

# Optimization and Validation of a Virtual Reality Orientation and Mobility Test for Inherited Retinal Degenerations

Jean Bennett<sup>1,2</sup>, Elena M. Aleman<sup>2</sup>, Katherine H. Maguire<sup>2</sup>, Jennifer Nadelmann<sup>1</sup>, Mariejel L. Weber<sup>1,2</sup>, William M. Maguire<sup>2</sup>, Ayodele Maja<sup>2</sup>, Erin C. O'Neil<sup>2,3</sup>, Albert M. Maguire<sup>1-3</sup>, Alexander J. Miller<sup>4</sup>, and Tomas S. Aleman<sup>1-3</sup>

<sup>1</sup> Scheie Eye Institute at the Perelman Center for Advanced Medicine, Philadelphia, PA, USA

<sup>2</sup> Center for Advanced Retinal and Ocular Therapeutics at the University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

<sup>3</sup> Division of Ophthalmology at the Children's Hospital of Philadelphia of the Department of Ophthalmology, Philadelphia, PA, USA

<sup>4</sup> Neurology Virtual Reality Laboratory of the Department of Neurology, University of Pennsylvania, Philadelphia, PA, USA

**Correspondence:** Tomas S. Aleman, Perelman Center for Advanced Medicine, University of Pennsylvania, 3400 Civic Center Boulevard, Philadelphia, PA 19104, USA. e-mail: [aleman@penmedicine.upenn.edu](mailto:aleman@penmedicine.upenn.edu)

**Received:** October 5, 2022

**Accepted:** November 20, 2022

**Published:** January 30, 2023

**Keywords:** virtual reality; mobility; orientation; Leber congenital amaurosis (LCA); retinitis pigmentosa (RP); multi-luminance mobility test (MLMT)

**Citation:** Bennett J, Aleman EM, Maguire KH, Nadelmann J, Weber ML, Maguire WM, Maja A, O'Neil EC, Maguire AM, Miller AJ, Aleman TS. Optimization and validation of a virtual reality orientation and mobility test for inherited retinal degenerations. *Transl Vis Sci Technol.* 2023;12(1):28. <https://doi.org/10.1167/tvst.12.1.28>

**Purpose:** To optimize a virtual reality (VR) orientation and mobility (O&M) test of functional vision in patients with inherited retinal degenerations (IRDs).

**Methods:** We developed an O&M test using commercially available VR hardware and custom-generated software. Normally sighted subjects ( $n = 20$ , ages = 14–67 years) and patients with IRDs ( $n = 29$ , ages = 15–63 years) participated. Individuals followed a dim red arrow path to a “course exit,” while trying to identify nine obstacles adjacent to, or directly in their path. Dark-adapted subjects completed 35 randomly selected VR courses at increasing luminances, twice per luminance step, binocularly, and unilaterally. Performance was graded automatically by the software. Patients with IRD completed a modified Visual Function Questionnaire (VFQ).

**Results:** Normally sighted subjects identified approximately 50% of the obstacles at the dimmest course luminance. Except for two patients with IRD with poor vision, all patients were able to complete the test, although they required brighter (by  $>2$  log units) luminances to identify 50% of the obstacles. In a single-luminance screening test in which normal subjects detected at least eight of nine objects, most patients with IRD underperformed; their performance related to disease severity, as measured by visual acuity, kinetic visual field extent, and VFQ scores. Test-retest differences in object detection were similar to the differences between the two eyes ( $\pm 2$  SD =  $\pm 2$  objects).

**Conclusions:** This VR-O&M test was able to distinguish subjects with IRDs from normal subjects reliably and reproducibly.

**Translational Relevance:** This easily implemented, flexible, and objectively scored VR-O&M test promises to become a useful tool to assess the impact that IRDs and their treatments have on functional vision.

## Introduction

The various attributes of vision, such as visual resolution or acuity and light sensitivity, are commonly tested in isolation, both in experimental environments as well as in the clinic. These attributes do not consistently translate into a level of functional vision, or the ability of a subject to use his or her version to perceive and interact with their environment.<sup>1–7</sup> A large number

of academic and industry groups are now testing approaches with which to measure functional improvement after experimental therapies, especially in individuals who are severely visually impaired or totally blind. Although inherently subjective from the point of view of the individual undergoing the tests, functional vision should be assessed as objectively as possible, especially when the tests are used as an outcome measure for many of the treatment strategies that are now in clinical trials in the inherited retinal degeneration (IRD)

clinic, including gene augmentation therapy, antisense oligonucleotide therapy, gene editing, optogenetics, transplantation, and visual prostheses.<sup>5,8–11</sup> There is thus a need for rapid, noninvasive, safe, and sensitive tests that allow quantitation of the individual's ability to perceive the physical environment. Such tests should properly incorporate the multiple aspects of vision that are vital for activities of daily living (visual resolution, luminance, visual field extent and sensitivity, movement, etc.) and be relevant to severely visually impaired individuals.

The trial that developed the Argus II retinal prosthesis introduced custom-designed functional assessments displayed on a computer monitor. These tests included measures of the ability to find a square, detect direction of motion and to see large objects. A real-world assessment included finding a door that was randomly placed on an opposing wall. A number of groups have used corridors as obstacle courses.<sup>12–14</sup> One group used an entire city block for such a purpose and another group used a (slightly smaller) Pedestrian Accessibility Movement Environment Laboratory (PAMELA) as a maze.<sup>3,15</sup> Jacobson et al. developed an obstacle course 27 m long with moveable ceiling-anchored wall segments and obstacles and measured accuracy at defined light levels.<sup>16</sup> The group at Children's Hospital of Philadelphia (CHOP) running a gene therapy clinical trial for Leber congenital amaurosis (LCA) due to *RPE65* mutations developed a standardized multi-luminance mobility test (MLMT), a physical test allowing for tracking functional vision changes at specified light levels over time in low-vision patients.<sup>17</sup> Results were scored by a reading center and a composite score assessing speed and accuracy resulted in grading of performances as pass or fail. The MLMT was the primary outcome measure used by Spark Therapeutics that showed improvement in vision after delivery of the gene therapy reagent now known as Luxturna (voretigene neparvovec-rzyl).<sup>18</sup> Ora, Inc. (Andover, MA) has since developed another test of the ability to navigate a physical obstacle course which, like the MLMT, is carried out under different light conditions (<https://www.oraclinical.com/resource/oras-keith-lane-discussing-oras-visual-mobility-course/>). Results are videotaped and sent to a grading center. The Ora-VNC test is being used as an outcome measure in several clinical trials (for example, see Ref. 19). Meanwhile, Institut de la Vision developed "Streetlab," an artificial street designed to reflect an urban environment, complete with audio recordings of urban soundscapes.<sup>20</sup> The subject wears a velcro jumpsuit with an array of tracking devices. The course utilizes 5 lighting conditions ranging from 1 lux to 235 lux and the physical obstacles in the course are modelled on

real-life items, such as a hose, a ladder, etc. A motion capture system records various aspects of the trajectory, timing, and collisions, and, like the MLMT and Ora-VNC tests, performance analyses are made from videotapes. Although these physical mobility tests can potentially be modified to probe specific visual conditions, they have great limitations, including difficulty in set-up, they present trip hazards to the subject, and there are physical requirements which limit their implementation at multiple centers. Furthermore, the scoring systems are cumbersome, time-consuming, and, because often the data are captured by video, include risk not only of bias by the person doing the scoring but also risk of divulging confidential patient information. In addition, the tests are often valid for only a subset of subjects. For example, individuals with choroideremia (who had good visual acuity late in their disease but had severely restricted visual fields and nyctalopia) scored well on the MLMT, whereas patients with LCA perform poorly on the same test.

Thus, limitations in the tests include the fact that they do not: (1) address diverse mechanisms of vision loss; (2) may not have the sensitivity to detect gradual changes in visual function, or (3) accommodate the wide spectrum of severity presented by the clinically heterogeneous group of IRDs. There is a need for a robust test which is easily modifiable to reflect disease- and stage-specific visual function, can quantify the function in a light-tight, homogeneously illuminated space, and can rapidly, sensitively, and reliably measure functional vision in a setting with minimal hardware and personnel.

In previous studies, we generated a virtual reality protocol that evaluates an individual's ability to navigate quickly and accurately through a set of virtual "obstacles."<sup>21</sup> We demonstrated in subjects with Leber congenital amaurosis due to *RPE65* mutations (*RPE65*-LCA) that individuals were able to navigate this system more accurately and at lower luminance after they received Luxturna gene therapy.<sup>21</sup> In the process, we identified details that could be improved upon in order to generate a system that can be used efficiently at multiple centers to both screen for deficits in functional vision and to monitor changes in functional vision due to disease progression and/or after therapeutic intervention.

There were two main concerns in the previous version of the test. Whereas both normally sighted control subjects and most affected individuals could track the path of arrows correctly, they frequently collided with objects as they moved forward. Furthermore, it was difficult to be certain which were accidental collisions of perceived objects and which occurred because the subject did not perceive the object at all.

To address this problem, in the current study we: (a) reduced the total number of objects (from 15 to 9) and (b) also instituted a mechanism whereby the subject was instructed to “tag” the obstacles if they saw them. The tagging of objects allows scoring of performance immediately and objectively, thereby eliminating the need to analyze video footage.

The current study includes not only the object detection system but also: (1) use of a tetherless virtual reality (VR) headset; (2) incorporation of virtual obstacles of the sort that present real life challenges in daily living to vision impaired individuals; (3) testing using a defined set of luminance values over a 3-log range of intensity; (4) implementation of a standardized test paradigm that randomly presents any of 35 different course designs in order to minimize potential learning effect; (5) an automated system for grading performance; and (6) a scoring process that accurately and reproducibly identifies presence and severity of visual impairment in a wide range of IRDs. The work paves the way to fully develop and validate this approach for use in a variety of diagnostic, natural history, and translational studies.

## Methods

### Subjects

Informed consent or assent and parental permission was obtained from all of the subjects; the procedures adhered to the Declaration of Helsinki and the studies approved by the Institutional Review Board of the University of Pennsylvania (protocol # 815348). Normally sighted subjects (from here onward “normal subjects”  $n = 20$ , ages = 14–67 years) and patients with different forms of IRDs ( $n = 29$ , ages = 15–63 years) participated (see the Table). Patients underwent a comprehensive eye examination, including best corrected visual acuities (VAs) and Goldmann kinetic visual fields (GVFs).

### Virtual Reality System and Environment

#### Headset and Trackers

The VR mobility test paradigm is modified from the VR test that we described in detail previously. Subjects are fitted with a VR headset and controllers (Oculus Quest 2, Burlingame, CA). Both the headset and hand-held controllers contain trackers that register the position of the subject 360 degrees in the virtual space. The controllers can be used to interact with objects within the virtual space. The head-mounted VR device is fitted under dim (red) illumina-

tion after 20 minutes of dark-adaptation. The subjects can wear corrective glasses comfortably under the VR headpiece so that the test can be performed using best-corrected distance visual acuity. The wireless headset consists of two organic LEDs (OLEDs) with a resolution of  $1832 \times 1920$  pixels (90 Hz refresh rate). The OLEDs displays render a better image quality with better contrast compared to conventional LED displays.<sup>22</sup> Each display can be turn on or off independently allowing for binocular or unocular testing conditions. The horizontal field of view subtends approximately 89 degrees and a vertical field of view of approximately 93 degrees (<https://smartglasseshub.com/oculus-quest-2-fov/>). Before starting the session and with guidance from the test-giver, the participant adjusts the inter-eye distance of the lenses (using the slider at the base of the headset) until the image that the participant sees in the goggles is maximally sharp. This adjustment is designed for individuals with different inter-pupillary distances (range = 58–68 mm). Although measuring luminance is easy in physical orientation and mobility (O&M) tests by placing sensors in the middle of a physical space and assuming that all object surfaces are illuminated equally, we worried that in the optically complex VR space such measures would require undertaking assumptions that could compromise the reproducibility of the test by other investigators. As we did in our last work and for simplicity, we measured instead the maximal luminance of the system by pointing our radiometer (ILT1700, International Light Technologies, Inc. Peabody, MA) to the empty achromatic background illuminated at the maximal output of the VR device. From here on, the values were then expressed as the aggregate luminances of all of the objects present in any given scene, each of them being fractions of the whole. The maximal luminance of the achromatic “empty” background was  $79.7 \text{ cd.m}^{-2}$  at the subject’s eye plane. Photometric comparisons with physical O&M courses will not be straight forward but the maximum illuminance of the system is 80 Lux (measured with a Sekonic Flash Master L-358; Tokyo, Japan). Using this methodology, we verified the overall luminance of the obstacle course under four “standard” luminance steps ( $-0.67$ ,  $-0.19$ ,  $+0.39$ , and  $+0.69 \text{ log phot.cd.m}^{-2}$ ). To extend the operating range of the instrument toward the low luminance range, neutral density (ND) filters positioned in front of the viewing screens attenuated the luminance of the virtual scene by 1.5 log units increments; two sandwiched ND filters (3 log units of attenuation) were required in to attenuate the luminance enough to challenge dark-adapted normal subjects in distinguishing objects within the virtual scene.

**Table.** Clinical Characteristics of the Patients

Subject ID	Age/Gender	Diagnosis <sup>†</sup>	Visual Acuity LogMar		Kinetic Perimetry <sup>§</sup>		VFQ Score <sup>#</sup>
			OD	OS	OD	OS	OD
VR05	33/M	<i>RPE65</i> -LCA <sup>‡</sup>	1.20	1.10	0.43	0.26	97
VR32*	17/F	<i>RPE65</i> -LCA <sup>‡</sup>	0.63	0.86	0.30 <sup>  </sup>	0.31 <sup>  </sup>	N/A
VR33*	15/M	<i>RPE65</i> -LCA <sup>‡</sup>	1.04	0.98	0.21 <sup>  </sup>	0.33 <sup>  </sup>	N/A
VR48	21/M	<i>RPE65</i> -LCA <sup>‡</sup>	0.62	0.76	0.47 <sup>  </sup>	0.37 <sup>  </sup>	N/A
VR61	21/F	<i>RPE65</i> -LCA <sup>‡</sup>	0.58	0.48	0.22 <sup>  </sup>	0.15 <sup>  </sup>	N/A
VR63	44/M	<i>RPE65</i> -LCA	1.60	1.60	0.51	0.88	163
VR64	43/F	<i>EYS</i> -arRP	0.02	0.02	0.12	0.15	247
VR65	35/F	RP simplex	0.80	0.70	0.18	0.14	235
VR66	41/M	RP simplex	0.14	0.06	0.15	0.21	ND
VR67	41/M	<i>EYS</i> -arRP	0.16	0.10	n/a	n/a	ND
VR68	34/F	<i>EYS</i> -arRP	0.04	0.06	0.18	0.05	ND
VR69	31/M	<i>RPE65</i> -LCA <sup>‡</sup>	1.92	1.13	0.00 <sup>  </sup>	0.00 <sup>  </sup>	ND
VR70	56/M	RP multiplex	5.00	4.00	0.16	0.29	16
VR71	62/F	RP simplex	1.60	0.80	0.82	0.85	111
VR72	17/M	<i>CNGB1</i> -arRP	0.06	0.10	0.74	0.80	203
VR73*	33/F	<i>NR2E3</i> -adRP	0.04	0.06	0.50	0.25	209
VR74*	63/M	<i>NR2E3</i> -adRP	0.06	0.34	0.04	0.03	211
VR75	46/M	<i>RPGR</i> -XLRP	3.00	3.00	0.29	0.90	72
VR76	32/M	Unilateral RP	0.24	0.14	0.75	0.51	185
VR77	42/M	<i>CRKL</i> -CRD	0.20	0.20	0.16	0.17	184
VR78	48/M	<i>PRPH2</i> -adRP	0.06	8.00	0.17	0.12	179
VR79	43/M	<i>RPGR</i> -CRD	1.90	2.20	0.00	0.00	103
VR80	41/M	CHM	0.10	0.54	0.18	0.35	123
VR81	36/M	<i>USH2A</i> -USH	0.50	0.52	0.19	0.12	88
VR82	16/M	<i>RPE65</i> -LCA <sup>‡</sup>	0.41	0.81	0.46	0.50	164
VR83	41/M	<i>RPE65</i> -LCA <sup>‡</sup>	1.00	1.10	0.26	0.18	96
VR84	43/M	<i>RPE65</i> -LCA	4.00	4.00	0.00	0.00	30
VR85*	26/F	<i>PRPF31</i> -adRP	N/A	N/A	N/A	N/A	N/A
VR86*	29/M	<i>PRPF31</i> -adRP	N/A	N/A	N/A	N/A	N/A

N/A, not available.

\*VR32, and VR33, VR85, and VR86 are siblings; V74 is the father of VR73.

<sup>†</sup>Clinical diagnoses: RP, retinitis pigmentosa; adRP, Autosomal Dominant RP; arRP, autosomal recessive RP; XLRP, X-linked RP; LCA, Leber congenital amaurosis; CRD, cone-rod dystrophy; CHM, choroideremia; USH, Usher syndrome. Molecular diagnoses are shown with gene name preceding associated clinical diagnosis/phenotype; otherwise, the presence (multiplex) or absence (simplex) of family history is specified.

<sup>‡</sup>Patients treated bilaterally with Luxturna.

<sup>§</sup>Goldmann kinetic visual field (GVF) is the visual field extent (in degrees) along the entire horizontal meridian on either side of fixation, expressed as a percentage of normal extent for the stimulus used (target V-4e for all patients, except, exceptions below).

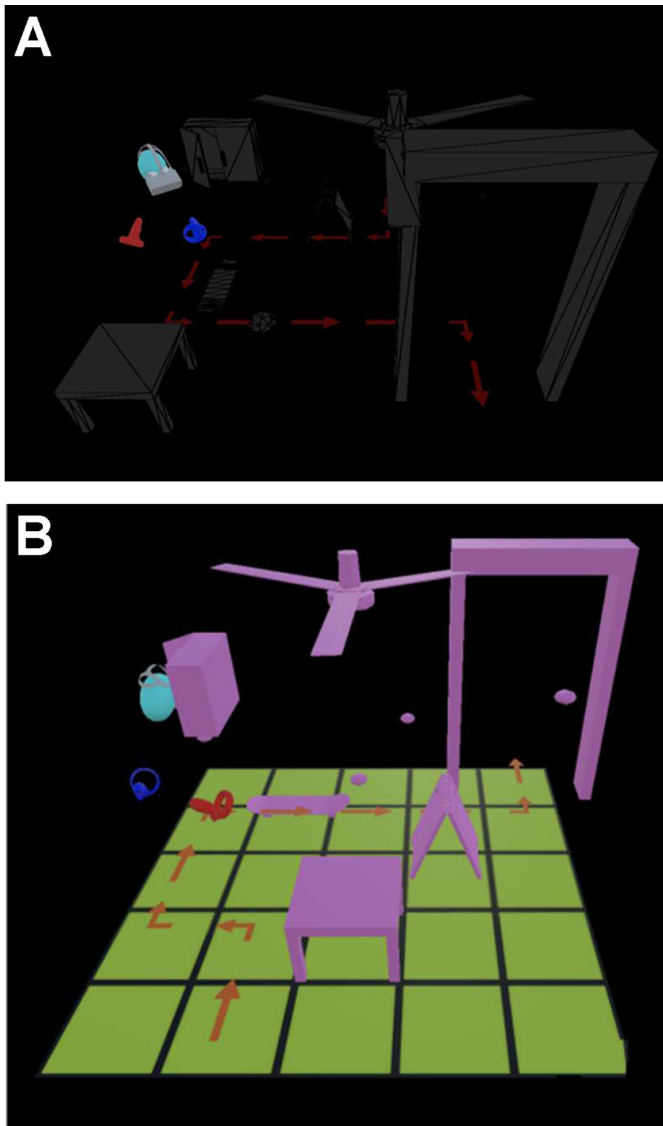
<sup>||</sup>GVF extent measured with size III-4e stimulus, expressed as the fraction of the normal extent for the size.

<sup>#</sup>VFQ maximal score = 250.

### Virtual Space and General Testing Configurations

A custom-built software was designed for use on a commercially available virtual reality hardware to simulate virtual scenarios and test the visual abilities of patients with vision loss from IRDs. Once the headset

and controllers were paired and the date and time were set (according to manufacturer's instructions), the headset was placed offline permanently. Only date and time stamps were used to label each data set (i.e. the data lacked identifiers) and no data were shared with



**Figure 1.** Virtual reality scenes. **(A, B)** Virtual reality scenery showing the avatar of the head of a participant wearing a headset (blue-green) and the hand-held controllers (blue and red). Path of arrows in red shown tracking into the exit threshold door. **A** Test with achromatic obstacles shown at low luminance levels (household objects, appear gray) in the dark-adapted state. **B** A participant taking a test under higher luminance conditions (with objects and tile floor colorized for display purposes) shown only during practice runs.

social media at any time. Details have been published.<sup>21</sup> Briefly, the virtual testing area consists of square tiles arranged in a rectangle 5 tiles wide and 5 tiles long (Fig. 1), occupying an area of  $2.3 \times 2.3$  m, centered by a black-patterned, orange polyhedron, used as a “starting platform.” The outer physical boundaries of the course are set at approximately  $3.9 \times 3.9$  m. This area occupies the center of a larger physical room ( $5 \times 5.2$  m), where the subject can move with enough room to avoid accidentally bumping into the physical walls. The

area also is close to a typical area available in clinical spaces. The position of the subject is tracked around 360 degrees.

The VR mobility course was initially designed to partially resemble the physical course assembled for the *RPE65-LCA* phase III gene therapy clinical trial whereby the subject follows a series of arrows, avoiding contact with obstacles in their path, as well as obstacles adjacent to the path and overhead. The test is delivered in two phases at increasing levels of “ambient luminance.” First, the software uses a “shrinking staircase method” to elicit the threshold luminance for detection of a track of relatively large (approximately 2 degrees of angular subtend) red arrows presented binocularly on a dark background that leads the subjects to an “exit door,” where the testing run ends (see Fig. 1). The arrows are displayed 3 seconds after the subject is guided to a large (approximately 5 degrees) orange starting platform. The starting platform “disappears” once the subject steps on it. The arrows disappear once the subjects cross the threshold of the “exit door” or ending point, and the platform is then shown at a different location or starting point. The arrows are randomly set in different path configurations (location of starting points, directions of arrows, and ending point), each with the same number of turns. Each individual arrow is large enough (approximately 4–6 degrees) so only one arrow is stepped on at a time, and so that they are visible at some luminance level even by subjects with low visual acuity (below the legal limit of blindness or 20/200 or logMAR 1.0). The system automatically recognizes and tabulates whether the subject follows the arrows or departs from the path and provides auditory and vibratory (through the handheld trackers) feedback. This “arrows only” first phase serves as a basic “orientation test” and to determine the lowest luminance at which the subject can track the path of arrows to the exit door without deviating from the path. If the subject cannot see the arrows or obstacles at a given luminance, they are instructed to hold both controllers above their head for 3 seconds, a movement which records the run as a fail and moves the testing to the next configuration and brighter light level.

Once the threshold for arrow detection has been measured, a second test phase evaluates the ability of the subject to navigate the path of arrows without colliding into obstacles, whereas the luminance of the arrows is kept invariant at approximately 1 log unit brighter than the “threshold” level measured in the “arrows only” phase. The test is delivered binocularly and then uni-ocularly to each eye separately by switching the contralateral VIR headset display off. There are 35 different course layouts that differ in the

configuration of the path of arrows and location of the objects, although each configuration has the same number of turns. There is a total of 16 different starting points. These are all located on the outer edge of the course so that the same number of turns and objects could be used in each configuration. Each course has a total of nine objects. The set of objects is identical in each run but the objects are placed at different locations or heights in each layout. The objects are of different sizes and include large (ceiling fan, table, and cabinet), medium (skateboard and wet floor sign), and small (spheres, including a swinging pendulum) objects (see Fig. 1), ranging from approximately 5 degrees to 20 degrees in angular subtend. Their height and vertical positions scale with the height of the subject so that low-level obstacles are always at foot/tracker level, mid-level obstacles are always at controller/hand level (subjects are instructed to hold controllers at their side), and high-level obstacles are always at or slightly above head level. The software tracks performance automatically and receives feedback from the subjects (see below).

### Test Administration and Scoring

Prior to embarking on the test, the subject navigates a practice course where both arrows and objects are displayed brightly illuminated (see Fig. 1B). During the practice test, the path of arrows as well as the obstacles are brightly illuminated so that the subject learns to identify them with both eyes open. The subject is taught to locate and step on the starting platform (“a red-orange orb”) as well as to identify the path of red arrows and the exit door through which they must pass. Once they pass through the door, the course disappears and a new orb (in a different location) appears. The subject again follows the path delineated by the arrows. Next, similar to the arrows-only test, the subject stands on the orb and the path of arrows (at the pre-set luminance – see above) as well as obstacles appear (see Fig. 1A). Subjects are warned that they might not be able to see some or all of the objects at the dimmest light levels or with one eye versus the other. The subjects are told to report their observations during the test but that the test-givers would not be able to comment except to assure that the system was functioning as planned or to remind the subject about details of the test. Test-givers record any comments along with the approximate time of the comment.

Once the subject is familiar with the test, the head-mounted VR headset is fitted under dim (red) illumination after 20 minutes of dark adaptation. During the dark adaptation period, patients with IRD provide verbal responses to the questions in the modified visual function questionnaire (see below). Those responses

are recorded by the test-giver. The initial test is the arrow-only course described above. Once the pre-set arrow course luminance is known, the second orientation and mobility phase can proceed. Our earlier VR-O&M design relied on automatic detection by the software of collisions with objects through trackers (handheld or at the ankles), resulting in scoring errors. To overcome this issue, the current design requires the subject to identify and “touch” each object by pushing a button on the controller (using either hand). When the subject touches an object and presses a controller button, the object disappears. The subject is given auditory feedback (a “positive” sound) if the object has been tagged successfully. The subject is given a different (“negative”) auditory feedback if an attempt was made to touch the object but if the subject did not locate it accurately. They can attempt to re-tag the object if they fail initially. The subject is reminded to look up, side-to-side, and down so that they do not miss an obstacle and to stay on the path while tagging each obstacle, avoiding a back-track. This is the expected behavior of a visually impaired patient as they walk into an unfamiliar place. The subject also receives auditory feedback when they step on the starting platform and when they exit each course. If the subject goes outside of the boundaries of the course, they receive tactile feedback (buzzing of the hand-held controllers).

The course is presented at step-wise increases of four standard luminance conditions ( $-0.67$ ,  $-0.19$ ,  $+0.39$ , and  $+0.69$  log phot.cd.m<sup>-2</sup>). The initial luminance step ( $-0.67$  log.cd.m<sup>-2</sup>) defines if additional attenuation of the virtual environment is needed as it always occurred in normal subjects who are always able to detect all objects at that luminance level. The addition of ND filters extends the operating range toward lower luminances with the addition of one 1.5 log unit ND filter in patients with IRD, or a sandwich of two ND filters for normal subjects, which brings the total of possible luminance steps from 4 to 8 or 12, with the addition of one or two ND filters, respectively. Each subject performs up to three runs per luminance level. The repeats per luminance level provides information regarding intra-session test-retest variability, and allows assessment for learning effects that may still exist after the initial training session. For each run, a different course configuration is used in order to minimize potential learning effect. After testing both eyes simultaneously, the obstacle test is then repeated through the entire luminance range for each eye individually.

### Scoring the Arrow and Obstacle Test Results

The software, generated using Unity (Unity Technologies, San Francisco, CA), automatically

recognizes and records numerous parameters, including the course configuration, details of the test paradigm and scene, height of the subject, luminance of arrows and objects, and all of the goggle and controller coordinates throughout the test. The data are embedded in csv files that are downloaded from the goggle headpiece after the session has been completed. These files can be opened at a later time with Unity software to “replay” the performance in each individual test on a desktop computer (Supplementary Video S1). Additional software, generated in Python, automatically tabulates the speed (amount of time necessary to complete the test) and accuracy (identification of each obstacle, departures from the path, direction of movement, collisions, and whether the subject missed any arrows or repeated them). This data set, including details of each individual test (arrow luminance, obstacle luminance, course configuration, time to complete each test, whether and when each of the different obstacles was tagged, etc.) is then encoded in a separate csv file for ease of statistical analyses.

## Visual Function Questionnaire

The 25-question visual function questionnaire (VFQ) was adapted for adult and pediatric subjects with inherited retinal disorders and developed to assess vision-dependent activities of daily living. The responses to each question utilize a numerical scale from 0 (worst performance/vision) to 10 (best performance/vision) such that a score reflecting “perfect” vision would be 250. A subset of questions (9) is mobility-specific and the remainder are visual acuity-specific.

## Results

### Subject Demographics

Participants ( $n = 49$ ) ranged in age from 14–62 years of age, including 20 normal subjects and 29 individuals with IRDs (see the [Table](#)). One of the objectives of the study was to assess the performance of the VR-O&M test over a wide range of functional deficits as estimated with common clinical measures, such as VA and GVF, independent of the clinical or molecular IRD diagnosis. Most patients had diseases with retina-wide, rod-greater than-cone photoreceptor dysfunction, or rod-cone dystrophies, including patients with phenotypes termed LCA, retinitis pigmentosa (RP), and choroideremia (CHM; see the [Table](#)). Most patients had molecularly confirmed disease. A third of the patients met the criteria of legal blindness (VA  $\leq 20/200$  or

$\geq 1.0$  Log MAR,  $< 20$  degrees visual fields), whereas several had VAs near 20/20 (LogMAR 0.0). Eight of 10 patients with *RPE65*-associated LCA (*RPE65*-LCA) had been treated bilaterally with voretigene neparvovec-rzyl (Luxturna). Absolute dark-adapted sensitivities by full-field sensitivity threshold (FST) testing were reduced by at least 1 log unit in each of these individuals and fields and/or acuities remained at the legal limit of blindness in 5 of them. Subjective functional vision estimated with the VFQ also sampled the spectrum of severity. About a third of the patients represented the upper and lower limit with four patients showing relatively high score values ( $> 200$ ), and five patients with low or very low scores values ( $< 100$ ).

## VR Mobility Test Performance

### Practice Session

Individuals were given as much time with the practice session as they needed to feel comfortable with the test and to minimize the potential interference of learning effects. The time required by different individuals to take the practice test varied between individuals (and previous experience with gaming devices), but the majority of individuals felt comfortable within 10 minutes. The time necessary to complete the practice test did not differ substantially between normal subjects and those with IRDs.

### VR Orientation Test (Arrows Only Test)

All normal subjects and all except two patients (VR70 and VR84) with the most severe vision loss, were able to orient themselves and locate the starting platform and trigger the start of the runs. Normal subjects could identify the path of arrows at 1.5 to 3.0 log units' attenuation compared to patients, who could only track the arrow path at the upper range of luminances ( $+0.11$  log phot-cd.m<sup>-2</sup>, median  $+0.15$  log phot-cd.m<sup>-2</sup>), without the attenuation with ND filters. Although most patients could orient themselves, individuals differed in the arrow luminances needed for them to be able to follow the path of arrows. Expectedly, subjects with normal vision performed well at arrow luminances of  $-0.67$  log phot-cd.m<sup>-2</sup> (without the ND filter attenuation), although somehow unexpectedly, some normal subjects favored brighter arrows at near maximum luminances when tested on the VR orientation (path of arrows only) test without ND attenuation. More than half of all patients also performed well at threshold arrow luminances of  $-0.67$  log phot-cd.m<sup>-2</sup>. Further, more than half of patients with LCA specifically performed at threshold arrow luminances  $\leq -0.22$  log phot-cd.m<sup>-2</sup>. In group analyses, there was no significant difference between the

arrow luminance thresholds of normal subjects tested without attenuation by neutral density filters versus patients. The average for normal subjects was  $-0.11$  log phot-cd.m<sup>-2</sup> and the median was  $-0.67$  log phot-cd.m<sup>-2</sup> and that for affected individuals was  $-0.27$  log phot-cd.m<sup>-2</sup> (with the same median of  $-0.67$  log phot-cd.m<sup>-2</sup>).

### Speed: The Time to Complete the Tests

Group analyses revealed no significant difference in the time to complete the arrow-only portion of the test between normal subjects and patients (Supplementary Fig. S1,  $P = 0.15$ ). However, there was a significant difference in the time to complete obstacle testing between those two groups (see Supplementary Fig. S1,  $P = 0.0027$ ). Normal subjects could complete the obstacle test (with the 4 different light settings) in approximately 3.3 minutes using both eyes and could complete the entire set of obstacles tests (using both eyes and each eye individually at all different light settings) in 11.8 minutes (range = 5.6–25.6 minutes). Normal subjects tested with the addition of the two ND still performed better than the patients. There was slightly more variability in patients when they used each eye individually than both eyes. On average, it took patients longer to complete the obstacle test than normal subjects (average time to complete the entire set of obstacle tests in patients was 22.2 minutes with a range of 4.5–54.9 minutes).

There was a wide range in the time necessary to complete the obstacle test at each of the different luminance levels between individuals in the affected group using both eyes (Fig. 2). There was only mild variation in normal subjects at the lowest ( $-0.67$  log phot-cd.m<sup>-2</sup>) luminance. Further, even if an affected individual could navigate the course quickly, that did not necessarily correlate with the ability of the individual to follow the path indicated by arrows. An example of the performance of an individual who navigated the course quickly but did not follow the path is shown in Supplementary Video S2. Some affected individuals followed the arrows and completed the course quickly because they did not see any obstacles (and thus did not spend the time to tag them). In summary, speed in itself does not necessarily reflect visual performance.

### VR Mobility and Orientation: Obstacle Detection

Figure 3 illustrates examples of the spectrum of severity of the abnormalities encountered. There was often a pigmentary retinopathy with variable extent and topographical distribution (see Fig. 3A). The retinopathy in VR64 exemplifies pericentral to midperipheral predilection with sparing the central retina (see Fig. 3A). On short-wavelength (SW) fundus

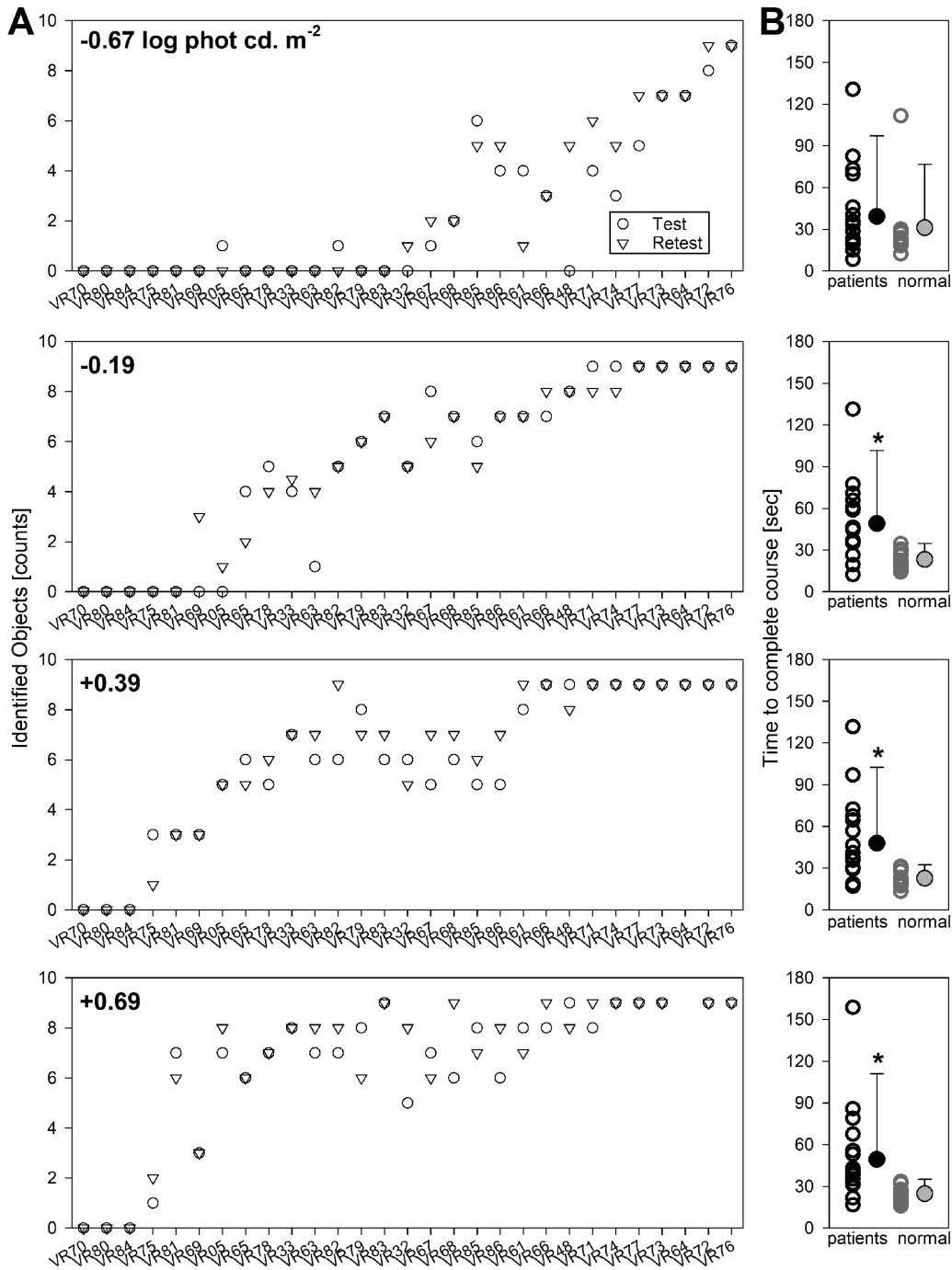
autofluorescence (FAF) there is an annulus of hypo-autofluorescence in the pericentral to midperipheral retina surrounding a better preserved center (see Fig. 3B). Consistent with this pattern, VAs and functional vision by VFQ were near normal (see the Table), although the patient's fields showed pericentral to midperipheral scotomas when measured with a large stimulus (Goldmann target size V-4e), and severely constricted fields when measured with a smaller I-4e target (see Fig. 3C). On VR-O&M, normal subjects were able to identify about half the total number of objects at the lowest luminances and showed a steep improvement in performance identifying nearly all objects (8/9 or 9/9 total objects) within approximately 1 log unit increment brightness from the dimmest level of luminance (see Fig. 3D). In contrast, and despite excellent acuities and relatively preserved visual fields to large targets, VR64 needed approximately 2 log units brighter object luminances compared to the average normal subject (see Fig. 3D, gray symbols = normal mean  $\pm$  3 SEM) to perform well and detect similar number of objects (see Fig. 3D).

VR74 showed a more extensive retinopathy with large areas of pigmentary changes and fundus hypo-autofluorescence extending from the pericentral to peripheral retina, but still sparing the central retina (see Figs. 3A, 3B). Accordingly, VAs and functional vision by VFQ are moderately abnormal (see the Table), although visual fields showed severe losses when measured with the large V-4e stimulus and were limited to a 5 degree tunnel visual field when measured with the smallest I-4e target (see Fig. 3C and the Table). On VR-O&M, VR74 can only perform like normal subjects at the highest two luminances ( $>2$  log units brighter luminances compared to normal; see Fig. 3D).

The third example (VR81) showed a retina-wide, diffuse retinopathy with retina-wide fundus hypo-autofluorescence (see Figs. 3A, 3B). VAs and functional vision by VFQ were moderately to severely reduced and visual fields were limited to a central and remnant temporal island of vision when measured with the large V-4e stimulus, limited to a 5 degree "tunnel" visual field with the smallest target (see Fig. 3C). On VR-O&M, VR81 can only detect objects at the highest two luminances and cannot detect all objects (see Fig. 3D). The result resembles the performance of normal subjects at the dimmest available luminances.

Figure 2A shows the number of obstacles (out of a total of 9) identified under binocular condition tested twice at each of 4 main luminance steps in all of 29 different patients. Two of them (VR70 and VR84) could not orient themselves and follow the path of arrows or identify objects and thus were not

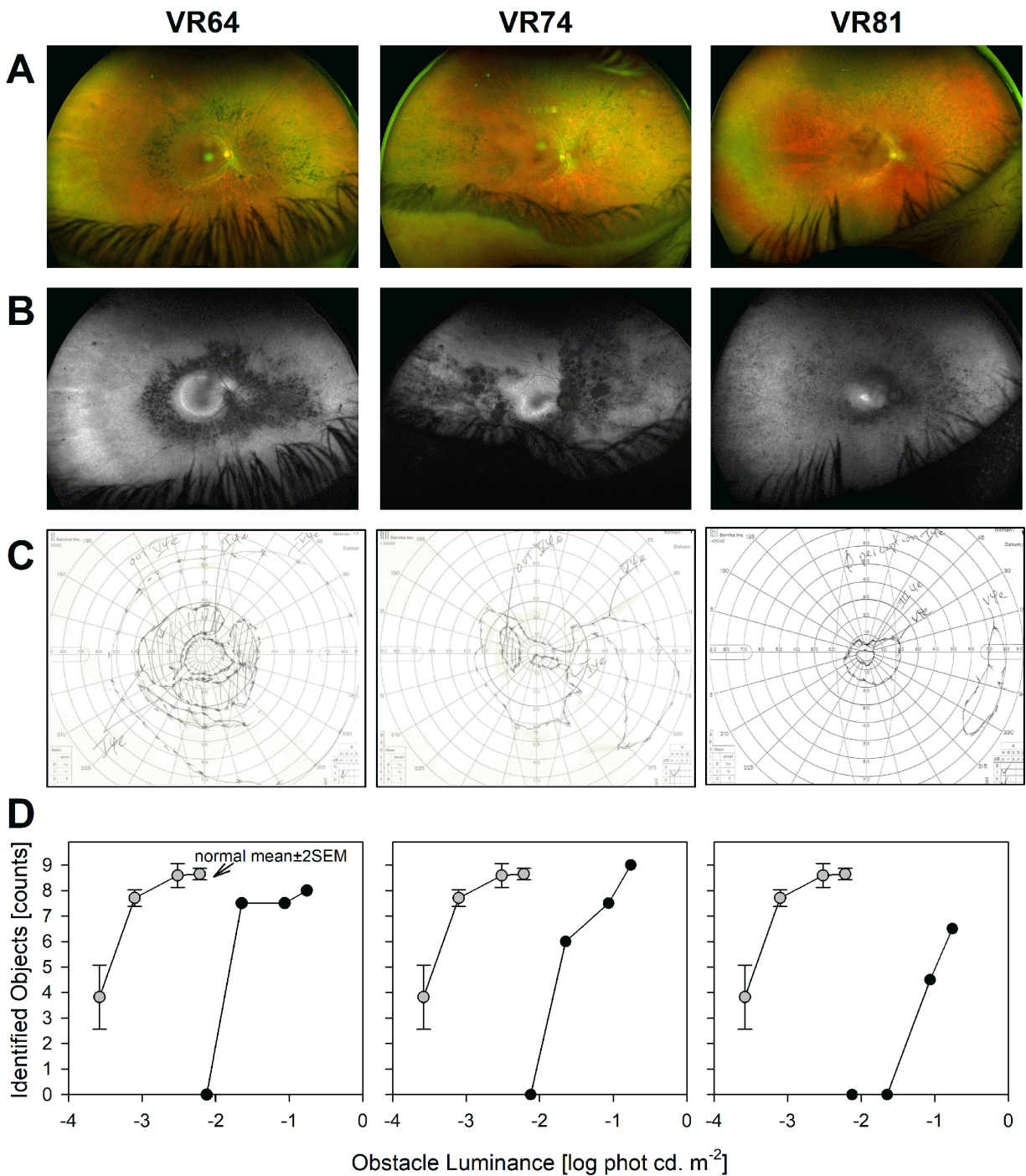




**Figure 2.** Performance on VR-O&M testing from all patients. **(A)** Counts of identified objects in all patients that underwent VR-O&M testing under binocular conditions. Pair of symbols in each patient represent test-retest evaluations. Panels represent the performance at each of four main luminance steps (without ND filters in place) ordered from dim (*top panel*) to brighter (*bottom*) objects luminances. Study IDs are shown on the horizontal axis ordered from poor (subjects to the left) to better performing (subjects to the right). **(B)** Time to complete the test at the lowest luminance condition in patients (*black outlined symbols*) compared to normal subjects (*grey outlined symbols*). Filled symbols represent mean +3 SD calculated from the individual data points to the left for patients (*black*) and normal subjects (*gray*). Asterisk represent statistically significant ( $P < 0.05$ ) difference between the distributions.

considered in the analyses of the overall performance on the object detection part of the test; VR 80 saw the arrows but could not identify any of the objects at any of the luminances. At the lowest tested object

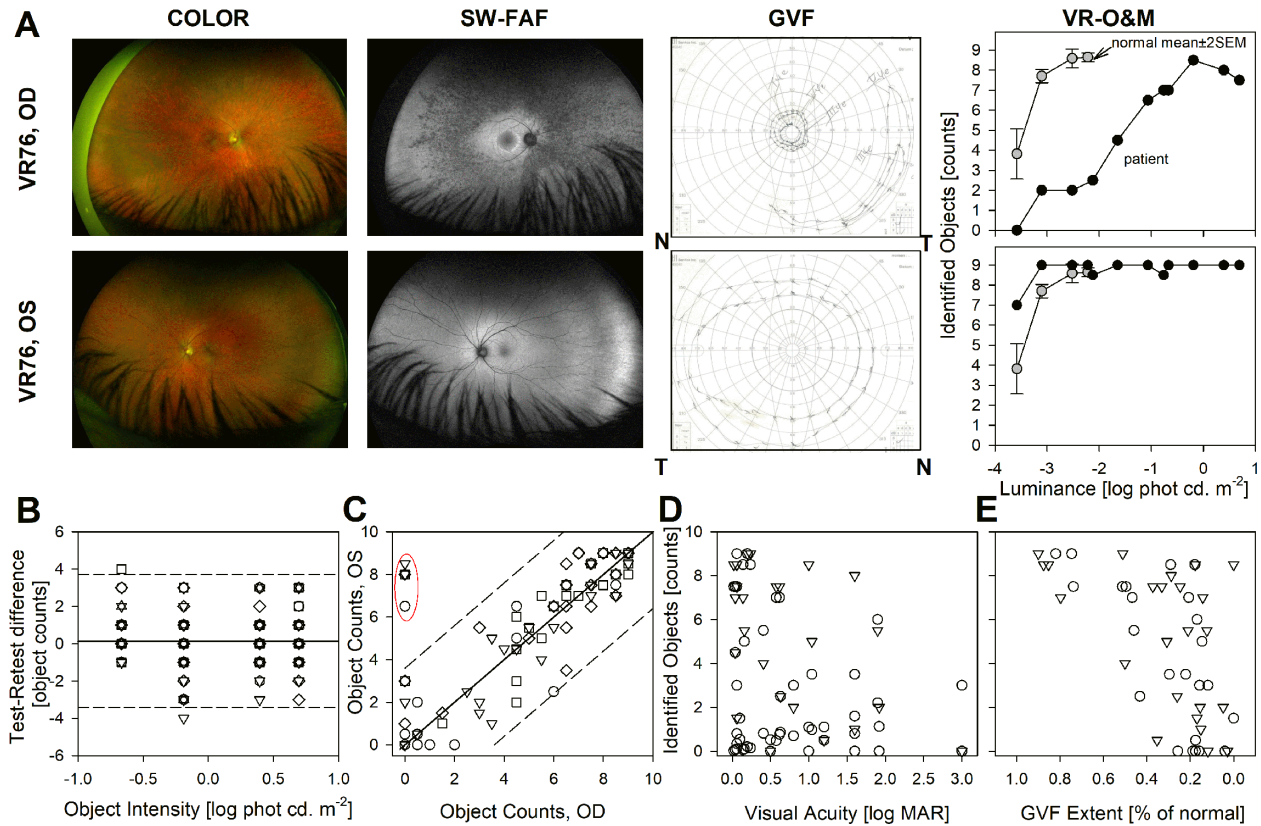
luminance ( $-0.67 \text{ log phot-cd.m}^{-2}$ ), normal subjects identified on average 97% of the objects (see Fig. 3D). In those rare cases where there were misses, the objects missed were globes (the smallest obstacle in the set)



**Figure 3.** Spectrum of severity of the structural and functional abnormalities in patients who underwent VR-O&M testing. **(A, B)** En face fundus appearance documented by color photography **A** and short-wavelength fundus autofluorescence **B**. **(C)** Kinetic Goldmann visual fields from each of the patients. The extent of the field is shown as a line or isopter, the size and intensity of the targets used to determine the visual field extent are penciled in (V4e and I4e) following conventional terminology; blind spots are hatched; VR81 could not perceive the smallest (I-4e) target and had an additional III-4e target size measured. **(D)** Number of objects identified by the patients (*black symbols*) plotted as a function of the luminance of the objects compared to the average performance of in normal subjects (*gray symbols*; error bars are  $\pm 2$  SD).

or the ceiling fan, both placed above eye level. All of the large and medium-sized objects were always tagged by normal subjects. In contrast, the majority of patients performed poorly at that luminance. Of

the 27 of 29 patients that could follow the path of arrows at the lowest luminance level, only 15 were able to identify some objects. They detected less than half the number of the objects as compared to normal



**Figure 4.** VR-O&M. Test-retest variability, interocular comparisons, and relationship with clinical measures of vision. **(A)** Color photography (first panel), short-wavelength fundus autofluorescence (second panel), kinetic visual fields (third panel), and VR-O&M performance (fourth panel) from each eye of a patient with interocular differences. The top row is the affected right eye, and the bottom row is the contralateral normal eye. VR-O&M data points in the patient (black symbols) are compared to the average object identification in normal subjects (gray symbols); error bars are  $\pm 2$  SEM. **(B)** Test Re-test difference in all patients as a function of object luminance. Solid line is the mean (visit 2–visit 1) Test – Retest difference; the dashed line represents  $\pm 2$  SD. **(C)** Comparison of the objects between the two eyes of each patient; the solid line represents the equality line; the dashed lines represent  $\pm 3$  SD of the interocular (the right eye minus the left eye) differences. **(D, E)** Counts of identified objects as a function of visual acuity **D** and visual field extent **E** for all patients for the second object luminance ( $-0.19$  phot. cd.m<sup>-2</sup>) for each, the right (circles) and left (triangles) eyes.

subjects (4/9, mean = 4.36 objects; see Fig. 3A, top panel). Performance improved on each of the next three steps of incremental luminances, with a greater proportion of patients (57%, 69%, and 77%), being able to identify an average greater number of objects (5.2, 6.2, and 7 objects) for the  $-0.19$ ,  $+0.39$ , and  $+0.69$  log phot-cd.m<sup>-2</sup> luminance steps, respectively (see Fig. 3A). Both small and large objects were missed as well as objects at all heights. The object that was most often tagged by affected individuals was the wet floor sign. At the lowest object luminance, patients, and normal subjects completed the VR-O&M test at a similar speed (patients mean  $\pm 2$  SD = 39 + 58 seconds; normal subjects = 31 + 46 seconds; see Fig. 3B). Patients were significantly slower at completing the course (49 + 53, 48 + 55, and 49 + 62 seconds) compared to normal subjects (23 + 12, 23 + 10, and 25 + 10 seconds,  $P < 0.05$ ) at greater luminance levels (see Fig. 3B).

As noted above, the lowest luminance within this set ( $-0.67$  log phot-cd.m<sup>-2</sup>) separated most patients from the behavior of normal subjects and was used to decide if further attenuation with the addition of neutral density filters was needed (i.e. when the performance was indistinguishable from normal subjects; see Fig. 3A). This occurred in two patients (VR72 and VR76), who performed normally under binocular conditions at this luminance step, and three others were near normal limits (VR64, VR73, and VR77). VR64 underwent further testing with the addition of ND filters to extend the operating range to lower luminances (one or two 1.5 log ND filters). At  $-2.12$  log phot-cd.m<sup>-2</sup>, the performance of VR64, a patient with EYS-arRP, was worse (detected 7 objects) than that of normal controls. VR76, on the other hand, presented a unique opportunity to test if the VR-O&M test could distinguish not only patients from normal subjects, but interocular differences in

dysfunction. This patient had a unilateral pigmentary retinopathy, or “unilateral RP” (Fig. 4A). His right eye (OD) had a subtle retina-wide pigmentary retinopathy that colocalized with hypo-autofluorescent lesions on SW-FAF, most obvious in the midperiphery, associated with a severely constricted fields and an island of vision in the temporal (T) peripheral field by GVF; his left eye (OS) was completely normal (see Fig. 4A). On VR-O&M testing, the patient performed similarly as the normal subjects, identifying 94% of objects with both eyes together, or with his normally sighted left eye (see Fig. 4A). In contrast, he performed poorly with affected right eye, identifying only 28% of the objects (Supplementary Video S2). Thus, VR testing was able to identify abnormal functional vision even if only one eye was affected.

We next asked if the VR-O&M test could produce similar results in each eye of each patient, who unlike VR76, had similar disease severity in each eye (see the Table). Specifically, we asked if the differences between the two eyes would exceed the variability of the test as determined by the test-retest differences in the detection of objects between two runs (run 2 minus run 1) of the test per intensity and per the eyes of each patient (see Fig. 4B). The test was reproducible with only minor improvement in object detection on the second run leading to minor positive test-retest differences (mean  $\pm$  3SD =  $+0.14 \pm 3.56$  objects,  $N = 183$  data points; see Fig. 4B). Note that there were slightly greater positive test-retest values at the dimmest luminance level suggestive of a minor learning effect (see Fig. 4B). The test-retest variability under binocular conditions ( $+0.22 \pm 3.16$  objects,  $N = 89$ , shown in Fig. 2A) was similar to the results under monocular conditions ( $t$ -test,  $P = 0.58$ ; see Fig. 4B). Normal subjects showed smaller but comparable ( $P = 0.23$ ) test-retest variability ( $+0.03 \pm 1.81$  objects,  $N = 132$  data points). Raw interocular differences (IODs) in object detection scores (OD minus OS, mean  $\pm$  3SD =  $-0.51 \pm 6.54$  objects) exceeded the test-retest variability ( $P = 0.001$ ), reflecting patients with interocular asymmetries in function (VR71, VR76, and VR83) (see Fig. 4C, outlined in red). Excluding these outliers from the analysis led to interocular differences ( $-0.08 \pm 3.63$  objects) that were comparable to the test-retest variability ( $P = 0.16$ ). Testing under binocular conditions improved performance. Patients who could not identify objects at certain luminances under unocular conditions were able to identify some objects binocularly. In addition, and although there were no major interocular differences in most patients, the binocular performance corresponded to the better performing eye as was illustrated in the extreme case for the patient with unilateral disease (see Fig. 4A). Last, the results of the VR-

O&M test related well with the results of basic clinical visual function tests. Patients with better visual acuities and larger visual field extents tended to perform better (see Figs. 4D, 4E). At the lowest 2 intensities, eyes with visual acuities at 20/40 or better detected on average 7 objects compared to patients with visual acuities at or below the legal limit of 20/200 who scored on average 2.2 objects (see Fig. 4D). Similarly, patients with visual fields extending over 70% of the normal visual field extent along the horizontal meridian scored on average twice the objects (8 objects) compared to those under 30% of the field (4 objects; see Fig. 4E).

### Visual Function Questionnaire Results

Total scores (maximum possible score of 250) and scores of questions that were visual acuity-related versus navigation-related (maximum possible scores of 160 and 90, respectively) are provided in Supplementary Table S1. There was a significant large positive relationship between performance at low light levels using both eyes on the VR test and the navigation score on the VFQ test (Pearson correlation coefficient = 0.59,  $P = 0.013$ ). The two outliers were VR74 and VR65, who self-assigned high scores on the VFQ (total of 211 with 79 for navigation and total of 235 with 85 for navigation, respectively) but tagged only 44% or 0% (respectively) of the objects at the low luminance on the standard VR test.

The average score of the affected individuals (excluding those with LP vision) was 157, the median was 164 and the range was 72 to 247. The average score for visual acuity-related questions of the affected individuals (excluding those with LP vision) was 106.12. The median score was 113. The range was 46 to 160 (out of a total possible of 160). The average score for navigation-specific questions of the affected individuals (excluding those with LP vision) was 50.94, the median was 49 and the range was 23 to 87 (out of a total possible of 90). The two individuals with LP vision (VR70 and VR84) had the lowest total scores (16 and 30 out of a total possible of 250) of all of the affected individuals. Their scores to visual acuity-related questions were low (3 and 20, respectively, out of a total possible of 120). Their scores to navigation-specific questions were also particularly low (13 and 10, respectively, out of a total possible of 90).

## Discussion

Here, we show that a relatively short (<1 hour including practice session and 20-minute dark adaptation) VR-O&M test can distinguish the majority of

individuals with IRDs included in this study from normal subjects. We had previously designed two sequential test paradigms with the first (“arrows only” sequence) fulfilling the role of an orientation task. A path of relatively large red arrows was chosen to bias detection by most central cones, minimize interference with rod-mediated vision, and allow perception by patients with low visual acuity. Once a “threshold” for successful tracking of the arrows was estimated, then the subjects were exposed to “arrow-plus-obstacle” tests, the actual navigation test. Note that we tried our best to simplify the test so that visual clues would be relatively simple and invariant across the range of luminances, facilitating the interpretation in terms of their relationship to clinically test, such as kinetic visual fields. The obstacles are simple in shape and achromatic, and changes in contrast likely minimal as increases in luminance of the “objects” and background occur simultaneously, reducing the complexity of the psychophysical mechanisms involved in their detection (mainly, luminance, size, and position). Objects are placed on, over, or adjacent to the path of arrows. The subjects have to simultaneously follow the arrows with their most central and sensitive vision, and detect obstacles located more peripherally in their field of vision. In the previous paradigm, these obstacles were to be avoided. The automatic scoring of collisions depended on the proper spatial registration of sensors attached to the subjects’ ankles or hand-held in relationship to the virtual space, which led to scoring errors. In the current paradigm, the subject is asked to “tag” the objects with their hand-held controller. This act confirms that the subject has seen the object and also allows the software to automatically register the results with less room for errors. The requirement to follow the path of arrows while searching for obstacles orients the subjects’ visual field to the scene similar to what someone might do while walking down a sidewalk or a path, while having to be aware of the overall scenery. The use of 35 different course templates minimized the potential of a learning effect. All normal-sighted controls carried out the screening test quickly and accurately. Both children (as young as 14 years old) and adults found it easy to learn how to take the test. Patients appreciated seeing obstacles that represented some of the challenges they encounter daily in navigating in the “real world.”

Although speed is an important measure of the individual’s ability to carry out the VR visual task, we found that accuracy was more important. Thus, by focusing on the number of obstacles that an individual could identify under different luminance conditions (while following the path delineated by arrows), we could rapidly assess the individual’s functional vision.

In fact, the vast majority of the patients could be distinguished from normal subjects by their performance when they used both eyes to navigate under the dimmest light level of the VR system without the use of additional ND filters, a setting that could be used as a single-luminance, abbreviated screening test, or as a starting point to rapidly screen for the degree of abnormalities, and then configure the VR-O&M system to only test an informative range of luminances, as was adopted in this study. In such a protocol, subjects who fail to identify objects correctly at that luminance level can proceed to higher luminances, whereas subjects that show a performance close to normal subjects, would then be tested over the lower range of intensities, as demonstrated for the normal subjects and two less affected individuals in this study. This approach promises to reduce testing time and frustration, especially from more severely affected individuals, variables often associated with poor performance and reproducibility in other psychophysics tests used in the clinic, such as visual fields.

In the current work, we formally assessed the short-term variability of the results by repeating the test for each luminance level in both binocular and monocular conditions. We found that the test was highly reproducible, with only slightly greater test-retest differences at the lowest luminance levels tested in both patients and normal subjects. Interestingly, patients were as reproducible as normal subjects when tested at the higher luminances. Clinical trials for IRDs often involve the delivery of therapeutic agents to one eye, using the contralateral eye as control. This trial design depends on interocular comparisons as a simplified way of testing safety and efficacy. In this study, interocular differences in subjects with similar disease severity in each eye were not greater than the short-term variability of the test. Further, the test was able to distinguish the abnormal eye of a subject with a mostly uni-ocular disease, suggesting this VR-O&M may be able to provide eye-specific outcome measures of functional vision.

Commonly used clinical measures of vision related well with the performance in the VR-O&M test, as well as with the patients’ self-assessed visual function as measured by the VFQ. Both visual acuity and visual field extent measured by kinetic perimetry related well with the patient’s performance on the VR-O&M course. The data obtained in this study also demonstrates that the VR-O&M test can easily accommodate a wide spectrum of disease severity encountered in patients with IRDs partially represented in this cohort. Notably, an individual with advanced choroideremia and cone-only vision was able to carry out the test

(even though he performed poorly on this test). In a validation study of the MLMT (the test used as the primary outcome measure in a phase III study of gene therapy for RPE65 deficiency) subjects with choroideremia passed the MLMT without any difficulty. This suggests the MMLT test may be insensitive in subjects with relatively preserved central vision, particularly visual acuity, as is often the case in subjects with choroideremia.<sup>23</sup> This preliminary observation indicates the VR-O&M test may be more sensitive than the MLMT test in a wider range of phenotypes, including patients with severe visual field constrictions, but relatively preserved small central islands of much better preserved vision (by sensitivity and/or visual acuity), a frequent scenario in inherited retinal degeneration. Formal direct comparisons between physical orientation and mobility tests in longitudinal studies are warranted to determine the pros and cons of each methodology.

Although patients with severe retinal degenerations were represented in this work, the VR-O&M system has yet to be challenged by testing patients with the most severe forms of these diseases, some of which are being treated or are in the planning stages of gene therapy clinical trials. We anticipate the upper range of luminances may need to be extended to accommodate greater losses of vision. There are numerous advantages of the VR platform compared to more commonly used physical navigation courses including ease of administration and rapid automatic unbiased scoring. It would be relatively easy to modify the visual attributes of both objects and surrounding scenery (for example, shape, size, height, contrast, color, luminance, and textures), in order to focus on specific variables affecting functional vision in specific IRDs. Additional modifications can be made to be able to include individuals with LP vision, such as incorporation of even brighter and larger obstacles. There may be a technical limit in currently available VR systems, although the technology to allow for these type of improvements likely exists. We also anticipate that there may be a limit to the subject's ability to orient themselves and move guided by vestiges of classical and even non-classical photoreception in end-stage retinal degenerations. Even in that scenario, the VR-O&M test will still serve to provide a qualitative and quantitative baseline functional vision outcome that may be tested after the delivery of therapies intended to restore some vision. On the other side of the spectrum, the current work proved that the impact of milder levels of visual dysfunction on functional vision can be detected with the current system. The observations, however, are limited in number. Further studies are thus needed to

test the performance of the test in greater number of patients, especially at the mildest and most severe ends of the spectrum, including children.

The VR course was designed to encompass a range (limited) of luminances from the high scotopic range to low photopic levels. Although we tested the system in a patient with a cone-rod dystrophy, evaluation of inherited retinal diseases with other mechanisms of dysfunction, such as severe cone dysfunctions in the spectrum of achromatopsias, and severe central diseases with preservation of peripheral retinal function, may require refinement of the current test or different set ups altogether, which awaits further development. Comparison of VR-O&M performance against more precise measures of vision, such as the topography of the sensitivity losses for photoreceptor-subtypes (rods versus cones), is needed to further determine the underlying mechanisms driving performance, better configure the test to specific phenotypes, as well as identify the limits of the test as a detector of relevant changes in functional vision. The steep functions that related object detection against luminance suggest smaller incremental steps are needed near the threshold for detection of objects in both patients and normal subjects, which cannot be done without proper optimization to avoid making the testing algorithm impractical. The current test design, however, promises to be useful in both natural history studies of disease progression and to assess any potential therapeutic effect in upcoming clinical trials in a large number of patients with IRDs.

In summary, the present work supports the utility of a VR-O&M test as a measure of functional vision in patients within a wide spectrum of disease severity, and, modestly, across a limited number of phenotypes. The equipment is easy to install and to wear, is comfortable and presents minimal challenge in teaching even young subjects how to use it. The test is sensitive and is able to objectively and rapidly differentiate patients with vision impairments from normally-sighted subjects. It is much easier to configure, deliver, and is perhaps more sensitive than the MMLT, currently accepted as the gold-standard for physical orientation and mobility test for inherited retinal degenerations. The next step will be to further validate this test focusing on the reproducibility and variability of the measures in the long-term (weeks to months) as well as on exploring the determinants of the functional vision performance by comparing with other measures of vision. This novel virtual reality test of functional vision promises to be a useful outcome measure for quantifying the impact of IRD disease and treatments thereof.

## Acknowledgments

The authors are deeply grateful to the families and patients for their participation in this study. The authors also extend thanks to Jules LaRosa and Shelby Brizzolara-Dove for accommodating some of the subjects in this study and to Nancy Bennett for invaluable input.

Supported by the F.M. Kirby Foundation and Center for Advanced Retinal and Ocular Therapeutics (CAROT), University of Pennsylvania.

**Declaration of Interest:** T.S.A., E.M.A., K.H.M., W.M.M., A.J.M., and J.B. are co-authors on intellectual property describing the virtual reality mobility test.

**Disclosure:** J. Bennett, None; E.M. Aleman, None; K.H. Maguire, None; J. Nadelmann, None; M.L. Weber, None; W.M. Maguire, None; A. Maja, None; E.C. O'Neil, None; A.M. Maguire, None; A.J. Miller, None; T.S. Aleman, None

## References

- Bennett CR, Bex PJ, Bauer CM, Merabet LB. The Assessment of Visual Function and Functional Vision. *Semin Pediatr Neurol.* 2019;31:30–40.
- Shingledecker CA, Foulke E. A human factors approach to the assessment of the mobility of blind pedestrians. *Hum Factors.* 1978;20:273–286.
- Marron JA, Bailey IL. Visual factors and orientation-mobility performance. *Am J Optom Physiol Opt.* 1982;59:413–426.
- Black A, Lovie-Kitchin JE, Woods RL, Arnold N, Byrnes J, Murrish J. Mobility performance with retinitis pigmentosa. *Clin Exp Optom.* 1997;80:1–12.
- Apfelbaum H, Pelah A, Peli E. Heading assessment by “tunnel vision” patients and control subjects standing or walking in a virtual reality environment. *ACM Trans Appl Percept.* 2007; 4:8.
- Kolarik AJ, Cirstea S, Pardhan S, Moore BC. A summary of research investigating echolocation abilities of blind and sighted humans. *Hear Res.* 2014;310:60–68.
- Peli E, Apfelbaum H, Berson EL, Goldstein RB. The risk of pedestrian collisions with peripheral visual field loss. *J Vis.* 2016;16:5.
- Finger RP, Ayton LN, Deverell L, et al. Developing a Very Low Vision Orientation and Mobility Test Battery (O&M-VLV). *Optom Vis Sci.* 2016;93:1127–1136.
- Shapiro A, Corcoran P, Sundstrom C, et al. Development and Validation of a Portable Visual Navigation Challenge for Assessment of Retinal Disease in Multi-Centered Clinical Trials. *Invest Ophthalmol Vis Sci.* 2017;58:3290. Presented as an abstract at the 2017 ARCO Annual Meeting, Baltimore, MD, May 7–11, 2017.
- Chang KJ, Dillon LL, Deverell L, Boon MY, Keay L. Orientation and mobility outcome measures. *Clin Exp Optom.* 2020;103(4):434–448.
- Mees L, Upadhyaya S, Kumar P, et al. Validation of a Head-mounted Virtual Reality Visual Field Screening Device. *J Glaucoma.* 2020;29:86–91.
- Velikay-Parel M, Ivastinovic D, Koch M, et al. Repeated mobility testing for later artificial visual function evaluation. *J Neural Eng.* 2007;4:S102–S107.
- Nau AC, Pintar C, Arnoldussen A, Fisher C. Acquisition of Visual Perception in Blind Adults Using the BrainPort Artificial Vision Device. *Am J Occup Ther.* 2015;69:6901290010p6901290011–6901290069.
- Geruschat DR, Richards TP, Arditi A, et al. An analysis of observer-rated functional vision in patients implanted with the Argus II Retinal Prosthesis System at three years. *Clin Exp Optom.* 2016;99:227–232.
- Bainbridge JW, Smith AJ, Barker SS, et al. Effect of gene therapy on visual function in Leber's congenital amaurosis. *N Engl J Med.* 2008;358:2231–2239.
- Jacobson SG, Cideciyan AV, Sumaroka A, et al. Outcome Measures for Clinical Trials of Leber Congenital Amaurosis Caused by the Intronic Mutation in the CEP290 Gene. *Invest Ophthalmol Vis Sci.* 2017;58:2609–2622.
- Chung DC, McCague S, Yu ZF, et al. Novel mobility test to assess functional vision in patients with inherited retinal dystrophies. *Clin Exp Ophthalmol.* 2018;46:247–259.
- Russell S, Bennett J, Wellman J, et al. Year 2 results for a phase 3 trial of voretigene neparvovec in biallelic RPE65-mediated inherited retinal disease. Baltimore, MD: ARVO; 2017;58(8). Presented as an abstract at the 2017 ARCO Annual Meeting, Baltimore, MD, May 7–11, 2017.
- Russell SR, Drack AV, Cideciyan AV, et al. Intravitreal antisense oligonucleotide seposfarsen in Leber congenital amaurosis type 10: a phase 1b/2 trial. *Nat Med.* 2022;28:1014–1021.

20. Sahel JA, Grieve K, Pagot C, et al. Assessing Photoreceptor Status in Retinal Dystrophies: From High-Resolution Imaging to Functional Vision. *Am J Ophthalmol.* 2021;230:12–47.
21. Aleman TS, Miller AJ, Maguire KH, et al. A Virtual Reality Orientation and Mobility Test for Inherited Retinal Degenerations: Testing a Proof-of-Concept After Gene Therapy. *Clin Ophthalmol.* 2021;15:939–952.
22. Chang Y-L, Lu Z-H. White Organic Light-Emitting Diodes for Solid-State Lighting. *J Display Technol.* 2013;9:459–468.
23. Chung DC, Bertelsen M, Lorenz B, et al. The Natural History of Inherited Retinal Dystrophy Due to Biallelic Mutations in the RPE65 Gene. *Am J Ophthalmol.* 2019;199:58–70.