risk of cortical cataract.¹⁵ Based on such data, the World Health Organization estimates that, globally, 20% of cataracts may be attributed to UV exposure. It is important to acknowledge, however, that many of the detrimental effects of UV exposure (particularly acute exposure) to the lens involve nonmutagenic mechanisms (such as type I photo-oxidation reactions¹⁶).

If somatic mutations indeed play a contributory role in cataract formation, one might expect that defects in DNA repair pathways, which typically result in increased mutational frequency,¹⁷ would be associated with an elevated risk of cataract. In keeping with this notion, cataract is a common feature of several syndromic conditions caused by mutations in DNA-repair genes, including Cockayne syndrome, Werner syndrome, Rothmund-Thompson syndrome, and Trichothiodystrophy.¹⁸ Similarly, polymorphisms in *XRCC1* (base excision repair pathway) and *XPD* (nucleotide excision repair pathway) are associated with the development of age-related cataract.¹⁹

Somatic variants, should they occur in the human lens epithelium, are likely to be present at low frequencies, making it technically challenging to detect them against the large background signal from the wild-type genome. In this study, therefore, we elected to use a targeted hybridization next-generation sequencing (NGS) strategy to screen a panel of 151 genes for the presence of somatic variants. By focusing on a restricted gene set, we were able to achieve sufficient depth of coverage to allow the detection of variants present at variant allele frequencies as low as 1%. Our data suggest that somatic variants are present in the human lens epithelium, at frequencies consistent with the presence of millimeter sized clones. The potential implications of this finding for cataract formation are discussed in relation to the clonal organization of the lens epithelium and the lifelong growth process of the lens.

MATERIALS AND METHODS

Lens Epithelial Samples

Intact, de-identified human eyes or isolated lenses were obtained from Mid-America Transplant Services (St. Louis, MO, USA), Saving Sight (Kansas City, MO, USA), and the autopsy service of the Department of Pathology and Immunology (Washington University, St. Louis, MO, USA). Samples were obtained less than 48 hours postmortem and dissected immediately on arrival at the laboratory. In addition to donor lenses, surgical capsulorhexis specimens also were used. Capsulorhexis samples are small flaps of central anterior lens capsule with adherent epithelial cells and are removed (and routinely discarded) in the course of extracapsular cataract surgery. Ethical approval for the capsulorhexis study was obtained from the Washington University Human Research Protection Office (HRPO), and written informed consent was provided by all participants before enrollment, in accordance with the tenets of the Declaration of Helsinki and Health Insurance Portability and Accountability Act (HIPAA) regulations. A description of the samples used in this study is provided in Table 1.

Dissection of the Lens Epithelium

The base of a 35-mm Petri dish was covered with four layers of Parafilm, and a 6-mm-diameter circle was imprinted into the surface by pressing the blunt end of a pipette tip into the Parafilm. The base of the dish was filled with sufficient PBS (NaCl 137 mM; KCl 2.7 mM; Na₂HPO₄ 10 mM; KH₂PO₄ 1.8 mM) to prevent dehydration of the lens tissue during dissection. Lenses were released from donor eyes by cutting the ciliary zonule. Lenses were transferred to the Petri dish and oriented such that the epithelium faced down. Using surgical scissors, a circular piece of the posterior capsule approximately 7 mm in diameter was removed and discarded (Fig. 1). A series of radial cuts was made in the remaining portion of the posterior capsule. Lenses were positioned over the marked circle in the Parafilm and the capsule pinned to the base of the dish. The fiber cell mass was gently removed and discarded. In some cases, the entire epithelium was used. In other cases, the central region of the epithelium, corresponding to the 6-mm circle on the Parafilm, was excised carefully and collected in a microfuge tube. The remaining epithelium, referred to here as the peripheral epithelium, was collected in a separate microfuge tube. In some samples, cells were harvested from that region of the epithelium corresponding to the lower nasal quadrant (LNQ). For that purpose, the original orientation of the eye in the head was determined from the external anatomy of the globe, and the orientation of the lens in the eye was monitored during dissection by making a small mark on the capsule, as described.²⁰ The remaining quadrants (RQ) of the epithelium were collected in another tube.

Library Preparation

DNA was isolated from fresh or frozen tissue using QIAamp DNA micro kit (QIAGEN, Inc., Valencia, CA, USA). Genomic DNA (0.2–1.0 µg) was fragmented by sonication (Covaris E210; Covaris, Inc., Woburn, MA, USA) to an average size of approximately 175 bp. DNA fragments were concentrated using AMPureXP beads (Beckman Coulter Genomics, Danvers, MA, USA), and DNA ends were repaired using T4 DNA polymerase, Klenow polymerase, and T4 polynucleotide kinase. The 3' ends of the fragments were adenylated using exo-minus Klenow polymerase, allowing sequencing adapters to be ligated to the ends of the fragments. Adapter ligated DNA was subjected to limited (7–10 cycles) PCR amplification. Sequencing libraries were hybridized at 65°C for 24 hours to the Washington University Cancer Mutation Profiling version 2 (WUCaMP2) gene set²¹ using custom SureSelect biotinylated cRNA capture baits (Agilent Technologies, Santa Clara, CA, USA), according to the manufacturer's instructions. The WUCaMP2 panel includes coding exons from 151 cancer-related genes (see Table 2) totaling 0.83 Mb of captured sequence. The hybridized product was amplified for 14 PCR cycles using Agilent postcapture primers and a custom indexing primer.

Sequencing and Base Calling

Multiplex sequencing of the DNA libraries was performed on a HiSeq 2500 sequencing platform (Illumina Inc., San Diego, CA, USA) to obtain paired end 101-bp reads. Sequence reads were aligned to the human reference genome (hg19) using Novoalign (v. 3.02.06; Novocraft Technologies, Petaling Jaya, Malaysia). Polymerase chain reaction duplicates were removed from the alignments with Picard tools (v. 1.46; <http://broadinstitute.github.io/picard/>, available in the public domain), and variants were called using SAMtools (v. 0.1.18-1; <https://github.com/samtools/samtools>, available in the public

TABLE 1. Description of Tissue Samples

Donor ID	Age/ Gender	Tissue/Source	Race	Description	Ocular History
N199	47/male	Lens/Eye bank	White	Individual lens	Unremarkable
N931	71/male	Lens/Eye bank	White	Individual lens	Unremarkable
N785	70/male	Lens/Eye bank	White	Individual lens	Unremarkable
N793	17/female	Lens/Eye bank	White	Individual lens	Unremarkable
N934	73/female	Lens/Eye bank	White scissors</td <td>Individual lens</td> <td>Unremarkable</td>	Individual lens	Unremarkable
N26	63/male	Capsulorhexis	Black	Central epithelium	Mixed nuclear (3+) and cortical cataract, keratoconus
N29	67/male	Capsulorhexis	White	Central epithelium	Nuclear cataract (3+), Fuchs' corneal dystrophy
N335	58/male	Whole eye/WU pathology	White	Intact lens	Unremarkable
N589a/c (Pool)	17, 23, 43, 18	Lenses/Eye bank		Pool control/Pool control	
N1643	6/male	Lenses/Eye bank	White	Right lens/Left lens	Unremarkable
N1128	13/male	Lenses/Eye bank	Asian	Right lens/left lens	Unremarkable
N2403	26/male	Lenses/Eye bank	White	Right lens/Left lens	Unremarkable
N2463	74/male	Lenses/Eye bank	White	Right lens/Left lens	Cataract/Macular degeneration
N4446	71/female	Lenses/Eye bank	White	Center/Periphery	Unremarkable
N24172	68/female	Lenses/Eye bank	White	Center/Periphery	Unremarkable
N41100	51/female	Whole eyes/WU pathology	Black	Lower nasal quadrant/ Remaining quadrants	Unknown
N1474	68/female	Whole eyes/WU pathology	Black	Cornea Central lens epithelium Peripheral lens epithelium	Unknown
N14303	61/male	Whole eyes/WU pathology	White	Cornea Central lens epithelium Peripheral lens epithelium	Unknown
N146	37/female	Posterior pole/Eye bank	White	Retina Central lens epithelium Peripheral lens epithelium	Unknown
N20	66/female	Whole eyes/WU pathology	Black	Cornea Central lens epithelium Peripheral lens epithelium	Unknown
N11	67/male	Whole eyes/WU pathology	Black	Cornea Central lens epithelium Peripheral lens epithelium	Unknown

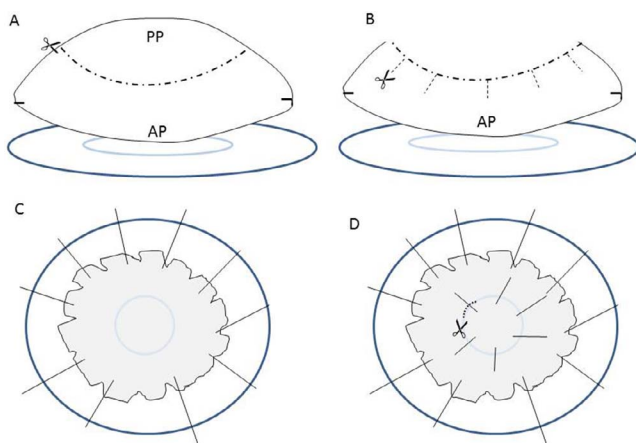


FIGURE 1. Dissection of the central and peripheral regions of the human lens epithelium. (A) The base of a 35-mm Petri dish (dark blue) is covered with four layers of Parafilm and imprinted with a 6-mm-diameter circle (light blue). The lens is positioned in the Petri dish with the anterior pole (AP) facing down. (B) A 7-mm-diameter circle is removed from the posterior pole (PP) and radial cuts are made in the remaining portion. (C) The lens is centered over the 6-mm circle, and the capsule is pinned to the base of the Petri dish bottom. The fiber cell mass is removed. (D) The central capsule (with its adherent layer of epithelial cells) is pinned to secure the tissue during dissection, and the central portion of the epithelium/capsule is removed using the 6-mm circle as a guide. The remaining part of the lens epithelium/capsule is collected separately.

domain). Variants were annotated with SnpEff software (<http://snpeff.sourceforge.net/>, available in the public domain).²² The mean coverages of the unique reads on target were obtained using Qualimap V2.1.1 (<http://qualimap.bioinfo.cipf.es/>, available in the public domain).²³ The coverage of targeted bases was calculated using Bedtools (<https://github.com/arq5x/bedtools2>, available in the public domain).²⁴ Genome visualization was performed using the Integrative Genome Browser data server.²⁵ A schematic showing the sequencing and bioinformatic workflow is shown in Figure 2.

Somatic Variant Calling

Somatic variants were identified using VarScan2 (<http://varscan.sourceforge.net/>). To remove false positives that may arise through sequencing or alignment errors, a false-positive filter is incorporated into the VarScan2 algorithm.²⁶ The filter evaluates putative variants according to nine empirically derived criteria designed to distinguish true somatic variants from artifactual calls. VarScan2 can operate in two modes: one for unpaired samples and the other for paired samples (somatic mode), where one of the two samples serves as a case and the other as a control.²⁷ For single tissue samples (Table 1), VarScan2 was run in unpaired mode, using default parameters.²⁶ To screen for putative somatic changes in unpaired samples, variants with frequencies lower than 40% were selected manually. From this group, statistically significant variants ($P < 0.05$; Fisher's Exact Test), covered to a

TABLE 2. Genes in the WUCaMP2 Cancer Panel Used for Targeted NGS of Lens Epithelial Samples

<i>ABCB1</i>	<i>CDKN2B</i>	<i>EGFR</i>	<i>GNAS</i>	<i>LAMA2</i>	<i>NOTCH1</i>	<i>RARG</i>	<i>SMARCB1</i>
<i>ABCC2</i>	<i>CEBPA</i>	<i>ERBB2</i>	<i>GSTP1</i>	<i>LCK</i>	<i>NPM1</i>	<i>RB1</i>	<i>SMO</i>
<i>ABL1</i>	<i>CHD7</i>	<i>ERBB3</i>	<i>H3F3</i>	<i>LTK</i>	<i>NRAS</i>	<i>RET</i>	<i>SNCAIP</i>
<i>ABL2</i>	<i>CHIC2</i>	<i>ERBB4</i>	<i>HNF1A</i>	<i>MAP2K1</i>	<i>PDGFRA</i>	<i>ROS1</i>	<i>SOS1</i>
<i>AKT1</i>	<i>CREBBP</i>	<i>ERG</i>	<i>HRAS</i>	<i>MAP2K2</i>	<i>PDGFRB</i>	<i>RPS6KB1</i>	<i>SPRED1</i>
<i>AKT2</i>	<i>CRLF2</i>	<i>ESR1</i>	<i>IDH1</i>	<i>MAP2K4</i>	<i>PHF6</i>	<i>RUNX1</i>	<i>SRC</i>
<i>AKT3</i>	<i>CSF1R</i>	<i>ESR2</i>	<i>IDH2</i>	<i>MAP3K1</i>	<i>PIK3CA</i>	<i>RXRA</i>	<i>STK11</i>
<i>ALK</i>	<i>CTNNB1</i>	<i>EZH2</i>	<i>IKZF1</i>	<i>MAPK1</i>	<i>PIK3R1</i>	<i>RXRB</i>	<i>SUFU</i>
<i>APC</i>	<i>CYP19A1</i>	<i>FBXW7</i>	<i>IL2RA</i>	<i>MED13</i>	<i>PSMB1</i>	<i>RXRG</i>	<i>TAS2R38</i>
<i>ASXL1</i>	<i>CYP2A6</i>	<i>FGFR1</i>	<i>IL2RB</i>	<i>MET</i>	<i>PSMB2</i>	<i>SHH</i>	<i>TET2</i>
<i>ATM</i>	<i>CYP2B6</i>	<i>FGFR2</i>	<i>IL2RG</i>	<i>MLH1</i>	<i>PSMB5</i>	<i>SHOC2</i>	<i>TP53</i>
<i>ATRX</i>	<i>CYP2C19</i>	<i>FGFR3</i>	<i>INPP4B</i>	<i>MLL</i>	<i>PSMD1</i>	<i>SLC22A1</i>	<i>TRRAP</i>
<i>BRAF</i>	<i>CYP2C9</i>	<i>FGFR4</i>	<i>JAK1</i>	<i>MPL</i>	<i>PSMD2</i>	<i>SLC22A2</i>	<i>TYK2</i>
<i>BRCA1</i>	<i>CYP2D6</i>	<i>FLT1</i>	<i>JAK2</i>	<i>MST1R</i>	<i>PTCH1</i>	<i>SLC31A1</i>	<i>UGT1A1</i>
<i>BRCA2</i>	<i>DDR1</i>	<i>FLT3</i>	<i>JAK3</i>	<i>MTOR</i>	<i>PTEN</i>	<i>SLC34A2</i>	<i>VHL</i>
<i>CBL</i>	<i>DDR2</i>	<i>FLT4</i>	<i>KDM6A</i>	<i>MYC</i>	<i>PTPN11</i>	<i>SLC45A3</i>	<i>WT1</i>
<i>CDA</i>	<i>DDX3X</i>	<i>FSTL5</i>	<i>KDR</i>	<i>MYD88</i>	<i>RAF1</i>	<i>SLCO1B1</i>	<i>YES1</i>
<i>CDH1</i>	<i>DNMT3A</i>	<i>GNA11</i>	<i>KIT</i>	<i>NELL2</i>	<i>RARA</i>	<i>SMAD4</i>	<i>ZMYM3</i>
<i>CDKN2A</i>	<i>DPYD</i>	<i>GNAQ</i>	<i>KRAS</i>	<i>NF1</i>	<i>RARB</i>	<i>SMARCA4</i>	

depth greater than 300 with a minimum of four reads supporting the variant in both forward and reverse strands were evaluated further using the Integrative Genomics Viewer (IGV).

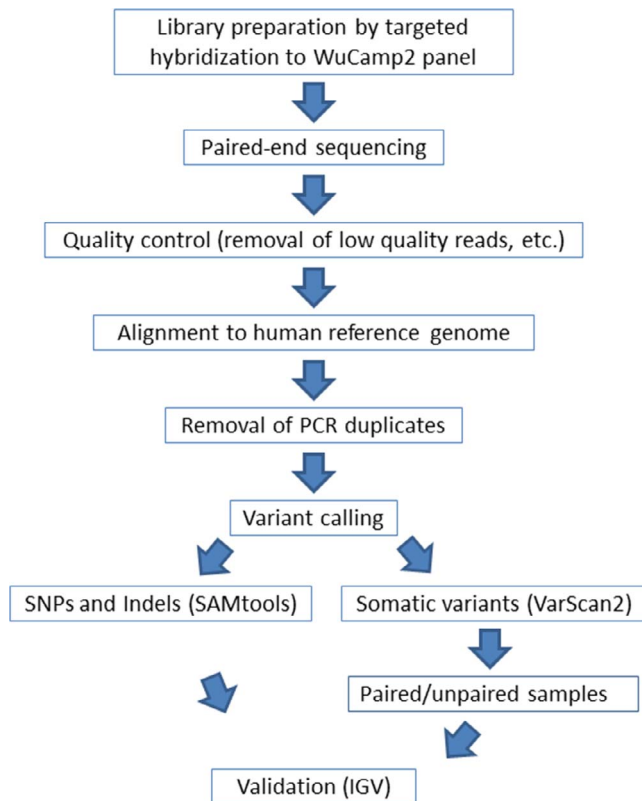


FIGURE 2. Work flow used to identify somatic variants in genomic DNA extracted from human lens epithelial cells. Sequencing libraries were enriched for genes of the WUCaMP2 panel (Table 2) by targeted hybridization capture. The resulting library was sequenced on an Illumina HiSeq 2500 platform to obtain paired-end reads. Sequencing results were filtered at several levels. The reads were then aligned to a human reference genome (hg19), and PCR duplicates were removed. Variants were called using SAMtools and VarScan2 software. Finally, selected variants were inspected manually using the IGV visualization tool.

VarScan2 was used in somatic (paired) mode when matched case/control tissue pairs (for example lens/cornea or lens/retina) from a single donor were available (Table 1). Developed originally as a method for identifying rare variants in tumor samples, here we used VarScan2 to identify variants in lens samples that were absent from control tissues, allowing us to exclude single nucleotide polymorphisms (SNPs) or germline variants. The mpileup input files for VarScan2 were obtained from SAMtools mpileup v. 0.1.18 using default parameters, with Phred Base quality ≥ 20 and Phred Mapping quality ≥ 30 . To call somatic variants, VarScan2 was run using the following parameters: Min coverage: 50x for Normal, 50x for “Tumor” (usually, lens); Min reads 2:2; Min strands 2:1; Min var freq: 0.01; Min freq for hom: 0.75; Normal purity: 1.0; Tumor purity: 1.0; Min avg qual: 15; *P* value thresh: 0.99; Somatic *P* value: 0.05, for all other settings, default parameters were used. Loss of heterozygosity (LOH) variants were taken as somatic variants. The following additional criteria were used to call a somatic variant: a minimum of 300 reads at the position of the variant for both case and control, a minimum of four reads supporting the variant present, in both forward and reverse strands, and the absence of the variant allele from the matched tissue. When a third specimen from the same donor was included as an extra control in the analysis (see Table 1), a minimum of 250 reads at the position of the variant for the case, a minimum of 300 reads at the position of the variant in the two controls, a minimum of four reads supporting the variant present in both forward and reverse strands, and the absence of the variant allele from the other two control specimens were required to score the variant as somatic. Finally, the reads supporting each putative somatic variant were evaluated manually using the IGV program. Variants found in reads with questionable mapping (more than three discrepancies), close to indels, or at the beginning or end of reads were discarded.^{25,28}

Droplet Digital PCR

In droplet digital PCR (ddPCR) a genomic DNA sample is partitioned into thousands of small droplets. A PCR reaction is performed within each droplet, using primers designed to differentiate between wild-type and variant sequences. The number of positive droplets is measured by fluorescence activated cell sorting (FACS) technology and the data are interpreted using Poisson statistics. The ddPCR technique was

developed originally to quantify rare variants in the presence of a large background signal from the wild-type genome²⁹ and variant allele frequencies as low as 1 in 200,000 have been measured successfully using the ddPCR approach.³⁰ Here, we used ddPCR to validate independently one of the variants detected in the NGS analysis.

Digital droplet PCR was performed on a QX200 system (Bio-Rad Laboratories, Hercules, CA, USA) using TaqMan chemistry. Assays were designed as per MIQE guidelines.³¹ Primers and probes were used at final concentrations of 900 nM and 250 nM, respectively. Approximately 10 ng of genomic DNA was added to each reaction. After droplet generation, the droplets (approximately 15,000 per reaction) were transferred to an Eppendorf TwinTec PCR 96-well plate. A temperature gradient (52° to 62°C) was used to determine optimum annealing temperature, using oligonucleotides for wild-type and variant sequences. The plate was run on a Bio-Rad C1000 thermal cycler: 95°C for 10 minutes, 95°C for 30 seconds, and 54.5°C for 1 minute × 39 cycles, 98°C for 10 minutes, and a 12°C hold. Droplets were read on a Bio-Rad QX200 droplet reader, and data were analyzed using QuantaSoft 1.7 (Bio-Rad). Thresholds were set manually, based on the location of clusters in the negative and positive controls. The negative control consisted of PCR reaction mix to which no DNA was added, whereas the positive control was a mixture of 10 ng/mL of genomic wild-type DNA and 2.5×10^{-5} ng of the variant synthetic DNA fragment. The variant was considered to be present in droplets with fluorescent signals of greater than 4000 in the FAM channel and less than 3000 in the HEX channel. The absolute copy number (copies/ μ L) of the variant allele (CN_{VAR}) and the wild-type allele (CN_{WT}) in each reaction was calculated from the Poisson distribution law³² and the variant allele fraction (VAF) was calculated as follows, $VAF = CN_{VAR}/(CN_{VAR} + CN_{WT})$.

RESULTS

Isolation of Genomic DNA From Ocular Tissue Samples

We obtained lens epithelia from several sources, including surgical specimens and cadaveric tissue (Table 1). It was helpful to have two tissue samples from a single donor (so that somatic and germline variants could be readily differentiated), but depending on the source, this was not always possible. Another challenge was obtaining sufficient DNA for sequencing experiments. The lens epithelium contains approximately 0.5×10^6 cells,³³ and this limited number of cells did not always yield sufficient DNA for the analysis. This was particularly true of the surgical (capsulorhexis) specimens, which constituted only a small (4–5-mm diameter) piece of the central lens epithelium. Nine pairs of capsulorhexis specimens had to be discarded for this reason. In postmortem samples, the lens cells did not always adhere tightly to the capsule, and this inevitably resulted in cell loss on dissection, with diminished yields of DNA as a consequence. Thus, only 50% of lens pairs, 33% of central/peripheral tissue pairs, and 25% of LNQ/RQ pairs provided sufficient high-quality DNA for sequencing. In total, 41 tissue samples from 24 donors/cataract patients were included in the study (Table 1).

Sequencing Assessment

The mean total number of reads for each of the tissue samples examined in this study was 32×10^6 (Supplementary Table S1). On average, 98% of the reads mapped to the reference sequence, and 97% were properly paired. An average of 56% of the reads mapped to the target region, implying high

capture efficiency during library preparation. A high read depth was achieved across the target region. For example, after removal of PCR duplicates, 99% of the targeted bases were covered at ≥ 100 -fold, 79% at ≥ 500 -fold, and 42% at ≥ 1000 -fold (Supplementary Table S2). Across all samples, the mean sequencing depth was 1018-fold (Supplementary Table S1).

To gauge the reproducibility of our sequencing/bioinformatics workflow, well-mixed DNA from a pooled sample of lens epithelial DNA (sample N598; Table 1) was divided into two aliquots (N589a and N589c), which were analyzed independently. The sequencing results, enrichment, alignment and coverage were comparable for N589a and N589c (Supplementary Table S1). In each sample, 269 known SNPs, 2 novel SNPs, and 12 Indels (insertions/deletions) were detected (Supplementary Table S3). Thus, at least for high-frequency variants (SNPs have a variant allele frequency of 50% or 100%, depending on whether they are heterozygous or homozygous), the sequencing technique was extremely reproducible.

Variant Calling From Unpaired Samples

For eight of our samples, no control tissue was available (these comprised six lenses and two capsulorhexis specimens from eight individual donors; see Table 1). Sequence data obtained from these samples were, therefore, examined with VarScan2 in unpaired mode (Table 3). In the eight unpaired samples, 123 variants with frequencies of less than 25% were identified. This list was refined further by selecting variants covered to a depth of greater than 300 and supported by both forward and reverse strands. Of the 34 variants that met those conditions, 23 (67.5%) were detected in multiple samples (Table 3). Notably, 21 of the 23 variants were located in *CYP* family genes (*CYP2D6* and *CYP2A6*). Moreover, the variants at positions 41354661 and 41354606 were observed at low frequency in 35 of the 39 tissue samples sequenced in this study. It is unlikely that the same point substitution would occur independently in multiple samples, implying that the variants detected in the *CYP* family genes reflected sequencing or alignment errors. Manual inspection of the seven repetitive variants (Table 3) revealed that the paired end reads did not map to a single gene, but rather to two members of the *CYP* gene family located on the same chromosome. These imprecise alignments, a consequence of the high-sequence homology between *CYP* genes, were a source of false positives, and variants associated with these genes were therefore discarded. Similar observations were obtained for the unique variants (see Table 3) in *CYP2D6* (sample N26) and *CYP2A6* (sample N931). From the 34 variants identified initially, only nine (from samples N29, N793, and N934; Table 3) passed all quality control filters, with frequencies ranging from 1.02% to 3.94%. Of these, four were nonsynonymous missense substitutions and a fifth, in *ALK*, represented a splice site donor mutation.

Calling Somatic Variants From Paired Samples

VarScan2 was designed for calling low-frequency somatic variants from deep sequencing experiments on matched tissue samples. The algorithm, intended primarily for cancer studies, compares data from pairs of tissues (in our case, left versus right lens epithelia, lens versus cornea, or LNQ of the epithelium versus the RQs) obtained from a single individual. The software compares the number of reads containing the variant in each tissue sample and calculates a corresponding *P* value, using Fisher's Exact Test. Variants are classified as germline (if a variant is found in both control and case tissues), somatic (if the variant is found in the case only), or LOH (if the

TABLE 3. Somatic Variants Detected in Unpaired Lens Specimens

	Position	RefSNP	Ref Allele	Depth	VAF	p-value	ADF	ADR	IGV	Gene	N29	N199	N335	N785	N793	N931	N934	N26
chr19	41354606	4986892	G	1632	6.25	2.0727E-30	33	69	fail	CYP2A6	X	X	X			X	X	X
chr22	41355849	2302990	A	797	7.65	1.3145E-19	6	55	fail	CYP2A6		X	X					X
chr22	42523003	116917064	A	612	18.3	7.1446E-37	38	74	fail	CYP2D6				X		X	X	X
chr19	41354661	79584535	T	1826	3.89	7.8724E-21	8	63	fail	CYP2A6	X	X	X		X	X	X	X
chr22	42523528	1058172	C	3165	32.86	0	377	663	fail	CYP2A6	X			X				
chr22	42524795	0	A	2281	24.2	1.0167E-178	370	182	fail	CYP2A6				X		X	X	X
chr9	80537112	0	T	1384	1.3	0.000036295	15	3	fail	GNAQ								
chr19	42523514	61745683	C	2592	6.87	4.8943E-52	85	93	fail	CYP2D6								
chr22	42525772	28371706	G	2868	31.73	6.1735E-307	420	490	fail	CYP2D6								
chr7	98574379	0	A	939	3.94	5.0677E-12	1	36	pass	TRRAP								
chr13	32914236	4987117	C	977	1.02	0.0009542	4	6	pass	BRCA2								
chr13	28624294	1933437	G	1521	1.05	0.00013261	7	9	pass	FLT3								
chr2	212251864	3748962	T	1657	1.15	0.000019114	5	14	pass	ERBB4								
chr1	226259111	0	T	597	1.34	0.0038151	7	1	pass	H3F3A								
chr9	139391636	2229974	G	2918	1.34	3.4932E-10	21	18	pass	NOTCH1								
chr6	160670282	316019	A	1470	1.5	2.6616E-06	19	3	pass	SLC22A2								
chr19	41354538	60605885	G	1364	2.05	4.9051E-08	22	6	fail	CYP2A6								
chr2	29462545	0	C	1346	1.34	0.000036244	17	1	pass	ALK								

NON_SYN_CODING MODERATE | MISSENSE | Acc/Ccc | T2720P |
NON_SYN_CODING MODERATE | MISSENSE | aCg/aTg | T1915M |
NON_SYN_CODING MODERATE | MISSENSE | aCg/aTg | T227M |
SYNONYMOUS_CODING (LOW) | SILENT | gtA/gtG | V1049 |
SYNONYMOUS_CODING (LOW) | SILENT | caT/caC | H114 |
SYNONYMOUS_CODING (LOW) | SILENT | gaC/gaT | D2185 |
NON_SYN_CODING (MODERATE) | MISSENSE | Tct/Gct | S270A |

SPlice_SITE_DONOR (HIGH) || || | 1620 | CODING |

variant is found in the control tissue only). In our study, LOH variants were included under the somatic variant designation.

A total of 9622 putative variants (ranging from 615 to 1038 variants per sample) were detected in the 13 sets of two- and three-tissue comparisons (Table 4). Most of these were known SNPs or Indels, which are inherited through the germline and not considered further here. However, VarScan2 also identified 1001 low-allelic fraction variants, which were classified, provisionally, as somatic variants. Evaluation of the control DNA samples N589a and N589c (consisting of aliquots from a single pooled sample that were sequenced independently) with VarScan2 yielded 55 putative variants (Table 4), an indication of the potential false-positive discovery rate in the initial screen. Additional filters (described in Materials and Methods) were applied to further suppress the false-positive discovery rate. Following filtration, the number of variants in the control sample was reduced to zero. In the other 12 sets of tissue, 33 somatic variants passed all filters and are described in detail in Table 5.

We examined the occurrence of somatic variants in four pairs of lens epithelia, obtained from donors (N1643, N1128, N2403, and N2463) ranging in age from 6 to 74 years (Table 5). Six somatic variants were identified in these samples, at frequencies of 1.02% to 2.72% (frequencies were derived from the number of reads supporting a variant/total number of reads). None of the variants were present in the SNP database and two were the result of C→T transitions at dipyrimidine sites, so-called UV signature mutations.

Central lens epithelial cells are located directly on the optical axis of the eye and, therefore, exposed to sunlight. Conversely, cells located in the peripheral lens epithelium are situated in the shadow cast by the iris. We explored whether differential exposure to sunlight results in an uneven distribution of somatic variants across the epithelium. For these experiments, we extracted DNA from cells located in the sunlight-exposed central epithelium and compared it to DNA extracted from cells in the shaded peripheral epithelium (samples N4446, N24172, N1474, N14303, N11, N146, and N20). In some cases, a third tissue sample (cornea or retina) was included in the analysis (Table 5). An example of one such analysis is shown in Figure 3, and tabulated data for all the comparisons are provided in Table 5.

In central and peripheral samples, most (>90%) variants were present at frequencies of 50% or 100%, representing heterozygous or homozygous SNPs, respectively (see Fig. 3). However, some variants (identified using VarScan2) were present at markedly lower frequencies (1%–10%). Many of the low-allelic fraction variants were excluded from further analysis because they did not pass the quality control filters. Somatic variants were identified in six of the seven pairs of matched central/peripheral lens samples. In only one pair (N1474) were more somatic variants detected in the central epithelium than in the peripheral epithelium (Table 5). Thus, at least in this limited sample, there was no indication of a preponderance of somatic variants in the sunlight-exposed central region of the epithelium. Optical modeling of light exposure in the eye indicates that light exposure may not be equal at all radial locations and irradiance is predicted to be greatest in the LNQ of the lens epithelium.³⁴ Using DNA from donor N41100, we compared the type and number of somatic variants in the LNQ with the remaining epithelial quadrants (RQ; Table 5). Five somatic variants were detected in the LNQ sample versus a single variant in the RQ sample.

In comparing sequence data from sets of matched ocular tissues obtained from 12 donors, 33 somatic variants were detected in 26 genes of the WUCaMP2 gene set. The spatial distribution of the variants is shown schematically in Figure 4.

TABLE 4. Variant Calling in Lens Specimens Based on Comparison With One or More Control Tissues

Donor ID	Tissue Description	Total Variants (Includes SNPs and Indels)	Putative Somatic Variants	Validated Somatic Variants	Validated Variants in Nonlens Tissue
N589	Pool a	1038	55	0	
	Pool c			0	
N1643	Right lens	647	38	1	
	Left lens			0	
N1128	Right lens	726	121	0	
	Left lens			3	
N2403	Right lens	687	37	0	
	Left lens			1	
N2463	Right lens	701	80	1	
	Left lens			0	
N4446	Central epithelium	615	41	0	
	Peripheral epithelium			0	
N24172	Central epithelium	630	32	0	
	Peripheral epithelium			1	
N41100	RQs	877	114	1	
	LNQ			5	
N1474	Central epithelium	943	214	11	
	Peripheral epithelium			1	
N14303	Cornea*	617	40		0
	Central epithelium			1	
N146	Peripheral epithelium	667	64	0	
	Retina*			3	
N20	Central epithelium	736	75		0
	Peripheral epithelium			2	
N11	Cornea*	738	90		0
	Central epithelium			0	
	Peripheral epithelium			1	
	Cornea*				0

* Nonlens tissue.

Droplet Digital PCR Analysis

For most tissue samples used in this study, the amount of DNA was limiting and, consequently, the entire DNA sample was generally used for sequencing. In one case (N146), however, there was a small surplus of DNA. This was used to verify independently the existence of one of the somatic variants identified by deep sequencing and obtain an absolute measure of variant allele frequency by ddPCR analysis. Sequencing analysis of the tissue samples from donor N146 identified a somatic variant (T > C; position 32907082) in the *BRCA2* gene (Table 5) that was present in DNA extracted from the peripheral epithelium (red circle in Fig. 3) but absent from the central epithelium or the retina (Fig. 4C). Next generation sequencing suggested that the variant was present at a frequency of 1.10%. Primers were designed to differentiate the variant from the wild-type sequence, and ddPCR was performed on lens and retinal DNA from donor N146. Positive controls (consisting of wild-type or mutant DNA) were used to set threshold values, as described in Materials and Methods. In test samples, droplets meeting the threshold conditions were

detected only in samples prepared from the peripheral lens epithelium and not from the central epithelium or retina (Fig. 5). In the peripheral lens epithelial sample, the variant allele frequency for the targeted *BRCA2* variant was 1.3%, close to the value of 1.1% obtained from the deep sequencing analysis (Table 5). Thus, using an orthogonal technique, ddPCR, we were able to substantiate the variant call from the deep sequencing data.

DISCUSSION

Epidemiologic studies have identified an elevated risk of developing cortical cataracts and PSCs after exposure to UV-B or ionizing radiation but the molecular mechanisms remain obscure.

The indeterminate threshold dose and increasing latency period between low-dose exposure and cataract formation suggests that, at least for radiation-induced opacities, the underlying process could involve a mutational component. It has been suggested that genotoxic damage may also contribute

TABLE 5. Detailed Annotation of Validated Somatic Variants Identified in Two- or Three-Way Tissue Comparisons

Donor ID	CHROM	Position	RefSNP	REF	ALT	P	Gene	Depth	VAF	FWD	REV	Depth	VAF	FWD	REV	Third Sample Variants/Reads
N589	(-)	(-)	0	(-)	(-)	0	(-)	0	0%	0	0	0	0%	0	0	0/0
N1643	chrX	44949963	0	C	T	0.03437	KDM6A	400	Right lens 1.25%	4	1	384	Left lens	0	0	0/0
N1128	chr4	54915245	0	G	A	0.04315	CHIC2	366	Right lens	0	0	308	Left lens	2	2	2/2
	chr4	143159058	0	A	G	0.04102	INPP4B	399	0%	0	0	328	1.22%	2	2	2/2
	chr21	39769161	0	A	G	0.04068	ERG	441	0%	0	0	361	1.11%	2	2	2/2
N2403	chrX	44950032	0	A	G	0.03378	KDM6A	471	Right lens	0	0	488	Left lens	2	3	3/3
N2463	chr1	179090780	0	C	T	3.18E-09	ABL2	956	Right lens 2.72%	16	10	1059	Left lens	0	0	0/0
N41100	chr3	38180223	0	C	T	0.00415	MYD88	834	RQ	0	0	420	LNQ	3	2	2/2
	chr7	98608635	0	A	G	0.04591	TRRAP	445	0%	0	0	385	1.04%	2	2	2/2
	chr8	128752436	rs182162908	C	T	0.00414	MYC	442	0%	0	0	452	1.77%	3	5	5/5
	chr8	128752448	0	C	T	0.00371	MYC	509	0%	0	0	506	1.58%	3	5	5/5
	chrX	70473134	0	A	G	0.03498	BCYRN1	519	0%	0	0	397	1.01%	1	3	3/3
	chr12	25368450	0	T	C	0.0296	KRAS	397	1.01%	2	2	558	0%	0	0	0/0
N4446	(-)	(-)	0	(-)	(-)	0	(-)	0	Central epithelium	0	0	0	Peripheral epithelium	0	0	0/0
N24172	chrX	44919509	0	A	G	0.04751	KDM6A	358	Central epithelium	0	0	315	Peripheral epithelium	2	2	2/2
N1474	chr1	98060708	0	T	C	0.01033	DPYD	272	Central epithelium	2	2	578	Peripheral epithelium	0	0	0/695 cornea
	chr2	30143174	0	C	A	0.01428	ALK	347	1.47%	2	2	654	0%	0	0	0/668 cornea
	chr5	112043498	0	C	A	0.01204	APC	371	1.15%	2	2	744	0%	0	0	0/585 cornea
	chr6	117715449	0	G	T	0.01595	ROSI	383	1.08%	2	2	692	0%	0	0	0/786 cornea
	chr7	140624434	0	C	T	0.02684	BRAF	266	1.04%	2	2	389	0%	0	0	0/363 cornea
	chr10	6059898	rs7899538	C	A	0.03315	IL2RA int	358	1.50%	2	2	479	0%	0	0	0/463 cornea
	chr10	6067944	0	T	C	0.01825	IL2RA	356	1.12%	2	2	610	0%	0	0	0/637 cornea
	chr10	123353427	0	C	A	0.00951	FGFR2	372	1.12%	2	2	816	0%	0	0	0/766 cornea
	chr19	3118975	0	T	C	0.01121	GNAI1	330	1.08%	1	3	681	0%	0	0	0/470 cornea
	chr19	11097546	0	G	C	0.02461	SMARCA4	353	1.21%	2	2	536	0%	0	0	0/463 cornea
	chr19	11100029	rs201099398	T	C	0.01332	SMARCA4	374	1.13%	2	2	724	0%	0	0	0/565 cornea
	chr19	4101298	0	G	A	0.0443	MAP2K2	345	1.07%	0	0	509	1.18%	3	3	3/3
N14303	chr17	60024371	0	A	G	0.02841	MED13	444	0%	3	2	471	Peripheral epithelium	0	0	0/459 cornea
N11	chr13	28578078	0	C	A	0.04798	FLT3	445	Central epithelium	0	0	393	Peripheral epithelium	2	2	2/2

TABLE 5. Continued

Donor ID	CHROM	Position	RefSNP	REF	ALT	P	Gene	Depth	VAF	FWD	REV	Depth	VAF	FWD	REV	Third Sample Variants/Reads
N146	chr6	117718203	0	A	G	0.03558	<i>ROS1</i>	490	Central epithelium	2	3	463	Peripheral epithelium	0	0	0/346 retina
	chr4	142950131	0	T	C	0.03636	<i>INPP4B</i>	487	Central epithelium	0	0	379	Peripheral epithelium	2	2	0/460 retina
	chr13	32907082	0	T	C	0.01666	<i>BRCA2</i>	643	Central epithelium	0	0	362	Peripheral epithelium	2	2	0/602 retina
	chr13	49039515	0	T	C	0.01329	<i>RBI</i>	567	Central epithelium	0	0	300	Peripheral epithelium	2	3	0/461 retina
N20	chr3	178937724	0	T	C	0.0419	<i>PIK3CA</i>	469	Central epithelium	0	0	389	Peripheral epithelium	2	2	0/438 cornea
	chr11	118360001	0	A	G	0.03548	<i>KMT2A</i>	491	Central epithelium	0	0	378	Peripheral epithelium	2	2	0/564 cornea

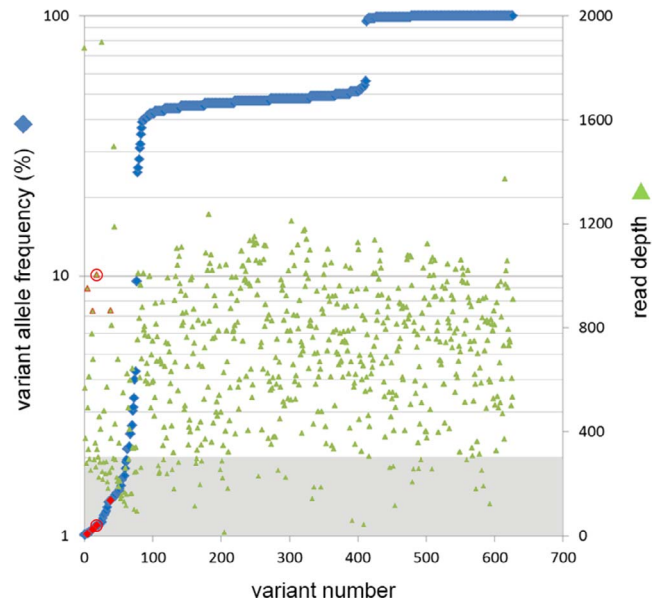


FIGURE 3. Variant allele frequency distribution in the peripheral lens epithelial sample from donor N146 (see Table 1). A total of 628 variants (blue diamonds) are detected with frequencies ranging from 1% to 100%. For each variant, the corresponding read depth is indicated (green triangle). Variants with frequencies of 50% or 100% represent heterozygous or homozygous SNPs, respectively. Seventy-eight low-allelic fraction (<10%) variants are detected. Many of the low-allelic fraction variants were not called as true somatic variants because they were associated with read depths of less than 300 (shaded area) or because they failed other quality control criteria (see Materials and Methods). Four variants (red diamonds) passed all filters (including sufficient read depth [red triangles]). The variant circled in red (the associated read depth value is also encircled in red) corresponds to a T > C substitution in the *BRCA2* gene (see Table 5).

to the formation of common, age-related cataracts and, in that regard, it may be significant that micronuclei, biomarkers of DNA damage, are present in epithelial cells from cataractous lenses.³⁵ Until recently, however, sequencing methods have lacked the sensitivity to determine directly whether somatic variants are present in the human lens.

Application of massively parallel sequencing technologies has begun to reveal the full burden of somatic mutations in untransformed mammalian cells. In postmitotic neurons, for example, each cell appears to have a distinct genome, harboring hundreds³⁶ or perhaps thousands of single nucleotide variants.³⁷ In sunlight-exposed tissues, such as skin, the mutational burden can be especially heavy, with a preponderance of C > T UV-induced signature mutations.³⁸ In such a setting, mutations that confer a growth advantage are positively selected, leading to the emergence of subclonal lineages.

The lens is a unique tissue, containing populations of actively cycling cells, growth-arrested cells, and postmitotic cells. By virtue of its role in image formation, the lens is also subjected to lifelong exposure to light. Irradiance is not uniform, however, and cells located in the pupil space are exposed to considerably more light than those at the periphery. Irradiance may be highest of all in the LNQ of the epithelium.³⁴ The current study was designed to assess whether somatic variants can be detected in the aging lens, the nature of the underlying sequence alterations, and the distribution of variants with respect to latitude and radial location in the epithelium (i.e., whether variants are more common in sunlight-exposed regions or regions harboring the proliferating cell populations).

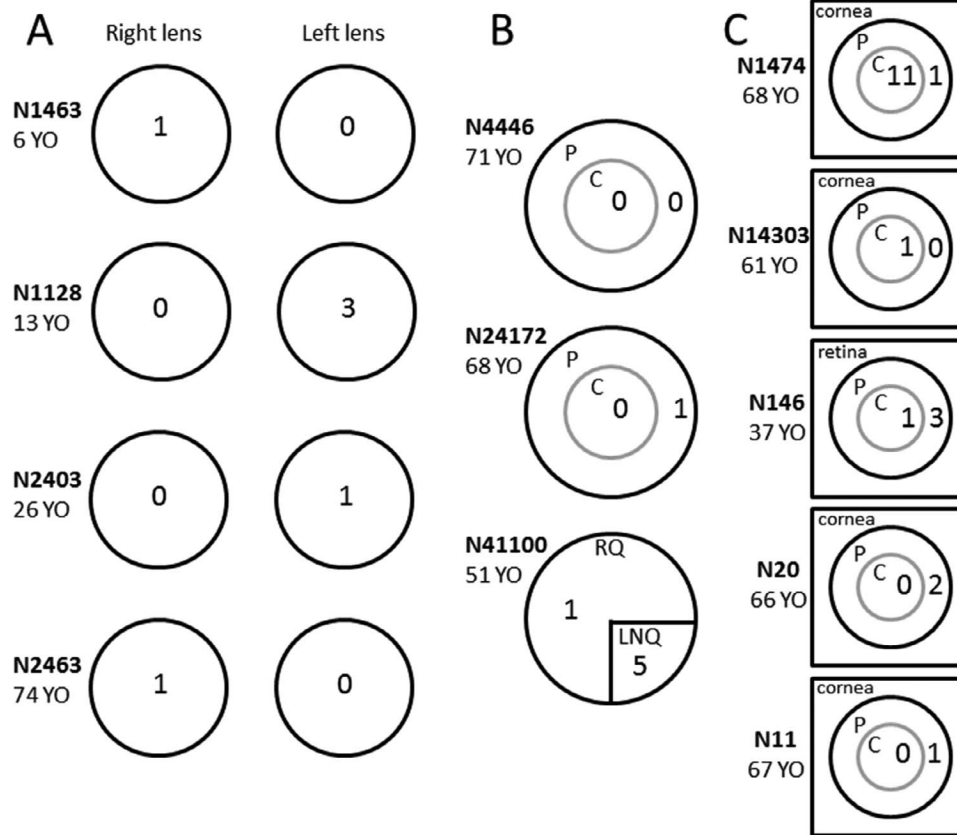


FIGURE 4. Somatic variants in matched ocular specimens. Tissue comparisons consisted of epithelia from right versus left lenses (A), central versus peripheral epithelium from individual lenses (B), or central versus peripheral lens epithelium versus cornea or retina (C). The number of somatic variants detected in each region is indicated. Additional information about the individual samples or the sequencing results is available in Tables 1, 4, and 5. P, peripheral epithelium; C, central epithelium; YO, years old.

In total, we detected 41 somatic variants in 151 genes from 29 human lens epithelial samples. The age of the donors varied from 6 to 74 years. In other model organisms, somatic variants accumulate with age, but interestingly, at different rates in different tissues.³⁹ Variants arise through misincorporation of nucleotides during DNA replication or from improperly repaired lesions arising between replication cycles, so-called clock mutations.^{40,41} In our study, however, we were unable to confirm that variants are more common in samples from older patients, although such an association may be detected in a more comprehensive analysis. Similarly, we did not observe a consistent pattern in the distribution of variants across the lens epithelium; sometimes variants were concentrated in the central epithelium, and sometimes they were more numerous in the periphery. A single experiment on the angular distribution of variants within the epithelium revealed that variants were most common in the LNQ. This is an interesting observation because that region is exposed to more sunlight than the rest of the epithelium,³⁴ and it is there that cortical cataracts often develop initially.⁴² It has been hypothesized that germinative zone cells in the LNQ may be particularly prone to UV damage through the phenomenon of peripheral light focusing.⁴³ Clearly, more experiments will be required to substantiate the initial sequencing results.

Seven of the 41 somatic variants detected in the lens samples were C > T substitutions. Of these, four occurred at dipyrimidine sites and may, therefore, constitute UV-induced substitutions. It is instructive to contrast the mutational burden in the lens epithelium with that of cells in the nearby eyelid epidermis. In a recent study of 234 eyelid biopsies from four

donors, 3760 somatic mutations were identified in a panel of 74 genes.³⁸ This is a much higher mutational burden than seen in lens samples in the current study. Moreover, many of the mutations identified in skin were “driver” mutations (i.e., mutations that confer a growth advantage and thereby lead to clonal expansion) in important signaling pathways (notably, the Notch pathway), although the samples were nominally free from cancer. Significantly, most mutations in eyelid skin were C > T or CC > TT UV-signature mutations, thought to result from misrepaired CPDs.¹³ A higher burden of mutations might be expected in skin because the epidermal cells are exposed directly to sunlight. In contrast, to reach the lens epithelium, light has first to pass through the cornea. Although transmissivity of human cornea to visible light is high, experiments on enucleated eyes suggest that much of the UV-B is absorbed as it passes through the tissue,⁴⁴ reducing the likelihood of CPD formation in the underlying lens cells.

The variant frequencies detected in the current study ranged from 1% to 4%, and it is pertinent to ask whether these values are reliable, and if so, what do they represent at the cellular level? The ability of massively parallel sequencing to identify somatic mutations is limited, ultimately, by the fidelity of the sequencing reactions⁴⁵ and extremely rare variants are currently undiscoverable using this technology. For allelic fractions of 1% and higher, however, the VarScan2 program has been shown to reliably call variants, provided a sufficiently high sequence depth is achieved.⁴⁶ We took all reasonable steps to exclude false positives (including sequencing the same pooled sample of DNA twice and manually inspecting the reads supporting every putative variant). In trying to be conserva-

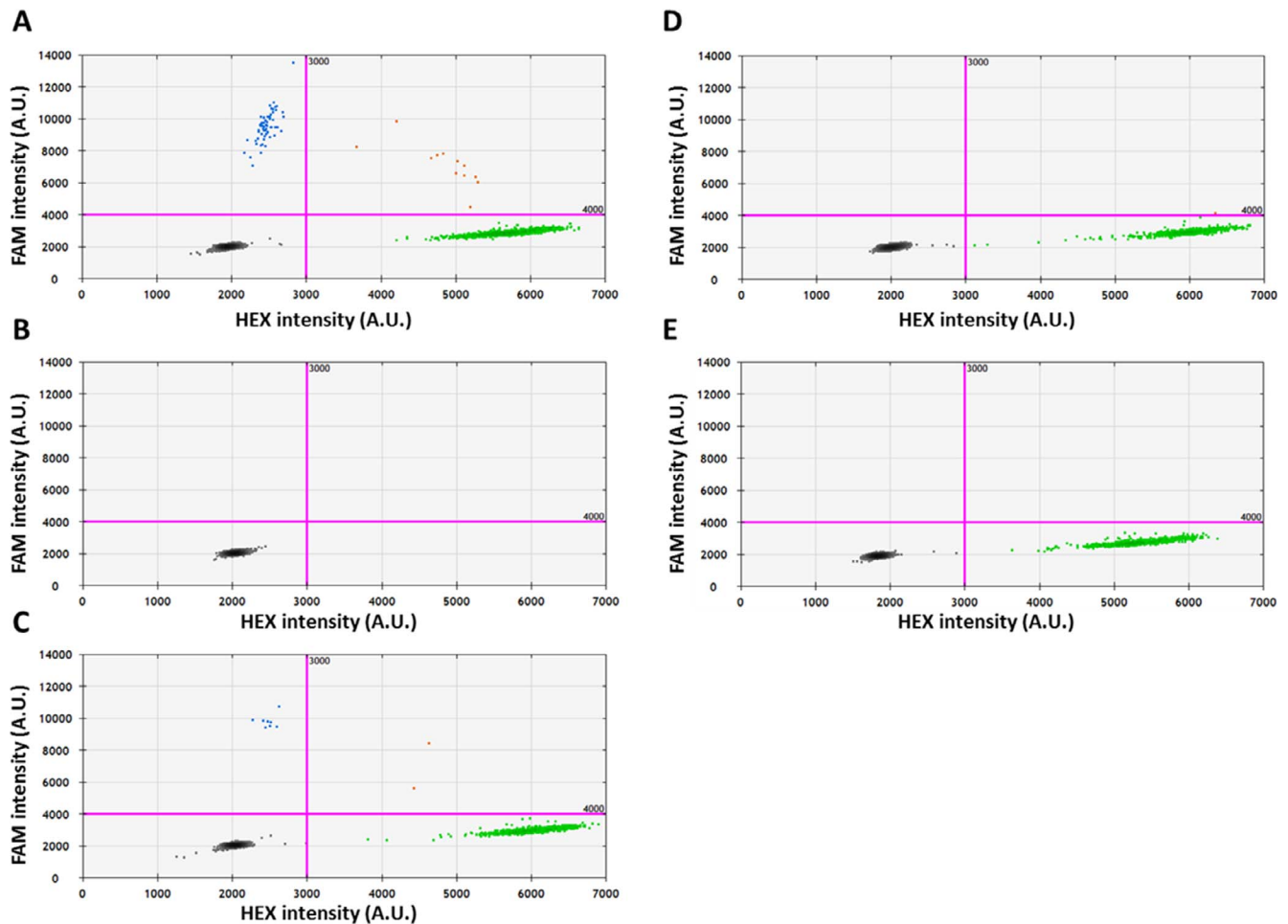


FIGURE 5. The presence of a *BRCA2* somatic variant detected in sample N146 (see Table 5; Fig. 3) is validated by ddPCR analysis. A TaqMan probe labeled with HEX fluorophore was designed to detect the wild-type (WT) *BRCA2* allele, and a probe labeled with FAM fluorophore was designed to detect the variant *BRCA2* sequence. The threshold lines (pink) were determined by the positive control (A), to separate clusters with WT or variant DNA. Green dots represent droplets with WT DNA, blue dots represent droplets with variant DNA, brown dots represent droplets with both WT and variant DNA, and black dots represent droplets with no target DNA. The variant was called as present in droplets with fluorescence units in FAM > 4000 and with fluorescence units in HEX < 3000. The DNA from the peripheral lens region (C) shows the presence of the variant, whereas samples from the central lens region (D) and from the retina (E) show the presence of WT DNA only. In the negative control (B), no HEX or FAM droplets were identified.

tive, we probably excluded some valid variants, as noted by others.²⁶ However, it is difficult to categorically exclude the possibility that some of the variants may have arisen from errors in sequencing or library preparation. On the single occasion we had sufficient DNA, we were able to use an orthogonal technique, ddPCR, to confirm independently the variant allele frequency of a *BRCA2* T > C mutation identified through NGS sequencing. Unfortunately, it is not practicable to devise custom ddPCR assays to independently validate each putative variant.

The presence of a variant at a frequency of approximately 1% (for example, the *BRCA2* T→C variant identified in sample N146 and confirmed by ddPCR) suggests that the variant was present in approximately 2% of the cells sampled (assuming that the variant allele was heterozygous). The primate lens epithelium contains approximately 0.5×10^6 cells. The peripheral lens epithelial sample in which the *BRCA2* variant was detected constituted approximately half of the epithelial area, implying that the variant was present in approximately 5000 cells. The cell density of the peripheral human lens epithelium is approximately 5500 cells mm^{-2} .⁴⁷ The *BRCA2* mutant clone would therefore cover an area of approximately 1

mm^2 . Eyelid epidermis is modeled as a complex mosaic of somatic mutations nested in a clonal architecture, and large (millimeter-sized) clones are thought to result from the presence of driver mutations that confer a cellular growth advantage.³⁸

Approximately half of the somatic variants detected in lens epithelial cells were nonsynonymous changes, with the potential to impact cell growth kinetics. As in the epidermis, the presence of such variants may confer a growth advantage that accounts for the emergence of large clones of cells. Alternatively, large clonally related cell clusters may form through the normal growth process of the lens. Lineage tracing studies in mouse lenses have shown that epithelial cells undergo multiple rounds of cell division as they traverse the germinative and pre-germinative zones, leading to the generation of clonal clusters numbering 30 to 50 cells.^{48–50} Although this is two orders of magnitude smaller than the predicted clone size in humans, we note that primate lenses contains many more cells than the mouse lens (0.5×10^6 vs. 4.3×10^4), and of course, humans have a significantly longer life span. It is likely that over the course of the human life span, lens epithelial cells undergo many more cell divisions than mouse

lens cells. As a consequence, it is plausible that somatic variants arising early in a lens epithelial cell lineage may reach frequencies of greater than 1% later in life.

In the current study, we screened a panel of genes (WUCaMP2) implicated in cancer for the presence of somatic variants. Although some of these genes may have relevance for cataract development, there is no reason to suspect, a priori, that sequence variations in these particular genes contribute directly to lens cell opacification. They were chosen primarily because they have been validated for use in hybridization capture analysis.²¹ However, a set of genes with essential roles in the maintenance of lens transparency has been identified through studies of inherited forms of cataract.⁵¹ Many of these genes encode proteins that are strongly or specifically expressed in lens tissue, such as α -, β -, or γ -crystallin, and gap junction proteins, such as connexin46 and connexin50. In brain tissue, single nucleotide variants are most common in actively transcribed genes, suggesting that mutations are introduced during transcription rather than DNA replication.³⁷ If this is also true in the lens, then genes encoding crystallins and other strongly expressed proteins may be particularly prone to somatic mutation. The Cat-Map, an up-to-date listing of genes implicated in Mendelian or age-related forms of cataract, now includes more than 200 genes distributed across all 22 autosomes and the X-chromosome.⁵² Future sequencing efforts will focus on this set of genes with demonstrated roles in the maintenance of lens fiber cell transparency. If somatic variants can be identified in Cat-Map genes, it will strengthen the hypothesis that somatic mutations may play a contributory role in human cataract formation.

Acknowledgments

The authors thank the staff at Saving Sight St. Louis and Mid-America Transplant Services for help with tissue procurement. Brandon McKethan (Bio-Rad) helped design the primers used in the ddPCR studies. We thank the Genome Technology Access Center in the Department of Genetics at Washington University for help with genomic analysis and, in particular, Paul Cliften for his assistance with the bioinformatics.

The Genome Technology Access Center is partially supported by National Cancer Institute Cancer Center Support Grant P30 CA91842 to the Siteman Cancer Center and by Clinical and Translational Science Award (CTSA) Grant UL1 TR000448 from the National Center for Research Resources, a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. This work was also supported by NIH grant National Eye Institute R01 EY09852 (SB), core grant for vision research P30 EY02687, and an unrestricted grant to the Department of Ophthalmology and Visual Sciences from Research to Prevent Blindness.

Disclosure: **R. Mesa**, None; **M. Tyagi**, None; **G. Harocopos**, None; **D. Vollman**, None; **S. Bassnett**, None

References

- Bourne RR, Stevens GA, White RA, et al. Causes of vision loss worldwide, 1990-2010: a systematic analysis. *Lancet Glob Health*. 2013;1:e339-e349.
- Richter GM, Choudhury F, Torres M, Azen SP, Varma R. Los Angeles Latino Eye Study Group. Risk factors for incident cortical, nuclear, posterior subcapsular, and mixed lens opacities: the Los Angeles Latino eye study. *Ophthalmology*. 2012;119:2040-2047.
- Truscott RJ, Friedrich MG. The etiology of human age-related cataract. Proteins don't last forever. *Biochim Biophys Acta*. 2016;1860:192-198.
- Klein BE, Klein R, Linton KL. Prevalence of age-related lens opacities in a population. The Beaver Dam Eye Study. *Ophthalmology*. 1992;99:546-552.
- Stewart FA, Akleyev AV, Hauer-Jensen M, et al.; for the International Commission on Radiological Protection. ICRP publication 118: ICRP statement on tissue reactions and early and late effects of radiation in normal tissues and organs—threshold doses for tissue reactions in a radiation protection context. *Ann ICRP*. 2012;41:1-322.
- Chylack LT Jr, Peterson LE, Feiveson AH, et al. NASA study of cataract in astronauts (NASCA). Report 1: cross-sectional study of the relationship of exposure to space radiation and risk of lens opacity. *Radiat Res*. 2009;172:10-20.
- Worgul BV, Kundiye YI, Sergiyenko NM, et al. Cataracts among Chernobyl clean-up workers: implications regarding permissible eye exposures. *Radiat Res*. 2007;167:233-243.
- Chodick G, Bekiroglu N, Hauptmann M, et al. Risk of cataract after exposure to low doses of ionizing radiation: a 20-year prospective cohort study among US radiologic technologists. *Am J Epidemiol*. 2008;168:620-631.
- Hall P, Granath F, Lundell M, Olsson K, Holm LE. Lenticular opacities in individuals exposed to ionizing radiation in infancy. *Radiat Res*. 1999;152:190-195.
- Shore RE. Radiation impacts on human health: certain, fuzzy, and unknown. *Health Phys*. 2014;106:196-205.
- Neriishi K, Nakashima E, Akahoshi M, et al. Radiation dose and cataract surgery incidence in atomic bomb survivors, 1986-2005. *Radiology*. 2012;265:167-174.
- Neriishi K, Nakashima E, Minamoto A, et al. Postoperative cataract cases among atomic bomb survivors: radiation dose response and threshold. *Radiat Res*. 2007;168:404-408.
- Brash DE. UV signature mutations. *Photochem Photobiol*. 2015;91:15-26.
- Mesa R, Bassnett S. UV-B-induced DNA damage and repair in the mouse lens. *Invest Ophthalmol Vis Sci*. 2013;54:6789-6797.
- McCarty CA, Taylor HR. A review of the epidemiologic evidence linking ultraviolet radiation and cataracts. *Dev Ophthalmol*. 2002;35:21-31.
- Kammeyer A, Luiten RM. Oxidation events and skin aging. *Ageing Res Rev*. 2015;21:16-29.
- Maher VM, Ouellette LM, Curren RD, McCormick JJ. Frequency of ultraviolet light-induced mutations is higher in xeroderma pigmentosum variant cells than in normal human cells. *Nature*. 1976;261:593-595.
- Dollfus H, Porto F, Caussade P, et al. Ocular manifestations in the inherited DNA repair disorders. *Surv Ophthalmol*. 2003;48:107-122.
- Chi XX, Liu YY, Shi SN, Cong Z, Liang YQ, Zhang HJ. XRCC1 and XPD genetic polymorphisms and susceptibility to age-related cataract: a meta-analysis. *Mol Vis*. 2015;21:335-346.
- Harocopos GJ, Alvares KM, Kolker AE, Beebe DC. Human age-related cataract and lens epithelial cell death. *Invest Ophthalmol Vis Sci*. 1998;39:2696-2706.
- Goyal B, Duncavage EJ, Martinez D, Lewis JS Jr, Chernock RD. Next-generation sequencing of salivary high-grade neuroendocrine carcinomas identifies alterations in RB1 and the mTOR pathway. *Exp Mol Pathol*. 2014;97:572-578.
- Cingolani P, Platts A, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6:80-92.
- Okonechnikov K, Conesa A, Garcia-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*. 2016;32:292-294.

24. Quinlan AR. BEDTools: the Swiss-Army tool for genome feature analysis. *Curr Protoc Bioinformatics*. 2014;47:11.12.11-11.12.34.
25. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14:178-192.
26. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22:568-576.
27. Koboldt DC, Larson DE, Wilson RK. Using VarScan 2 for germline variant calling and somatic mutation detection. *Curr Protoc Bioinformatics*. 2013;44:15.14.11-15.14.17.
28. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29:24-26.
29. Vogelstein B, Kinzler KW. Digital PCR. *Proc Natl Acad Sci U S A*. 1999;96:9236-9241.
30. Pekin D, Skhiri Y, Baret JC, et al. Quantitative and sensitive detection of rare mutations using droplet-based microfluidics. *Lab Chip*. 2011;11:2156-2166.
31. Bustin SA, Benes V, Garson JA, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem*. 2009;55:611-622.
32. Pinheiro LB, Coleman VA, Hindson CM, et al. Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Anal Chem*. 2012;84:1003-1011.
33. Kuszak JR, Costello MJ. The structure of the vertebrate lens. In: Lovicu FJ, Robinson ML, eds. *Development of the Ocular Lens*. Cambridge: Cambridge University Press; 2004:71-118.
34. Coroneo MT, Müller-Stolzenburg NW, Ho A. Peripheral light focusing by the anterior eye and the ophthalmohelioses. *Ophthalmic Surg*. 1991;22:705-711.
35. Worgul BV, David J, Odrich S, et al. Evidence of genotoxic damage in human cataractous lenses. *Mutagenesis*. 1991;6:495-499.
36. Hazen JL, Faust GG, Rodriguez AR, et al. The complete genome sequences, unique mutational spectra, and developmental potency of adult neurons revealed by cloning. *Neuron*. 2016;89:1223-1236.
37. Lodato MA, Woodworth MB, Lee S, et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science*. 2015;350:94-98.
38. Martincorena I, Roshan A, Gerstung M, et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*. 2015;348:880-886.
39. Dolle ME, Snyder WK, Dunson DB, Vijg J. Mutational fingerprints of aging. *Nucleic Acids Res*. 2002;30:545-549.
40. Gao Z, Wyman MJ, Sella G, Przeworski M. Interpreting the dependence of mutation rates on age and time. *PLoS Biol*. 2016;14:e1002355.
41. Alexandrov LB, Jones PH, Wedge DC, et al. Clock-like mutational processes in human somatic cells. *Nat Genet*. 2015;47:1402-1407.
42. Sasaki H, Kawakami Y, Ono M, et al. Localization of cortical cataract in subjects of diverse races and latitude. *Invest Ophthalmol Vis Sci*. 2003;44:4210-4214.
43. Coroneo M. Ultraviolet radiation and the anterior eye. *Eye Contact Lens*. 2011;37:214-224.
44. Lombardo M, Pucci G, Barberi R, Lombardo G. Interaction of ultraviolet light with the cornea: clinical implications for corneal crosslinking. *J Cataract Refract Surg*. 2015;41:446-459.
45. Gundry M, Vijg J. Direct mutation analysis by high-throughput sequencing: from germline to low-abundant, somatic variants. *Mutat Res*. 2012;729:1-15.
46. Stead LF, Sutton KM, Taylor GR, Quirke P, Rabbitts P. Accurately identifying low-allelic fraction variants in single samples with next-generation sequencing: applications in tumor subclone resolution. *Hum Mutat*. 2013;34:1432-1438.
47. Karim AK, Jacob TJ, Thompson GM. The human anterior lens capsule: cell density, morphology and mitotic index in normal and cataractous lenses. *Exp Eye Res*. 1987;45:865-874.
48. Shi Y, Barton K, De Maria A, Petrash JM, Shiels A, Bassnett S. The stratified syncytium of the vertebrate lens. *J Cell Sci*. 2009;122:1607-1615.
49. Shi Y, De Maria A, Lubura S, Sikic H, Bassnett S. The penny pusher: a cellular model of lens growth. *Invest Ophthalmol Vis Sci*. 2015;56:799-809.
50. Sikic H, Shi Y, Lubura S, Bassnett S. A stochastic model of eye lens growth. *J Theor Biol*. 2015;376:15-31.
51. Shiels A, Hejtmancik JF. Molecular genetics of cataract. *Prog Mol Biol Transl Sci*. 2015;134:203-218.
52. Shiels A, Bennett TM, Hejtmancik JF. Cat-Map: putting cataract on the map. *Mol Vis*. 2010;16:2007-2015.