tvst

Retina

Addressing Multiplicity in Retinal Sensitivity Analysis: An Alternative Approach to Assessing Gene Therapy Efficacy in Inherited Retinal Diseases

Antonio Yaghy¹, David G. Birch², Yunchan Hwang³, Edmund Luo¹, JungAh Jung¹, Darin Curtiss¹, and Nadia K. Waheed¹

¹ Beacon Therapeutics, Alachua, FL, USA

² Retina Foundation of the Southwest, Dallas, TX, USA

³ Department of Electrical Engineering and Computer Science, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA

Correspondence: Nadia K. Waheed, MPH New England Eye Center, Tufts Medical Center, 260 Tremont St., Boston, MA 02116, USA. e-mail: nadiakwaheed@gmail.com

Received: May 15, 2024 Accepted: January 27, 2025 Published: March 27, 2025

Keywords: gene therapy; inherited retinal disease (IRD); microperimetry; retinal sensitivity; statistical analysis; multiplicity control; x-linked retinitis pigmentosa (XLRP); clinical trial

Citation: Yaghy A, Birch DG, Hwang Y, Luo E, Jung J, Curtiss D, Waheed NK. Addressing multiplicity in retinal sensitivity analysis: An alternative approach to assessing gene therapy efficacy in inherited retinal diseases. Transl Vis Sci Technol. 2025;14(3):25, https://doi.org/10.1167/tvst.14.3.25 **Purpose:** The purpose of this study was to propose an alternative statistical approach that addresses the issue of multiplicity in microperimetry data analysis, offering a more balanced and sensitive measure of the efficacy of gene therapy in inherited retinal diseases (IRDs).

Methods: We analyzed microperimetry data from a phase II trial of AGTC-501 in patients with X-linked retinitis pigmentosa (XLRP). The Macular Integrity Assessment (MAIA; CenterVue, Padova, Italy) device was used to evaluate test-retest repeatability. A binomial model was used to calculate the probability of \geq 7 decibel (dB) improvements due to chance alone across a 68-locus grid. We proposed an alternative approach to detect changes using a threshold of \geq 7 loci with \geq 7 dB mean improvement.

Results: Test-retest repeatability analysis showed a probability of < 5% for observing pointwise improvements ≥ 7 dB between 2 baseline visits. Applying the binomial distribution model, we found that the probability of observing improvements ≥ 7 dB in at least 7 unspecified loci purely by chance was 5.3%.

Conclusions: The proposed approach provides a balanced way to address multiplicity while maintaining reasonable statistical significance. Using \geq 7 unspecified loci as the criterion for assessing sensitivity changes, offers a comprehensive assessment that can detect genuine treatment effects without being overly conservative.

Translational Relevance: This alternative statistical method has the potential to improve the evaluation of retinal sensitivity changes in gene therapy trials for IRDs, providing a more accurate measure of therapeutic efficacy and enhancing clinical decision making.

Introduction

Microperimetry is an imaging technique that measures the sensitivity of the retina to light stimuli in different locations of the visual field.¹ It is a useful tool for assessing the functional outcomes of gene therapy for inherited retinal disease (IRD), a group of disorders that cause progressive vision loss due to mutations in genes involved in photoreceptor function or survival, including retinitis pigmentosa and Leber congenital amaurosis.^{2,3} The ideal outcome measure would identify changes in visual function due to novel therapeutic interventions, such as gene therapy, without being influenced by the natural patient variability.⁴

However, evaluating the efficacy of gene therapy for IRD poses several challenges, one of which is the issue of multiplicity. Multiplicity refers to the increased chance of observing a false positive (risk for type I errors) when multiple tests or comparisons

Copyright 2025 The Authors tvst.arvojournals.org | ISSN: 2164-2591



are made.⁵ Previously, the US Food and Drug Administration (FDA) recommended that for clinical trials using standard automated perimetry, like those in glaucoma studies, a between-group mean difference between the treated and untread cohorts of at least 7.0 decibel (dB) in mean sensitivity change for the entire field is considered clinically significant.⁶ This guideline reflected the FDA's traditional reliance on functional measures, particularly visual field testing, as primary end points in glaucoma clinical trials. However, as discussed at the 2010 NEI/FDA Glaucoma Clinical Trial Design and Endpoints Symposium, the FDA has shown openness to considering new end points, including structural measures, provided they demonstrate a strong correlation with clinically relevant functional outcomes and are validated by the research community.⁶ Lately, the FDA has also indicated that a positive outcome should be based on the mean improvement of at least 7 dB from baseline in at least 5 prespecified loci within the central 30 degrees of the visual field, and this improvement should be sustained over time (FDA Clinical IR for IND 17634, October 30, 2020). However, as demonstrated in the XIRIUS phase II/III study on Cotoretigene Toliparvovec, achieving this outcome can be challenging.⁷ The study failed to meet the primary end point, with no significant difference in the percentage of participants meeting the responder criteria between the treatment and control groups.⁷ This outcome highlights the difficulties inherent in prespecifying loci, such as the variability in patient responses and the potential for high false positive rates. Moreover, selecting specific loci prior to treatment might overlook broader improvements in retinal sensitivity outside five prespecified loci, thereby underestimating the therapeutic effect. These challenges underscore the need for alternative approaches that balance multiplicity control with the ability to detect genuine treatment effects across the entire visual field.

Herein, we propose an alternative approach that does not require prespecifying loci, but instead uses a statistical method that adjusts for multiplicity. We demonstrate that this method can reduce the false positive rate and increase the power of detecting a true treatment effect. We also discuss the clinical relevance and implications of this method for patients with IRDs who undergo gene therapy.

Understanding the FDA's Value Derivation

The FDA considers a positive outcome when a mean improvement of at least 7 dB from baseline in at least 5 prespecified loci within the central 30 degrees of the visual field is sustained over time. To address the FDA's concern regarding the multiplicity issue in defining responders for the visual field efficacy end point, we conducted analyses using microperimetry data collected from patients enrolled in a randomized, dose masked phase II dose expansion study whose primary objective is to assess the efficacy, safety, and tolerability of high dose and low dose of AGTC-501 (rAAV2tYF-GRK1-RPGR) administered through a subretinal injection to patients with X-linked retinitis pigmentosa (XLRP).8 Specifically, we assessed the test-retest repeatability of pointwise sensitivity using the Macular Integrity Assessment (MAIA; CenterVue,



Figure 1. Bland-Altman plot of pointwise sensitivity between screening visits two and three in the study (A) and fellow eyes (B).

New Approach to Addressing Multiplicity in IRDs

Padova, Italy) device during the screening two and screening three visits. There was a 4.5% probability that a repeat (screening 3) pointwise test results exceeded the first test (screening 2) by \pm 7 dB in the study eye (Fig. 1A). Similarly, there was a 3.5% probability that a repeat (screening 3) pointwise test results exceeded the first test (screening 2) by \pm 7 dB in the fellow eye (Fig. 1B). In other words, there is a < 5% probability chance that, for any point tested, more than \pm 7 dB difference in between follow-up visits is observed.

Addressing Multiplicity in Retinal Sensitivity Analysis: The Need for an Alternative Approach

In a 68-locus grid used in MAIA microperimetry to assess retinal sensitivity, each locus can either show an improvement greater than or equal to 7 dB or not (Figs. 2A, 2B). The probability of observing this change purely due to variability (and not an actual treatment effect) is 5% for each locus, assuming an a = 0.05. If the loci are not prespecified, the probability of observing one locus with an improvement of \geq 7 dB purely by chance is quite high. Because only improvements of \geq 7 dB are considered clinically meaningful, and the probability of observing such a large improvement by chance is likely less than 5%, we use 5% as

Baseline

A

a conservative upper bound in our calculations. Using
the concept of binomial distribution, which describes
the number of successes in a fixed number of indepen-
dent Bernoulli trials, we can calculate the probability of
observing at least one such change in a 68-locus grid.⁹
A Bernoulli trial is a random experiment with exactly
two possible outcomes: "success" and "failure." In this
context, a "success" is observing an improvement of
$$\geq$$
7 dB purely due to variability at a specific locus. The
binomial distribution allows us to calculate the proba-
bility of observing a specific number of successes (loci
with an improvement of \geq 7 dB) in the n-locus grid:

$$Prob\left(X=k\right) = \binom{n}{k} p^{k} (1-p)^{n-k}$$

where:

В

- X is the random number of total number of successes
- k is the observed number of successes
- n is the number of loci
- p is the probability of success for one locus.

when n = 68 (grid of 68 loci), P = 0.05 (a = 0.05),

Prob (X
$$\ge$$
 1) = 1 - Prob (X = 0)
= 1 - (0.95)⁶⁸ \approx 0.97

This means that there is a 97% probability of observing at least one locus with an improvement of

Month 3



Figure 2. Comparison of the FDA's method and our proposed method for assessing retinal sensitivity improvement. (**A**) Baseline 68-loci microperimetry grid with 5 randomly prespecified loci (*orange squares*) selected for analysis per the FDA's method (which specifies the selection of at least any 5 loci). (**B**) At month 3, the 68-loci microperimetry grid showing loci with an improvement of at least 7 dB from baseline. The *green circles* indicate loci meeting the \geq 7 dB improvement threshold, whereas the *gray circles* represent loci not meeting the threshold. Using the FDA's method, 4 out of the 5 prespecified loci achieved the \geq 7 dB improvement, which would classify this eye as a nonresponder. In contrast, our proposed method identified 10 nonspecified loci with \geq 7 dB improvement, classifying this eye as a responder.

 \geq 7 dB purely by chance when the loci are not prespecified albeit the probability of observing an improvement of \geq 7 dB by chance at each locus is only 5%. This is a typical example of multiplicity.

To mitigate this issue, the FDA requires the use of 5 prespecified loci out of 68 to ensure that any mean improvement of at least 7 dB is a true positive and not due to chance. The probability of making 5 consecutive false positive claims in prespecified loci is 0.00003125% (0.05^5), which is 160,000 times smaller than the historically used a value of 0.05, making the FDA's approach extremely conservative and stringent.

Note that in our analysis of retinal sensitivity changes using a 68-locus grid in MAIA microperimetry, we made 2 key assumptions to apply a binomial distribution model. First, we assumed that the probability of observing an improvement in sensitivity of \geq 7 dB is identical across all 68 loci. Although there may be some biological variation in the potential for improvement across different retinal regions, assuming a uniform probability is a reasonable simplification that enables the use of the binomial model. Second, we assumed that each measurement across the 68 loci is independent. This means that the occurrence of a sensitivity change in one locus is presumed not to affect the probability of observing a change in any other locus. Although there might be some spatial correlation between neighboring loci, treating the measurements as independent is a justifiable approximation given the sparse sampling across the retina. These assumptions, whereas not perfect, allow the binomial model to provide valuable insights into the likelihood of observing improvements across multiple loci by chance alone. Any deviations from these assumptions are likely to have a limited impact on the overall conclusions, as the binomial model serves as a useful approximation for assessing the statistical significance of observed multi-locus improvements.

Proposing an Alternative Approach

Whereas the FDA's approach of prespecifying five loci effectively controls for multiplicity, it is overly conservative. An alternative approach is to consider the number of loci showing an improvement of ≥ 7 dB in the entire 68-locus grid, without prespecification.

To compute the probability Prob ($X \ge 5$), we use the binomial distribution equation above to sum the probability of getting 5, 6, 7... up to 68 successes.

$$Prob (X \ge 5) = Prob (X = 5) + Prob (X = 6)$$
$$+ Prob (X = 7) \dots + Prob (X = 68)$$

Using SAS version 9.4 statistical software, we can compute *Prob* ($X \ge 1$) through *Prob* ($X \ge 10$; Fig. 3).

As can be seen from the sigmoid curve in Figure 3, the *Prob* ($X \ge k$) decreases as the number of successes (k) increases. For instance, the *Prob* ($X \ge 5$) is 25%, the *Prob* ($X \ge 6$) is 12.4%, and the *Prob* ($X \ge 7$) is 5.3%. In other words, if 7 unspecified loci with a mean improvement of at least 7 dB from baseline were used to define



Figure 3. Line plot showing the probability of observing \geq k loci with an improvement of \geq 7 dB.

Tabla	Commentions of Ctatisti	مما ۸ به به ممام ۸ ا	fay M	Coursetion in	Dational Cam	~:+:. /:+. / A	mal vaia
laple	Comparison of Statisti	cal Approaches	tor ivil litidiicity	Correction in	Refinal Sen	SITIVITV A	naivsis
I GINICI	companyon or statisti	cui / (ppi ouciics		CONCLUONING	1101101 3011	JICIVICY / 1	

Approach	Description	Advantages	Limitations
FDA's current approach	Requires an improvement of at least 7 dB from baseline in at least 5 prespecified loci within the central 30 degrees of the visual field	 Strong control of familywise Type I error rate Simplicity in implementation and interpretation 	 5 loci need to be pre-specified Overly conservative, reducing statistical power Totally ignore the improvements in nonspecified loci
Bonferroni correction	Adjusts the significance level for each test by dividing the desired overall alpha by the number of tests	 Strong control of familywise type I error rate Simple to implement 	 Extremely conservative, especially with 68 tests Reduces statistical power, potentially missing real effects
Hochberg's step-up procedure	A commonly used testing procedure that is less conservative alternative to Bonferroni by testing the largest <i>P</i> value at 0.05 significance level and continuing the testing procedure with reduced significance level for any negative test results	 Less conservative than Bonferroni, offering better power Controls familywise type I error rate 	 More complex to implement and interpret Still very conservative with 68 tests
Our proposed approach	≥7 unspecified loci with ≥7 dB improvement	 Balances type I and type II error control Allows for detection of improvements across the entire visual field More sensitive to global treatment effects Does not require prespecification of loci 	 May be less effective for localized retinal defects or diseases with regional variability The fixed 7 dB threshold may not be optimal for all disease stages and progression rates Does not account for spatial correlations between adjacent loci

a positive outcome, then there is a 5.3% probability that the positive outcome is due to chance alone.

This alternative approach provides a more balanced way to address multiplicity while maintaining a reasonable level of statistical significance. It allows for the consideration of the entire 68-locus grid without being overly conservative like the FDA's prespecified loci method. The Table provides a comprehensive comparison of our proposed approach with the FDA's current method and other established statistical approaches for handling multiplicity, highlighting the relative advantages and limitations of each method.^{10,11}

Exploring Data From the Skyline Phase II Clinical Trial

Skyline is a phase II, randomized, masked, multicenter clinical trial comparing 2 doses of ACTC-501 gene therapy administered to patients with XLRP.⁸ In the high-dose group, 4 out of 6 patients (67%) demonstrated a mean improvement of at least 7 dB from baseline in at least 5 loci within the central 30 degrees of the visual field and were classified as "responders." Figure 4 illustrates the change from baseline in sensitivity for the 68 individual loci of each patient, plotted in a waterfall plot. Among the 4 responders, the number of responding loci ranged from 9 to 22. Additionally, there was a robust increase in mean sensitivity across the entire grid, which was sustained at 3, 6, 9, and 12 months, as shown in the inset line plots.

Therefore, choosing seven unspecified loci would still accurately distinguish responders from nonresponders. This means that the four responders observed in the high dose group are not due to multiplicity (i.e. false positive rate inflated from the nominal 5% level). In addition, there was no robust increase in mean sensitivity across the entire grid in the low dose group or the untreated fellow eye group, further adding to the conclusion that choosing more than five unspecified loci can be effective in controlling for multiplicity. In other words, the specificity of the increased mean sensitivity to the high dose group, and the absence of this effect in the low dose and control groups, strongly suggests that the observed response is a real treatment effect rather than a spurious result arising from multiplicity; if the effect were merely a consequence



Figure 4. Waterfall plots showing the change from baseline in sensitivity of each of the individual 68 loci in each study eye as well as an inset line plot comparing the change from baseline in mean sensitivity for the whole grid for the study eye (*black line*) and the fellow eye (*grey line*) by study month. CFB, change from baseline; dB, decibel.

New Approach to Addressing Multiplicity in IRDs

of testing multiple loci, one would anticipate seeing similar increases in mean sensitivity across all groups due to chance alone.

Extracting Statistical and Clinical Significance

We selected a 5% false positive rate threshold, which is a standard practice in statistics to strike a balance between avoiding false positive rates and having the statistical power to detect true effects.¹² This threshold ensures that the probability of observing an improvement of \geq 7 dB in at least 7 unspecified loci out of the total 68 due to random variability alone is limited to 5%. In other words, if such an improvement is observed, there is a high degree of confidence that it is likely attributable to a genuine treatment effect rather than random chance.

Using five prespecified loci for measuring improvements in sensitivity is an overkill for multiplicity control, and might unintentionally focus the analysis on those areas alone, possibly missing improvements in other parts of the visual field. To avoid this, we propose using seven unspecified loci as the criterion for assessing sensitivity improvements. This method does not limit the focus on prespecified loci and allows for a more comprehensive assessment that is open to detecting improvements in sensitivity wherever they may occur in the visual field. This flexibility is particularly valuable because the prespecified loci might not necessarily be the areas that exhibit the most substantial improvement in sensitivity.

Furthermore, in the Skyline data, we found a global improvement in sensitivity across the entire visual field, as illustrated by the responders in Figure 4. In such situations, selecting unspecified loci that reflect this overall enhancement in sensitivity can be more clinically meaningful than solely focusing on a few prespecified loci. This approach acknowledges that therapeutic improvements in visual function can manifest holistically across the visual field, and by using unspecified loci, the analysis remains open to capturing these broader improvements.

The proposed alternative statistical method for analyzing retinal sensitivity improvements has significant practical implications for the design and interpretation of gene therapy clinical trials in IRDs. Implementing this approach could lead to more efficient and informative trial designs. For example, by using \geq 7 unspecified loci with \geq 7 dB improvement as the responder criterion, trials could potentially enroll fewer patients while still maintaining adequate power to detect clinically meaningful treatment effects. This is because the method is more sensitive while being at least as equally specific to capturing genuine therapeutic responses across the entire retina, rather than being limited to a small number of prespecified loci. In terms of trial outcomes, this approach provides a more comprehensive assessment of visual function improvements that is less prone to underestimating treatment efficacy. As illustrated in the Skyline phase II trial data, considering improvements across multiple loci revealed robust and sustained sensitivity gains in the high-dose group that may have been missed by focusing on only five prespecified points. Applying this methodology to future IRD gene therapy studies, such as ongoing trials of optogenetic therapies or CRISPRbased treatments, could yield a more accurate picture of their therapeutic potential and help guide clinical decision making. Furthermore, the method's flexibility to detect clinically relevant responses regardless of their specific retinal location makes it well-suited for evaluating therapies targeting different IRD subtypes with varying patterns of retinal degeneration. This could streamline the development of gene therapies for rarer IRDs by enabling a consistent and reliable approach to assessing efficacy across diverse disease phenotypes.

Limitations and Scope of Applicability

Whereas our proposed approach offers several advantages over the current FDA-recommended method, it is crucial to thoroughly examine its limitations and carefully consider its scope of applicability across different scenarios in retinal disease research.

Our method's effectiveness may vary significantly depending on the specific retinal disease being studied. For diseases with diffuse retinal involvement, such as XLRP, our approach is more likely to be suitable. In XLRP, improvements can occur across the entire visual field, aligning well with our method's consideration of multiple unspecified loci. However, for diseases characterized by localized retinal defects, such as Stargardt disease or certain forms of cone-rod dystrophy, our method may be less applicable. In these cases, improvements might be concentrated in specific retinal areas, potentially leading to an underestimation of treatment effects if the affected areas do not meet our threshold of >7 loci with >7 dB improvement. In addition, the effectiveness of our method may be significantly influenced by the rate of disease progression in different inherited retinal diseases. In rapidly progressing diseases, the 7 dB threshold we have established might prove too stringent. This could result in

missing clinically meaningful improvements that fall short of this threshold but are still significant given the aggressive nature of the disease. Conversely, in slowly progressing diseases, our method can still differentiate between treatment-induced improvements and inherent loss of retinal sensitivity, as it is designed to distinguish between natural variability and true progression or treatment effects. However, the slow pace of change in these diseases may require a longer follow-up period to accomplish sufficient magnitude of changes that exceed the natural variability of the disease. Moreover, whereas our method demonstrates robust statistical foundations based on the binomial distribution and shows promising applicability across different inherited retinal disease contexts, we acknowledge that additional validation studies across diverse datasets would further strengthen the generalizability of our findings. Future research could valuably expand upon our work by testing these methods in varied patient populations and clinical settings.

Our approach has been optimized for randomized controlled trials with a parallel group design, which is a common setup in gene therapy studies. However, its applicability to other trial designs, such as crossover trials or single-arm studies, requires further investigation. In crossover designs, for instance, the potential for carryover effects between treatment periods could complicate the interpretation of results using our method. For single-arm studies, the lack of a concurrent control group might make it challenging to distinguish treatment effects from natural disease variability using our approach.

It is important to note that the applicability of our method can be substantially influenced by various patient-specific factors. Age, for example, can affect both the baseline retinal sensitivity and the potential for improvement, with younger patients potentially showing greater capacity for recovery. Disease duration is another critical factor; patients with longer disease durations might have more advanced retinal degeneration, potentially limiting the scope for the substantial improvements our method is designed to detect. Additionally, the relatively small sample size in our study, whereas typical for trials involving rare conditions like XLRP, represents a limitation that future multi-center studies could address. However, the statistical significance of our findings, even with this sample size, suggests meaningful treatment effects that warrant further investigation in larger cohorts.

Whereas our method effectively addresses the issue of multiplicity by requiring improvements across multiple loci, it has limitations in its statistical approach. Notably, it does not account for spatial correlations between adjacent loci, which could be particularly important in some retinal diseases where effects might cluster in specific regions. This spatial independence assumption might lead to over- or under-estimation of treatment effects in diseases with regionally variable progression. Furthermore, the binary nature of our criterion (>7 loci with >7 dB improvement) might not capture the full spectrum of treatment effects. This approach could potentially overlook smaller but clinically meaningful improvements distributed across many loci, or fail to differentiate between marginal improvements just meeting the threshold and more substantial gains well exceeding it. Another limitation of our current analysis is that we did not explicitly track changes in the number of measurable loci over time, which could provide additional insights into treatment effects, particularly in cases where sensitivity values fall below measurable thresholds. Future studies would benefit from incorporating analyses of both sensitivity changes and the temporal dynamics of measurable versus nonmeasurable loci to provide a more comprehensive understanding of treatment outcomes. Also, while our current analysis focused on identifying improvements in retinal sensitivity, a subset of loci exhibited declines of \geq 7 dB. Although we did not explicitly analyze these declines, future work will investigate whether such reductions exceed those expected from natural disease progression.

Finally, whereas we developed this method with a focus on inherited retinal diseases, its applicability to other ophthalmic conditions that utilize microperimetry for assessment remains to be established. Conditions such as age-related macular degeneration or glaucoma, which also benefit from microperimetry assessments, may have different patterns of visual field loss and potential for improvement. The threshold of \geq 7 loci with \geq 7 dB improvement may not be equally meaningful or achievable across these diverse conditions. Adapting our method to these other conditions would require careful consideration of disease-specific factors and potentially the development of condition-specific thresholds and criteria.

Conclusions

In conclusion, our proposed alternative approach offers a robust and clinically meaningful method for assessing the efficacy of gene therapy for IRD. It balances the need to control for multiplicity with the desire to capture treatment effects across the entire visual field, ultimately benefiting patients by providing a more accurate representation of treatment outcomes. New Approach to Addressing Multiplicity in IRDs

Acknowledgments

The authors thank Beacon Therapeutics for generously providing data from the Skyline clinical trial, a phase II study investigating the safety and efficacy of AGTC–501 gene therapy in male participants with X-linked retinitis pigmentosa (NCT06333249). These data enabled us to conduct test-retest repeatability of pointwise sensitivity of the study and fellow eyes between consecutive visits and explore the alternative approach to address the issue of multiplicity.

Disclosure: A. Yaghy, Beacon Therapeutics (C); D.G. Birch, None; Y. Hwang, None; E. Luo, Beacon Therapeutic (E); J. Jung, Beacon Therapeutic (E); D. Curtiss, Beacon Therapeutic (E); N.K. Waheed, Carl Zeiss Meditec (F), Topcon (F), Nidek (F), Topcon (C), Complement Therapeutics (C), Olix Pharma (C), Iolyx Pharmaceuticals (C), Hubble (C), Saliogen (C), Syncona (C), Ocuyne (I), Beacon Therapeutics (E)

References

- 1. Rohrschneider K, Bültmann S, Springer C. Use of fundus perimetry (microperimetry) to quantify macular sensitivity. *Prog Retin Eye Res.* 2008;27(5):536–548.
- 2. Yang Y, Dunbar H. Clinical perspectives and trends: microperimetry as a trial endpoint in retinal disease. *Ophthalmologica*. 2021;244(5):418–450.
- 3. Buckley TMW, Jolly JK, Josan AS, Wood LJ, Cehajic-Kapetanovic J, MacLaren RE. Clinical applications of microperimetry in RPGR-related retinitis pigmentosa: a review. *Acta Ophthalmol* (*Copenh*). 2021;99(8):819–825.
- Taylor LJ, Josan AS, Jolly JK, MacLaren RE. Microperimetry as an outcome measure in RPGR-associated retinitis pigmentosa clinical trials. *Transl Vis Sci Technol.* 2023;12(6):4.

- Ranganathan P, Pramesh CS, Buyse M. Common pitfalls in statistical analysis: the perils of multiple testing. *Perspect Clin Res.* 2016;7(2):106– 107.
- 6. Weinreb RN, Kaufman PL. Glaucoma research community and FDA look to the future, II: NEI/FDA Glaucoma Clinical Trial Design and Endpoints Symposium: measures of structural change and visual function. *Invest Ophthalmol Vis Sci.* 2011;52(11):7842–7851.
- Lam BL, Pennesi ME, Kay CN, et al. Assessment of visual function with cotoretigene toliparvovec in X-linked retinitis pigmentosa in the randomized XIRIUS phase 2/3 study. *Ophthalmology*. Published online February 28, 2024:S0161-6420(24)00162-3, doi:10.1016/j.ophtha.2024.02.0 23.
- Applied Genetic Technologies Corp. A Phase 2/3, Randomized, Controlled, Masked, Multi-Center Study to Evaluate the Efficacy, Safety and Tolerability of Two Doses of AGTC-501, a Recombinant Adeno-Associated Virus Vector Expressing RPGR (rAAV2tYF-GRK1-RPGR), Compared to an Untreated Control Group in Male Subjects With X-Linked Retinitis Pigmentosa Confirmed by a Pathogenic Variant in the RPGR Gene. clinicaltrials.gov; 2021. Accessed September 7, 2023, https: //clinicaltrials.gov/study/NCT04850118.
- 9. Edwards AWF. The meaning of binomial distribution. *Nature*. 1960;186(4730):1074.
- Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*. 1988;75(4):800–802.
- 11. Bender R, Lange S. Adjusting for multiple testing– when and how? *J Clin Epidemiol*. 2001;54(4):343– 349.
- 12. Di Leo G, Sardanelli F. Statistical significance: p value, 0.05 threshold, and applications to radiomics—reasons for a conservative approach. *Eur Radiol Exp.* 2020;4(1):18.