tvst

Retina

Quantifying the Progression of Stargardt Disease in Double-Null ABCA4 Carriers Using Fundus Autofluorescence Imaging

Ivana Mihalek[1], Hanna De Bruyn[2], Tomislav Glavan[1], Annie M. Lancos[2], Caitlin M. Ciolfi[2], Katarzyna Malendowicz[2], Sigrid Aslaksen[3], Laurie L. Molday[3], Robert S. Molday[3], and Anne B. Fulton[2,4]

[1] Department of Molecular Medicine and Biotechnology, Faculty of Medicine, University of Rijeka, Rijeka, Croatia
[2] Department of Ophthalmology, Boston Children's Hospital, Boston, MA, USA
[3] Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia, Canada
[4] Department of Ophthalmology, Harvard Medical School, Boston, MA, USA

Correspondence: Ivana Mihalek, Department of Molecular Medicine and Biotechnology, Faculty of Medicine, University of Rijeka, 51000 Rijeka, Croatia. e-mail: ivana.mihalek@uniri.hr
Anne B. Fulton, Department of Ophthalmology, Harvard Medical School, Boston, MA 02115, USA. e-mail: anne.fulton@childrens.harvard.edu

Received: July 31, 2024
Accepted: January 27, 2025
Published: March 17, 2025

Keywords: ABCA4; Stargardt disease; fundus autofluorescence; disease onset; disease progression

Citation: Mihalek I, De Bruyn H, Glavan T, Lancos AM, Ciolfi CM, Malendowicz K, Aslaksen S, Molday LL, Molday RS, Fulton AB. Quantifying the progression of Stargardt disease in double-null ABCA4 carriers using fundus autofluorescence imaging. Transl Vis Sci Technol. 2025;14(3):16, https://doi.org/10.1167/tvst.14.3.16

Purpose: To score real-world fundus autofluorescence (FAF) images of pediatric patients with ABCA4-related Stargardt disease (STGD1), in a way that is automatable, scales with the disease progression, and is applicable to a wide time interval in the natural history of the disease.

Methods: We developed the score based on a series of Optos wide-field FAF images of pediatric STGD1 patients (73 images; 14 individuals) and controls (27 images; 8 individuals). The patients' images were obtained over up to 6 years, and the controls over up to 5 years. In each image, we manually selected an artifact-free region, within which we evaluated an average of the pixel-level intensity score, constructed so that the average increases with progression of the disease.

Results: The score we propose provides a statistically robust measure of disease progression (91% Spearman correlation with the absolute age, 97% with the estimated time from onset, when averaged over both eyes), comparable across timepoints and patients.

Conclusions: FAF is a reliable tool in STGD1 diagnostics, but its quantitative description must be modified to be applicable to tracking the disease progression. Analyzing images obtained in the course of clinical care of pediatric patients poses special challenges that make complete automation difficult.

Translational Relevance: Our methodology provides a quantitative tool for investigating the natural progression of the Stargardt disease, and, potentially, the effects of genotype, environment, and therapeutic intervention on its course.

Introduction

Retinal flecks are a clinical hallmark of ABCA4-related Stargardt disease (STGD1),[1] the most common of the heritable macular degenerations.[2] Flecks are attributable to the failure of the ABCA4 protein to clear the byproducts of the visual cycle.[3] The byproducts, bis-retinoids and lipofuscin,[4,5] are autofluorescent and thus visible using en face imaging procedures readily available in the clinic. With the course of the disease, the flecks change from hyper- to hypoautofluorescent, grow, and fuse into larger affected areas.[6]

These changes make the STGD1 fundus autofluorescence (FAF)[7] spatially heterogeneous and variable over time.[6] Investigators have exploited the variation in FAF to track progression of the disease.[8–10] Cukras et al.,[11] in particular, recognized progression of the overall intensity from hyperfluorescent to hypofluorescent. The FAF-based progression studies have relied on manual labeling of the observable changes, making them inapplicable to large input sets. Conversely, the proposed automatable methods[12–14] were applicable to a limited time window in the disease history or did not behave monotonically with disease progression.

Herein, we quantify clinical FAF images at the posterior pole of young patients with ABCA4 allele pairs known to result in no functional protein. In particular, we define an intensity-based score that is agnostic of the structural details of the FAF image, and changes monotonically from early to advanced stages of the disease.

Methods

Participants

This study was conducted in accordance with the Declaration of Helsinki and Health Insurance Portability and Accountability Act regulations, and approved by the Boston Children's Hospital Institutional Review Board.

Demographics

We studied 14 patients in whom STGD1 had been diagnosed in childhood. The reported age of onset of visual difficulties ranged from 4 to 8 years (median, 5 years) (Supplementary Table S1). We used the age of onset to estimate the duration of disease. The 8 controls were 25 to 42 years old (median, 26 years old) at the time of imaging.

Genotyping

All 14 participants were carriers of at least two pathogenic variants in ABCA4. Seven had the segregation of ABCA4 variants confirmed; for the remaining seven, the clinical features made it highly probable that the assigned variant segregation was correct (Supplementary Table S2). We required strong evidence that each genotype is effectively double null; namely, that both alleles have either (1) an early stop codon, or (2) experimentally confirmed lack of expression, or (3) experimentally confirmed lack of protein function. In both (2) and (3), we defined the lack of functionality as less than 10% of the wild-type value.

The variants newly characterized for this work in the Molday laboratory, using methods previously described,[15] are shown in Figure 1, and listed in Supplementary Table S3. The variants with previously published experimental characterization are listed in Supplementary Table S2.

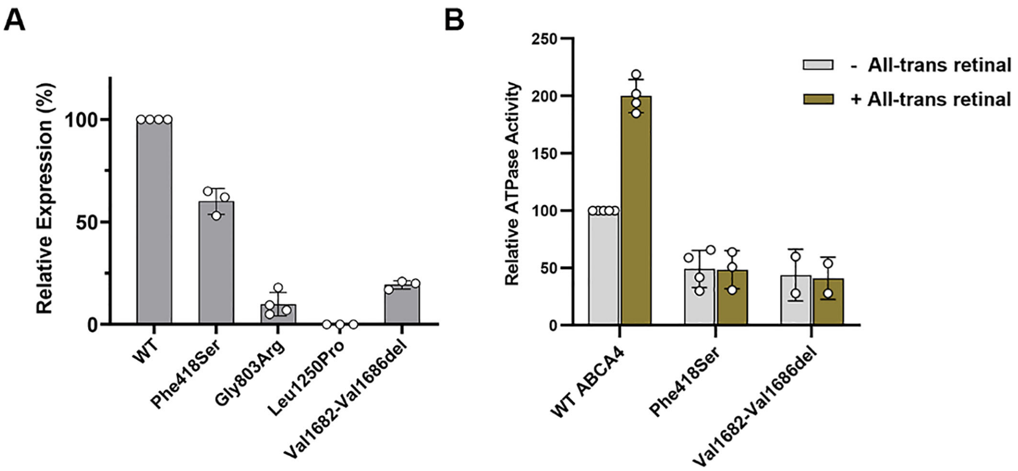

Figure 1. Biochemical characterization of wild-type (WT) ABCA4 and STGD1 variants from transfected HEK293T cells. (**A**) Protein expression levels of the variants relative to WT ABCA4 were determined from Western blots labeled for ABCA4. Data is the mean ± SD for 3 or more independent experiments (data points). (**B**) ATPase activity of purified WT ABCA4 and variants determined at the same protein concentration in the absence and presence of all-trans retinal. Data is expressed relative to WT ABCA4 in the absence of all-trans retinal. A two-fold increase in ATPase activity was observed measured for WT ABCA4 in the presence of all-trans retinal. No significant increase in ATPase activity was found for the ABCA4 variants in the presence of all-trans retinal. Data expressed as a mean ± SD for the number of independent experiments shown by the data points. The expression levels of the Gly803Arg and Leu1260Pro variants were too low for ATPase activity measurements. See also Supplementary Table S3.

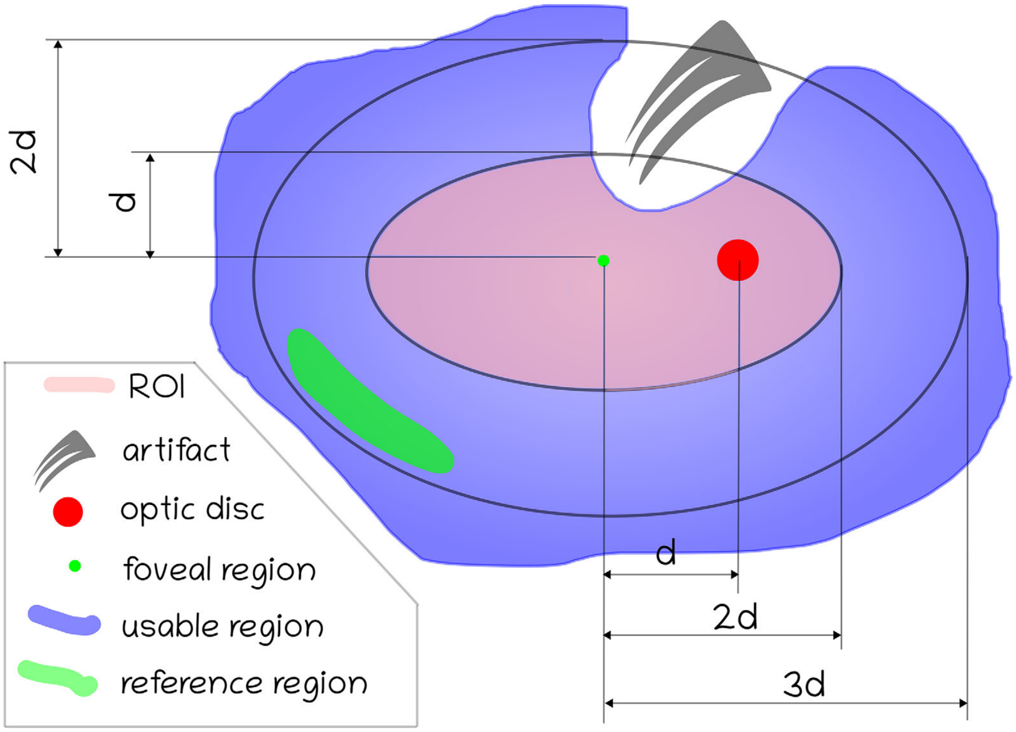

Figure 2. Diagram of the region of interest (ROI; *shaded pink*). The ellipse is centered on the fovea (*green dot*). However, both the fovea and the optic disc (*red*) are excluded from the ROI. The intersection of the inner ellipse with the usable region (the boundary outlined in *blue*) defines the ROI. We selected the reference retinal area (*green*) laying between the inner and the outer ellipse.

Image Acquisition

In 38 sessions, we obtained 73 Optos wide-field autofluorescent images with fovea and optic nerve head identifiable and thus acceptable for analysis. Eight patients were seen in two or more sessions (median, 3; range, 2–7). At the majority of sessions, both eyes were imaged, but in the patient with seven sessions, at three visits only one eye yielded analyzable images. In those patients with two or more sessions, the median interval between visits was 0.9 years (range, 0.3–4.4 years). The age at first, or only, imaging session ranged from 5.2 to 19.4 years (median, 9.6 years). The median acuity was 0.88 logarithm of the minimum angle of resolution (logMAR) (20/150 Snellen) and ranged from 0.18 to 1.60 logMAR (20/30 to 20/800 Snellen). For the patient imaged at age 5 years, acuity was then normal for age (log MAR, 0.18 −0.30; 20/30–20/40 Snellen), but deteriorated later on. No patient had high refractive error; over the period of all observations the median spherical equivalent was +0.88 (range, −4.00 to +2.75) diopters. As expected in patients with low vision and central scotoma, fixation tended to be eccentric and unsteady. In the seven patients who had fixation examinations on the MAIA microperimeter, the 95% bivariate contour ellipse area (a measure of fixation stability) of each exceeded the normal mean.[16]

For controls, we analyzed a total of 27 images gathered in 1 to 5 sessions, spanning up to 5 years. All of them had excellent visual acuity (mean 0.0 logMAR), and median spherical equivalent of −0.9 diopters (range, −0.5 to −3.4 diopters).

We acquired the autofluorescent images using California Optos Panoramic Ophthalmoscope (Optos, Dunfermline, Scotland, UK), a 200° confocal scanning laser ophthalmoscope in the autofluorescence mode (532 nm). All images were acquired through dilated pupils (combination drop: phenyleprine 2.5% mydriacyl 1%, and cyclogyl 1%; Leiters, Englewood, CO). The patients were coached to open eyes wide; nonetheless, eyelash artifacts, blinks, and so forth occurred, as Optos images take approximately 0.4 seconds to acquire. For analysis, we required that an image showed fovea, parafovea, and optic disc.

Image Analysis

Region of Interest (ROI) Geometry

We define the ROI in each FAF image as an ellipse at the posterior pole (Fig. 2). The ellipse is centered at the fovea. Its major axis passes through the optic disc, and has the length of $2d$, where d denotes the distance between the centers of the fovea and the optic disc. The length of the minor axis is d. We exclude from the ROI the circular areas centered at the fovea and disc, with the radii of $d/9$ and $d/3$, respectively. We also exclude the portions of the ellipse falling outside the manually determined usable region.

Scoring an FAF Image of a STGD1 Fundus

To each pixel p in the ROI, we assign value $s(p)$,

$$s(p) = \begin{cases} 10[I_{bg} - I(p)] & \text{if } I(p) < I_{bg} \text{ (dark pixels)} \\ [I(p) - I_{bg}] & \text{if } I(p) > I_{bg} \text{ (light pixels)} \end{cases} \quad (1)$$

where $I(p)$ is the pixel intensity (a number between 0 and 255), and I_{bg} is the average intensity in the reference region. Giving greater weight to the dark pixels ensures that the score keeps increasing with time, even as the pixels gradually change fom light to dark. The details behind our choice of the scoring function can be found in the Discussion. For the impact of varying the relative weight of dark and light pixels (the factor 10 in Equation 1), see Supplementary Figures S1–S5.

The score S assigned to each image, is simply the average of $s(p)$ values for all pixels within the ROI:

$$S = \frac{1}{N} \sum_{p \text{ in ROI}} s(p), \quad (2)$$

where the sum (Σ) runs over all pixels within the ROI, and N is their total number.

Technically, the score S measures the severity of the FAF aspect of the STGD1 phenotype. When applying S to track the progression of the disease, we are making a tacit assumption that the disease develops without periods of regression of the symptoms visible in FAF. So far, this assumption has not been contradicted by the available data.

The Reference Intensity

To estimate the reference (background) intensity, I_{bg}, we define an outer ellipse (Fig. 2), with the same characteristics as the ellipse used in defining the ROI, except that the lengths of its major and minor axes are $3d$ and $2d$, respectively. The I_{bg} is then evaluated as the mean intensity in the manually selected sampling region (green outline in the outer ellipse in Figs. 2 and 3), corrected for the average difference between the intensities in the inner and the outer ellipse in healthy subjects, a fixed number equal to 15 in the current parametrization.

Image Preprocessing

The original images were not rescaled or resampled other than changing the [0,1] to [0,255] grayscale representation. Because the values used in the calculation of $s(p)$ are measured from the average, they are invariant to the overall offset in the distribution. Therefore, we chose not to apply any other recalibration, to avoid introducing image processing artifacts into our analysis.

Manual Steps

Our protocol depends on human curation. To include imperfect images from our clinical practice, we resorted to several manual preparatory steps.

The locations of fovea and optic disc centers were estimated manually. This process enabled us to include images in which these two geometry-defining points were obscured by an artifact or by the disease-related hypofluorescent area.

The reference region was also outlined manually, because any predefined region might become affected in some patients, especially in the later stages of the disease. Whenever possible, we sought to satisfy these two requirements: (1) the reference region should be outside the inner ellipse, but within the outer one and (2) the pixel intensity distribution within the reference region should result in a distribution close to normal (gaussian).

A Step Toward Automating the Analysis

As a proof of concept, in the subset of images with known location of the fovea and ROI free of artifacts, we were able to perform the analysis without manual intervention. Specifically, to find a suitable reference region, we searched computationally the outer ellipse for a segment (Supplementary Fig. S11), with pixel intensity distribution that is close to normal and centered close to the mean value of distributions in all segments.

Dataset Size Considerations

Through a simulation, we estimated the number of data points needed to meet the usual criterion of statistical power (the ability to reject the null hypothesis, evaluated as the fraction of rejections in the simulated data) being greater than 0.8 (Supplementary Fig. S6). In the simulation, we allowed a wide range of fluctuations, and took the Spearman correlation of 0.8 and the p value of 1.e−6 as rejection criteria. With this parametrization the power criterion was met by 20 to 30 data points, well below the size of our image set.

Furthermore, we investigated through subsampling the effect of the sample size on the fluctuation of the correlation between the score and the age at test (Supplementary Fig. S7). Again, sample sizes over 20 result in consistent correlation estimate, and the interquartile range falls under 5% of the reported correlation value for the sample sizes of 70 or more.

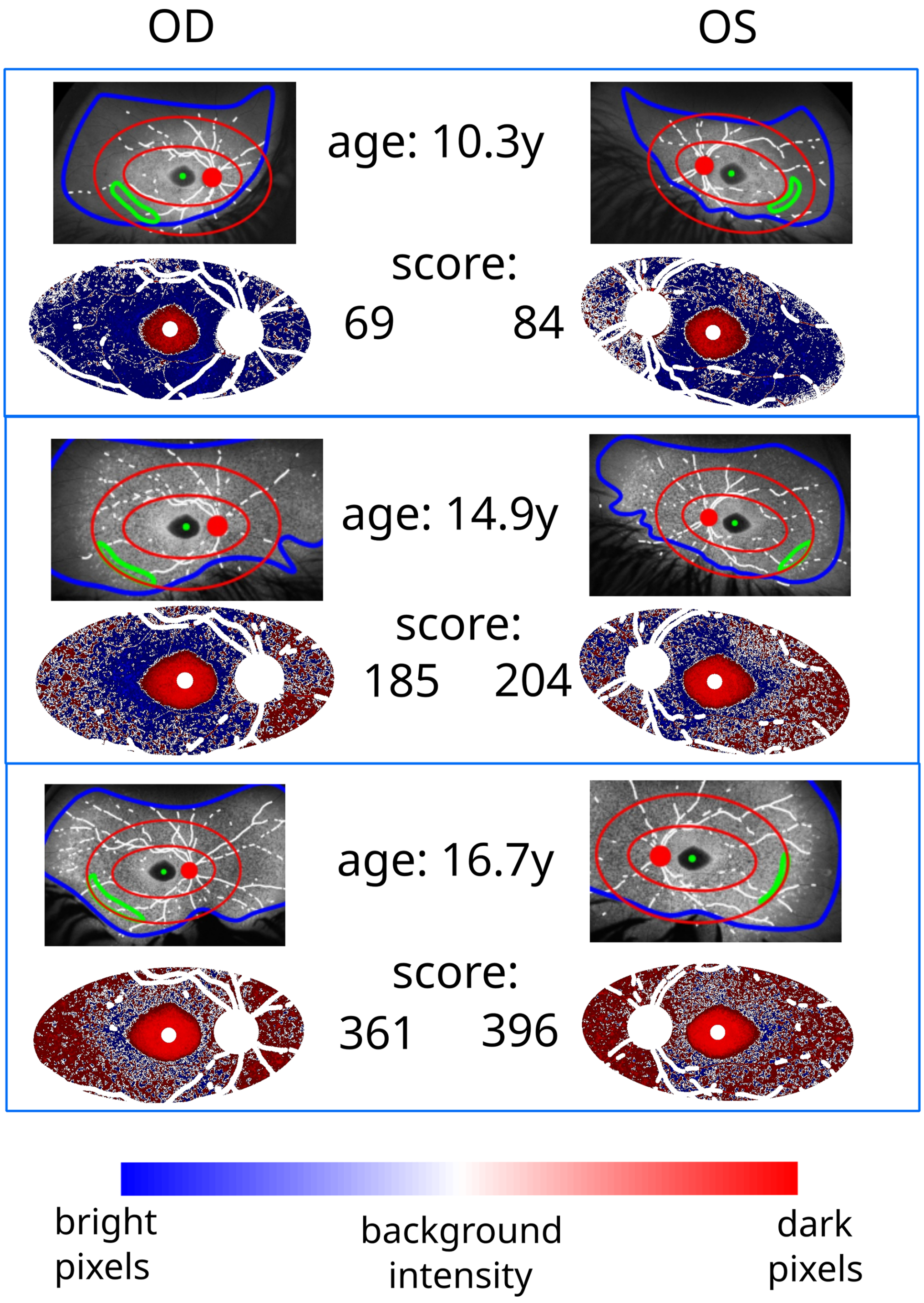

Figure 3. Illustrative timepoints of disease progression, Patient 10 in the Supplementary Table S2. In each panel the top shows the region of the image with the ROI (the inner ellipse, *red*); the image section withing in which we searched for a usable reference region (the outer ellipse, *red*); the reference region (*green line*); optic disc and fovea (*red* and *green dots*, respectively); and the border of the usable region of each image (*blue line*, see also Fig. 2). The pixel score (Equation 1) maps in the bottom row of each panel refer to the inner ellipse. The image score (Equation 2) calculated based on each pixel intensity map (see the main text) is indicated.

Results

As our key result, we show the strong correlation between the score *S* (Equation 2) and the time from disease onset. Before delving into this correlation, however, we present an illustrative set of longitudinal data, in which we highlight the changes in pixel brightness in FAF images that result from the progression of Stargardt's disease. Additionally, we evaluate the possibility of selecting computationally the reference region for images with artifact-free ROI, and demonstrate that knowing the exact foveal position is not critical to our results.

A Sample Case

Figure 3 shows an illustrative case from our cohort (patient 10, Supplementary Table S2). We show the original images, manually annotated with usable and reference regions, and the centers of the fovea and optic disc. The intensity maps within the ROI (Fig. 2), are also shown. The score, *S* (Equation 2), shown in the middle of Figure 3 increases with the age and thus with the progression of the disease.

Time Correlation

In Figure 4, we show the score *S*, averaged over the left and the right eye, as a function of the age of each patient, as well as the time from reported age of onset. The same score can be inspected on the linear scale and without the averaging in Supplementary Figure S8, and for the subset of longitudinal data in Supplementary Figure S9. The correlation between the left and the right eye score is shown in Supplementary Figure S10.

The desired behavior, the monotonic increase of the score *S*, is present both for the chronological age (Spearman correlation 86%), and for the time from onset (Spearman correlation 89%).

In both cases the *p* value is close to zero. The correlation seems to be better if we limit our attention to the patients with confirmed allele segregation, although this reduction of the input size predictably increases the *p* value (Supplementary Table S4). Supplementary Table S5 provides the same statistics for the case when the score values are not averaged over eye pairs. Taken

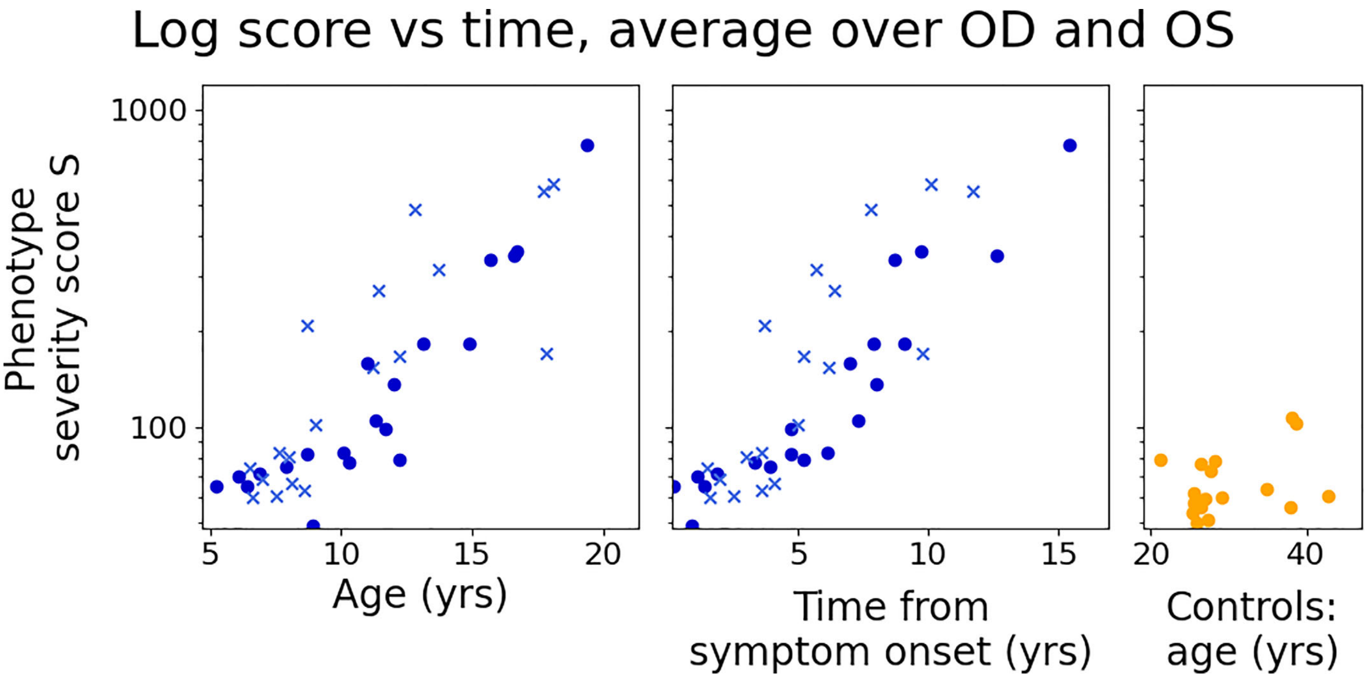

Figure 4. Phenotype severity score *S* (Equation 2), averaged of the left and the right eye, as a function of time. Note that the time in the left-most panel is the absolute age of the patient at the time the FAF image was taken, while in the middle panel the time is measured from the reported age of onset. Points marked with circles: confirmed (Sanger or pedigree) variant segregation; points marked with Xs: reasonably certain but unconfirmed variant segregation. The Spearman correlation is 86% for the correlation with the absolute age, and 89% for the correlation with the time from onset. For the points with confirmed phasing, these values 91% and 97% respectively. All p-values are below 1E-10. For detailed statistical evaluation (see Supplementary Table S4). The same data shown on the linear scale and the accompanying statistics can be found in Supplementary Figure S8 and Supplementary Table S5, respectively.

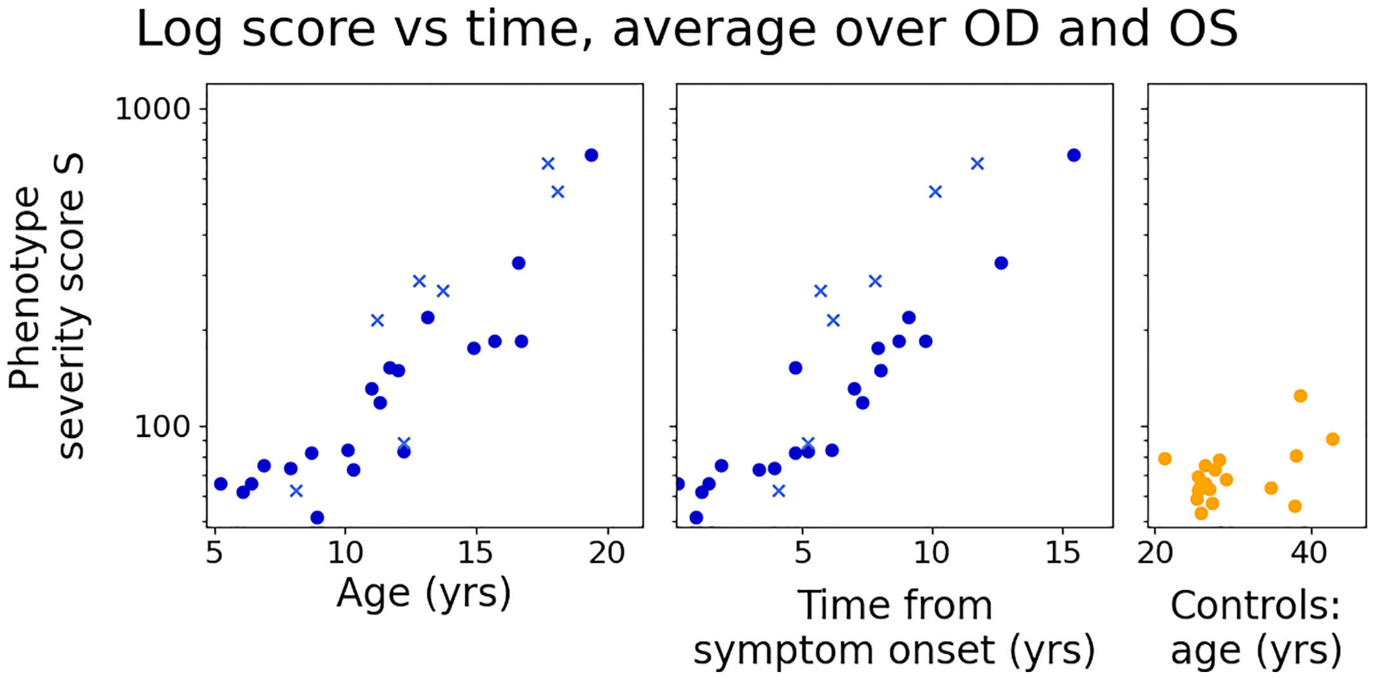

Figure 5. The same as Figure 4, except scores calculated using computationally defined reference regions (Supplementary Fig. S11).

together, the statistics indicate that our approach is robust against details of data grouping and the selection of statistical descriptors.

Automating the Scoring Process: A Proof of Concept

In Figure 5, we show the results of the scoring of the images with an artifact-free ROI (45 images, 60% of the original set). In such cases, we can automate the selection of the reference region, see Supplementary Figure S11. In this automated version of our protocol, the score S has the statistical strength comparable to that achieved by manual curation (Supplementary Table S6).

Impact of the Fovea Location

At the level of precision that we currently achieve, the exact location of the fovea is not critical. In Figure 6, we show the results of the computational experiment in which we recalculate the Spearman correlation between the FAF intensity score (Equation 2) and the time from disease onset, after moving, in an arbitrary direction, the manually assessed fovea location by a fraction of the distance d (Fig. 2). The set of images in this experiment is the same as used in Figure 4. The results indicate that as long as we can locate the fovea to within 10% of d, our results will not change by more than 2%. Conversely, we would have to misplace all foveal centers by 30% of d, to degrade the correlation shown in Figure 4 by 10%.

Discussion

The Scope of the Proposed Approach

We have shown that it is possible to use clinical FAF 532-nm images to track quantitatively the progression of STGD1, from the appearance of hyperfluorescent flecks to the later stages of the disease with predominantly hypofluorescent regions (Fig. 7). In essence, to construct a score that increases monotonically with time, we modified the qFAF approach proposed by Delori et al.[12] Specifically, we measured the signal intensity as an absolute value from the average intensity in an unaffected (reference) region, and gave greater weight to the pixels darker than this average (see Fig. 8). We found a correlation of over 90% between the average score from both eyes and the estimated disease duration (Fig. 4 and the Supplementary Table S4).

We have also shown that the precise location of the fovea is not critical to our results (Fig. 6). The sources of larger uncertainty are the estimate of the age of onset, and the selection of the reference region (Fig. 2).

The major obstacle to fully automating our analysis flow is the presence of artifacts. In a set of images without artifacts, we have shown that we can automate

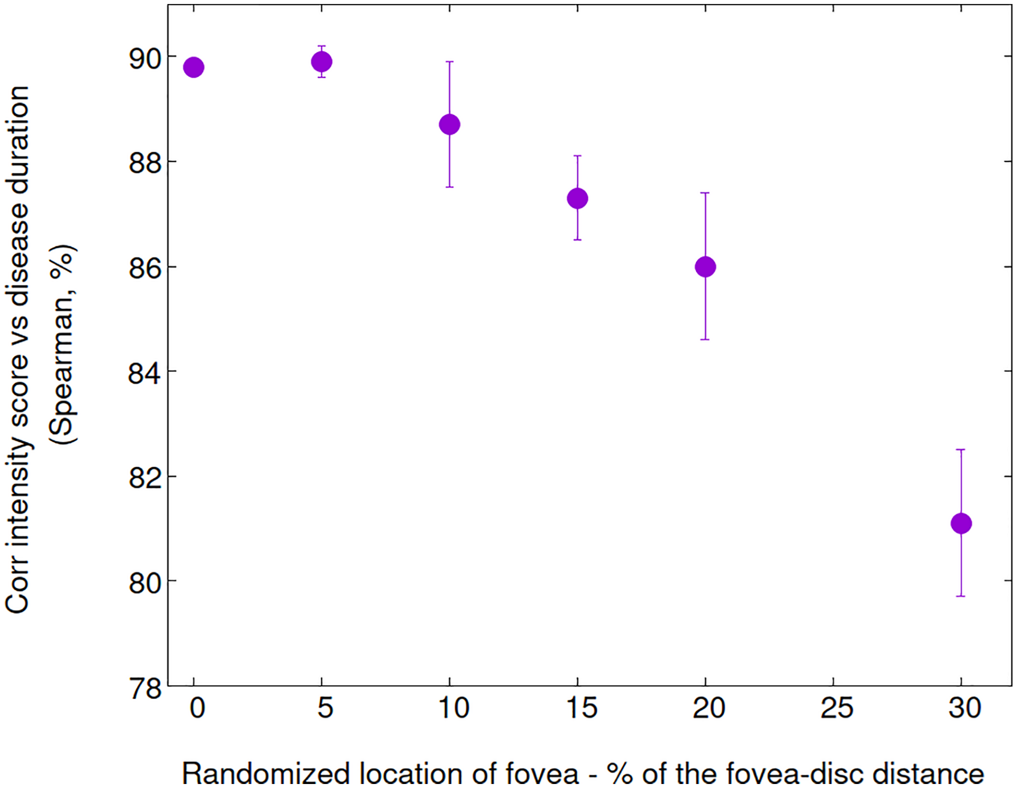

Figure 6. The change in the Spearman correlation, upon randomization of the fovea location. The location is randomized over 10 computational experiments, in which, in each image, we move the manually estimated fovea location in the direction of a randomly chosen angle with the major semiaxis in the range $[0, 2\pi\rangle$, by the amount specified on the x-axis in the graph. The average correlation for each direction size is shown. The error bars represent one standard deviation within each experiment. The set of the images is the same as in Figure 4.

part of our protocol and obtain the results as good as those based on the manually annotated set (Fig. 5). Even so, we still needed the locations of the fovea and optic disc as the input.

They could be obtained, for example, by registration of FAF and optical coherence tomography images.[8]

Previous Work

A number of insightful ways were proposed previously to account for the growth of the central scotoma and appearance of dark flecks observed in clinical practice, most notably AF texture,[13,17] leading disease front expansion[8], and the area of decreased autofluorescence.[18] Decreased autofluorescence, delineated on FAF images acquired using 488-nm excitation,[9,19] has been used by the ProgStar study group[9,18–20] as a tool for communicating the degree of progression in the middle to later stages of disease. This intuitive method relies on measuring the physical area of the foveal region darkened in the FAF images, known to increase as the STGD1 progresses. In our context, we note the need for extensive manual input into image analysis it requires, both in choosing cutoffs between definitely dark, dark, and the remaining levels of brightness, and in delineating the region itself.

Cukras et al.[6] studied a small patient cohort, and documented a centrifugal, radial expansion of the flecks visible at both 488 nm (blue) and 795 nm (near infrared autofluorescence). They further noted that the observable rate of progression and area of abnormal FAF was wavelength dependent.

Importantly, Cukras et al. noted that the flecks progressed from hyperfluorescent to hypofluorescence but did not attempt to incorporate the kinetics of hyper- to hypofluorescence into a numeric summary of the course of an individual patient's FAF.

It should be noted that these efforts, as well our own presented here, were aimed at quantitative and continuous scoring of STGD1 progression. Qualitative labeling of major stages in the progression of the disease, such as proposed by Varela et al.,[21] is expected to agree well with a quantitative score (Supplementary Fig. S12), but has a different purpose, which is enabling effective communication between clinicians.

Deep learning is also making incursion into this subject. In the pioneering work from the National

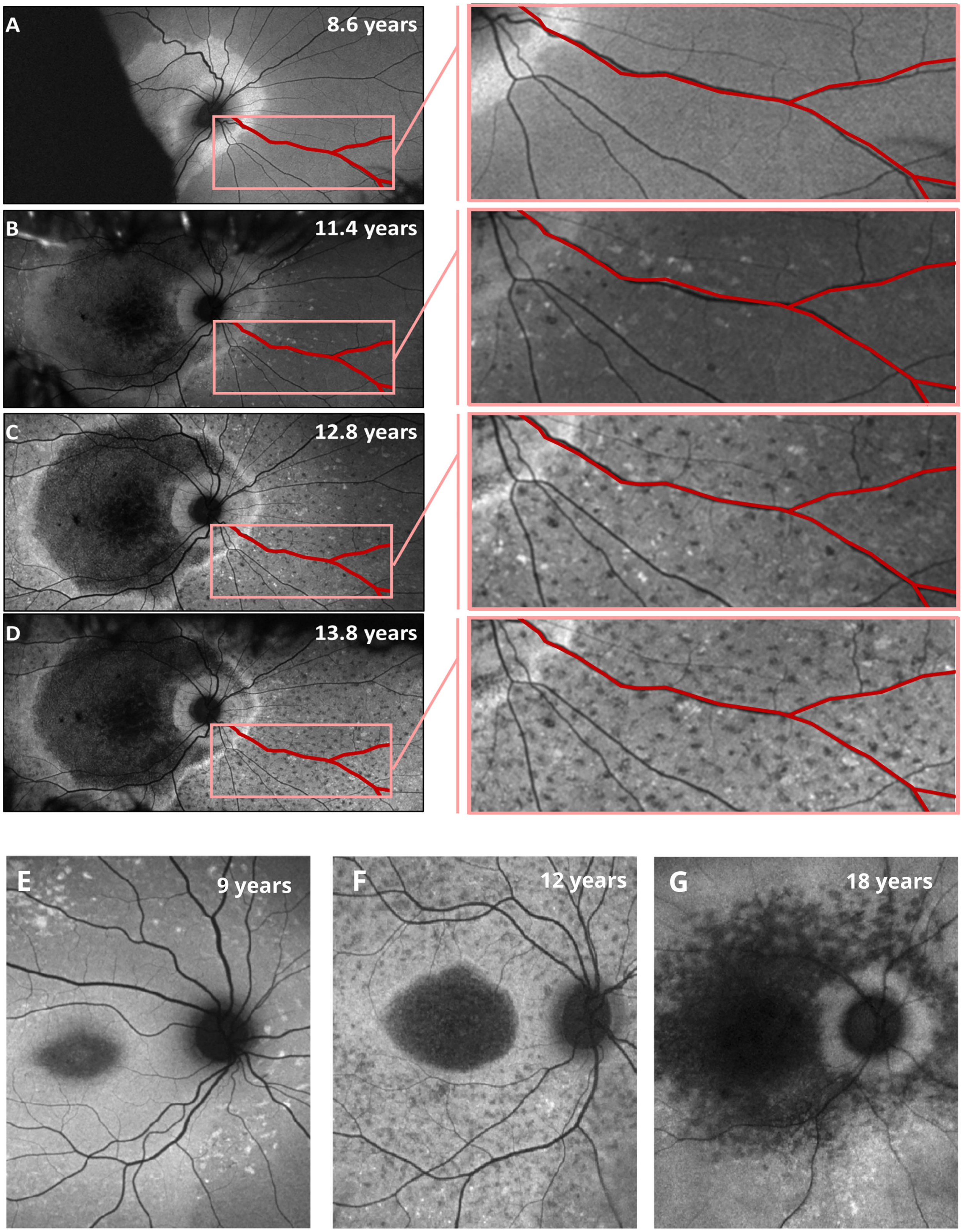

Figure 7. Illustrative points in the disease progression. (**A**–**D**) Same patient (patient 13 in the Supplementary Table S1), detailed view of the disease progression. One blood vessel is highlighted for easier comparison across images. Note the small bright flecks that get replaced by the dark ones as the disease progresses. (**E**–**G**) Three different patients (patients 2, 10, and 1 in the Supplementary Table S1), right eye. Notable features of these images are the bright flecks shortly after disease onset, the appearance of dark flecks a couple of years later, and the central retinal region covered in dark pixels in the advanced disease. For a thorough visual analysis of the phenotype, also see Fakin et al.[6]

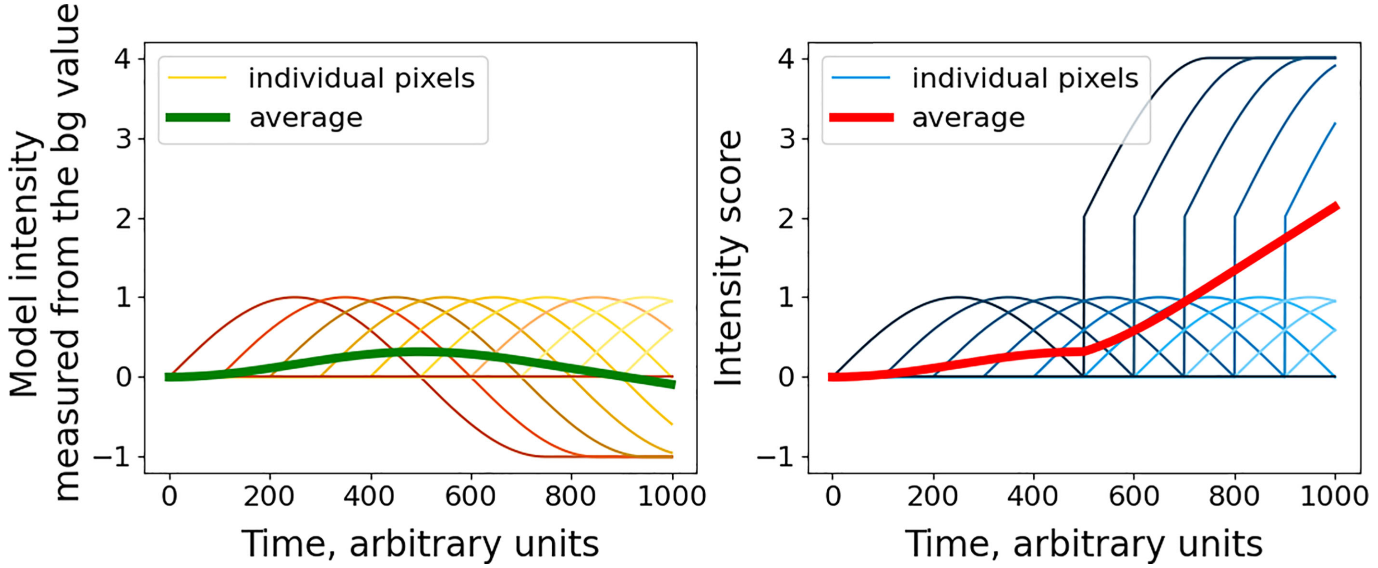

Figure 8. The scoring system, illustrated using a hypothetical case of smoothly changing intensity on a single pixel level. In this toy model, we also assume that we can perfectly register images from different timepoints and track over time the intensity change of individual pixels. If the individual pixel intensity (*orange* and *yellow lines* in the left panel) moves up and down from the refernce (background) level—and individual pixels start changing at random points in time—the average pixel intensity (*thick green line*) is going to do the same. This is precisely what happens with the qFAF. However, if the individual pixel intensity is turned into a score described in the Equation 1, except for visualization purposes we use the multiplication factor of 2, rather than 10 used therein, *blue lines* in the panel on the right, the average over all pixels (*thick red line*) will grow monotonically with time.

Eye Institute,[22] the methodology enabled turning again the community's attention to the hyperautofluorescent aspect of the STGD1 images[7], and Woof et al.[23] reported successful AI assisted FAF image segmentation and detection of the increased hypofluorescence region in patients with STGD1.

In the majority of the reported analyses, the age correlation with the proposed disease-quantifying score has been obscured by the genetically determined severity, which makes the disease appear at different points in life, and progress at variable pace. Several notable attempts to factor out this variable were published recently.[10,24] In its spirit, the work of Sajovic et al.[24] is close to our approach, in that the authors note that specifying the cutoff in the pixel intensity is sufficient to detect the hypoautofluorescent regions of the retina, and no further manual delineation is needed. We further suggest that this last piece of parametrization can also be dispensed with, if the contribution of each pixel tends to zero as their intensity approaches the average value.

Challenges of the Problem

Quantifying STGD1 using FAF has proven elusive for a couple of reasons. Because of the hyper- to hypoautofluorescence transition, Figure 7, the AF pixel intensity distribution initially distorts toward higher intensity, and then shifts toward lower values, Supplementary Figure S13. This means that a number constructed to faithfully track the changes in the intensity, as in the case of quantitative FAF (qFAF),[12,13] will not change monotonically (increase or decrease exclusively) with time.

Furthermore, image brightness and contrast are not calibrated to standard values across patients or timepoints. Shadows of unspecified origin are often present, along with artifacts such as hairs, eyelashes and tears.

Finally, obtaining full retinal view images in children is challenging due to reduced cooperation compared to adults (Supplementary Fig. S14). We could have chosen to limit our input set to high quality images only. This choice would drastically reduce our input set, so we suggest devising, instead, a robust quantitative score that will enable us to work with the real-world data we have at hand.

To address these challenges, we need a score that (1) monotonically changes with disease progression, (2) works with partially obstructed views, and (3) has comparable values across patients and timepoints.

Furthermore, to reduce the arbitrariness, bias, and overfitting, we have adhered to the following rules: all parameters should be whole numbers or rational fractions, and all distances should be expressed using the fovea–optic disc distance as the unit.

Adapting qFAF to STGD1 Specific Challenges

In Equation 2, we have defined a score that satisfies the requirements outlined above. In particular, in the equation for $s(p)$ (Equation 1), the value assigned to each individual pixel, giving ten times the weight to the dark pixels, as they start replacing the bright ones in the later stages of the disease, ensures that the average score S keeps increasing with time (Fig. 4), satisfying our requirement (1) above. Without this adaptation, the average intensity would follow the growing-and-receding pattern observed and discussed by Cukras et al.[11] (Fig. 6 in that work, and illustrated in Fig. 8 here). The wide-angle technology is critical for this point too, because it enables us to track the disease even after the large region centered around the macula becomes hypoautofluorescent.

The choice for the whole image score S is motivated by the observation that the common trait of different STGD1 cases at comparable points in the disease history is not the specific shape or location of bright flecks or dark regions. Rather, the common trait is the overall fraction of the retinal surface with increased or reduced autofluorescence. This independence from the detailed structure of the image makes possible the comparison across different patients, satisfying requirement (2).

It is quite possible that taking into account the fine-grained structure of the retinal FAF image might improve this approach at some time in the future. At this point however, the noise brought in by the artifacts and the delineation of the reference region overwhelm the size of this plausible improvement.

Because we are taking all pixels within a ROI into account, albeit as a small contribution to the overall score if their intensity is close to the reference value, we do not need to invoke a cutoff in either low or high intensity values to evaluate the score. Additionally, taking the average over an ROI ensures that even an incomplete view results in a usable score, as long as it provides a reasonably fair sample of the central retina, addressing the requirement (3).

Future Directions

In this work we limited our attention to double-null genotypes, thus removing one determinant of the observable phenotype – the degree of variant pathogenicity. Extending the input set of images and the accompanying analysis to other, milder, genotypes is one of the future extensions of our approach.

Our input was also restricted to the images acquired using a 532-nm (green) ophthalmoscope mode, dictated mostly by the availability of the images from our clinical practice. Although we believe the approach is applicable to other wavelengths, it remains to be tested.

To give the clinical strength to the scoring system we have proposed, further work is needed in establishing the quality standards for the FAF images, and in making sure that a portion of unaffected background is visible, that no shadows are present, and that standard brightness and contrast transformations are applied to stored images. Learning from the example of Delori et al.,[12] the instrument dependent correction may be applicable, although this might require a coordinated standardization effort.

Nonetheless, even in its present form the score in Equation 2 quantifies the progression in a way which is automatable on high-quality images, and thus potentially useful tool in further genotype–phenotype investigations. The question perhaps remains—why bother with the automation? We know that machine learning is changing the way we analyze biomedical data, with its promise to minimize human bias, automate the analysis, and scale it up to the volumes seen in expert diagnostic systems and clinical trial settings. It is, however, data hungry, and the methodology we propose brings us a step closer to systematically extracting information from large numbers of clinical-grade STGD1 FAF images.

Acknowledgments

Supported by Canadian Institutes of Health Research (CIHR) Grant PJT 175118 to R.M. and NIH-NEI EY 10597 to A.B.F.

Code and Data Availability: The code used in the image analysis in this work is available on GitHub, https://github.com/dogmaticcentral/ABCA4-faf. The data available upon reasonable request from anne.fulton@childrens.harvard.edu.

Disclosure: **I. Mihalek**, None; **H. De Bruyn**, None; **T. Glavan**, None; **A.M. Lancos**, None; **C.M. Ciolfi**, None; **K. Malendowicz**, None; **S. Aslaksen**, None; **L.L. Molday**, None; **R.S. Molday**, None; **A.B. Fulton**, None

References

1. Verdina T, Tsang SH, Greenstein VC, et al. Functional analysis of retinal flecks in Stargardt disease. *J Clin Exp Ophthalmol*. 2012;3, 10.4172/2155-9570.1000233.

2. Rahman N, Georgiou M, Khan KN, Michaelides M. Macular dystrophies: clinical and imaging features, molecular genetics and therapeutic options. *Br J Ophthalmol*. 2020;104(4):451–460.
3. Molday RS, Garces FA, Scortecci JF, Molday LL. Structure and function of ABCA4 and its role in the visual cycle and Stargardt macular degeneration. *Prog Retin Eye Res*. 2022;89:101036.
4. Sparrow JR, Fishkin N, Zhou J, et al. A2E, a byproduct of the visual cycle. *Vis Res*. 2003;43(28):2983–2990.
5. Sparrow JR, Marsiglia M, Allikmets R, et al. Flecks in recessive Stargardt disease: short-wavelength autofluorescence, near-infrared autofluorescence, and optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2015;56(8):5029–5039.
6. Fakin A, Robson AG, Fujinami K, et al. Phenotype and progression of retinal degeneration associated with nullizigosity of ABCA4. *Invest Ophthalmol Vis Sci*. 2016;57(11):4668–4678.
7. Pole C, Ameri H. Fundus autofluorescence and clinical applications. *J Ophthalmic Vis Res*. 2021;16(3):432.
8. Cideciyan AV, Swider M, Schwartz SB, Stone EM, Jacobson SG. Predicting progression of ABCA4-associated retinal degenerations based on longitudinal measurements of the leading disease front. *Invest Ophthalmol Vis Sci*. 2015;56(10):5946–5955.
9. Kong X, West SK, Strauss RW, et al. Progression of visual acuity and fundus autofluorescence in recent-onset Stargardt disease: ProgStar study report #4. *Ophthalmol Retina*. 2017;1(6):514–523.
10. Heath Jeffery RC, Thompson JA, Lamey TM, et al. Classifying ABCA4 mutation severity using age-dependent ultra-widefield fundus autofluorescencederived total lesion size. *Retina*. 2021;41(12):2578–2588.
11. Cukras CA. Centrifugal expansion of fundus autofluorescence patterns in Stargardt disease over time. *Arch Ophthalmol*. 2012;130(2):171.
12. Delori F, Greenberg JP, Woods RL, et al. Quantitative measurements of autofluorescence with the scanning laser ophthalmoscope. *Invest Opthalmol Vis Sci*. 2011;52(13):9379.
13. Burke TR, Duncker T, Woods RL, et al. Quantitative fundus autofluorescence in recessive Stargardt disease. *Invest Opthalmol Vis Sci*. 2014;55(5):2841.
14. Sparrow JR, Duncker T, Schuerch K, Paavo M, de Carvalho JRL. Lessons learned from quantitative fundus autofluorescence. *Prog Retin Eye Res*. 2020;74:100774.
15. Curtis SB, Molday LL, Garces FA, Molday RS. Functional analysis and classification of homozygous and hypomorphic ABCA4 variants associated with Stargardt macular degeneration. *Hum Mutat*. 2020;41(11):1944–1956.
16. Morales MU, Saker S, Wilde C, et al. Reference clinical database for fixation stability metrics in normal subjects measured with the MAIA microperimeter. *Transl Vis Sci Technol*. 2016;5(6):6.
17. Cideciyan AV, Aleman TS, Swider M, et al. Mutations in ABCA4 result in accumulation of lipofuscin before slowing of the retinoid cycle: a reappraisal of the human disease sequence. *Hum Mol Genet*. 2004;13(5):525–534.
18. Strauss RW, Ho A, Jha A, et al. Progression of Stargardt disease as determined by fundus autofluorescence over a 24-month period (ProgStar Report No. 17). *Am J Ophthalmol*. 2023;250:157–170.
19. Strauss RW, Muñoz B, Ho A, et al. Progression of Stargardt disease as determined by fundus autofluorescence in the retrospective progression of Stargardt disease study (ProgStar Report No. 9). *JAMA Ophthalmol*. 2017;135(11):1232.
20. Strauss RW, Muñoz B, Ho A, et al. Incidence of atrophic lesions in Stargardt disease in the Progression of Atrophy Secondary to Stargardt Disease (ProgStar) study: report No. 5. *JAMA Ophthalmol*. 2017;135(7):687.
21. Varela MD, Laich Y, Hashem SA, Mahroo OA, Webster AR, Michaelides M. Prognostication in stargardt disease using fundus autofluorescence: improving patient care. *Ophthalmology*. 2023;130(11):1182–1190.
22. Charng J, Xiao D, Mehdizadeh M, et al. Deep learning segmentation of hyperautofluorescent fleck lesions in Stargardt disease. *Sci Rep*. 2020;10(1):16491.
23. Woof W, de Guimarães TA, Al-Khuzaei S, et al. Quantification of fundus autofluorescence features in a molecularly characterized cohort of more than 3500 inherited retinal disease patients from the United Kingdom. medRxiv. 2024:2024-03.
24. Sajovic J, Meglič A, Fakin A, et al. Natural history of Stargardt disease: the longest follow-up cohort study. *Genes*. 2023;14(7):1394.