Progression of *ABCA4*-Related Retinopathy – Prognostic Value of Demographic, Functional, Genetic, and Imaging Parameters

Philipp L. Müller MD[1,2,3,*], Maximilian Pfau MD[1], Tim Treis BSc [4], Isabel Pascual-Camps MD[5], Johannes Birtel MD[1,2], Moritz Lindner MD[1,6], Philipp Herrmann MD PhD,[1,2] Frank G. Holz MD[1,2]

[1] *Department of Ophthalmology, University of Bonn, Bonn, Germany*
[2] *Center for Rare Diseases, University of Bonn, Bonn, Germany*
[3] *Moorfields Eye Hospital NHS Foundation Trust, London, UK*
[4] *BioQuant, University of Heidelberg, Heidelberg, Germany*
[5] *FISABIO Oftalmología Médica, Valencia, Spain*
[6] *Oxford Eye Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford, UK*

Funding/Grants/Support: This work was supported by the PRO RETINA Deutschland, Aachen, Germany; the RWA, Recklinghausen, Germany; the German Ophthalmological Society (DOG); and the German Research Foundation (DFG, Grant MU4279/1-1 and MU4279/2-1 to PLM and PF950/1-1 to MP). The Department of Ophthalmology, University of Bonn, received technical support from Heidelberg Engineering, Heidelberg, Germany, Carl Zeiss Meditec, Jena, Germany, and CenterVue, Padova, Italy. The sponsor or funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Meeting Presentation: n/a

Conflict of Interest / Disclosures: Frank G. Holz is consultant of Heidelberg Engineering, Heidelberg, Germany. No other disclosures were reported.

Abbreviated Title: Progression of *ABCA4*-Related Retinopathy

*** Corresponding author:**
Philipp L. Müller
University of Bonn, Dept. of Ophthalmology
Ernst-Abbe-Str. 2
D-53127 Bonn, Germany
Tel: +49 / (0)228 / 287-15505
Fax: +49 / (0)228 / 287-14817
Mail: philipp.mueller@ukbonn.de

KEY WORDS

Retina

Fundus autofluorescence

Stargardt disease

Retinal pigment epithelium

Hereditary dystrophy

SUMMARY STATEMENT

This study underscores the prognostic value of demographic, functional, genetic, and lesion geometry parameters in ABCA4-related retinopathy. Trial design including inclusion criteria based on these factors may provide economic benefits by selection of appropriate participants.

ABSTRACT

Purpose: To investigate the prognostic value of demographic, functional, genetic, and imaging parameters on retinal pigment epithelium (RPE) atrophy progression secondary to *ABCA4*-related retinopathy.

Methods: Patients with RPE atrophy secondary to *ABCA4*-related retinopathy were examined longitudinally with fundus autofluorescence imaging. Lesion area, perimeter, circularity, caliper diameters, and focality of areas with definitely decreased autofluorescence were determined. A model was used to predict the lesion enlargement rate based on baseline variables. Sample size calculations were performed to model the power in a simulated interventional study.

Results: Sixty-eight eyes of 37 patients (age range, 14-78 years) with a follow-up time of 10-100 months were included. The mean annual progression of RPE atrophy was 0.89mm^2. The number of atrophic areas, the retina-wide functional impairment, and the age of onset category constituted significant predictors for future RPE atrophy growth, explaining 25.7% of the variability. By extension of a simulated study length and/or specific patient preselection based on these baseline characteristics, the required sample size could significantly be reduced.

Conclusion: Trial design based on specific shape-descriptive factors and patients' baseline characteristics as well as the adaption of the trial duration may provide potential benefits in terms of required cohort size and absolute number of visits.

ABCA4-related retinopathy (Online Mendelian Inheritance in Man # 601691) is the most common single gene retinal dystrophy. It is caused by biallelic mutations in the *ATP-binding cassette sub-family A member 4 (ABCA4*) gene.[1] Dysfunction or loss of function of the encoded protein product (ABCA4, an integral transmembrane protein) leads to excessive accumulation of visual cycle byproducts, and potentially degeneration of the retinal pigment epithelium (RPE) and photoreceptors early in disease progression associated with progressive visual loss.[2]

Using short-wavelength excitation light, different molecular constituents of these visual cycle end products can be visualized with fundus autofluorescence (AF) imaging.[3,4] Despite variable age of onset and clinical manifestation that requires sophisticated classification,[5] *ABCA4*-related retinopathy therefore is characterized by elevated AF intensities.[6] Thus, areas with RPE atrophy theoretically contrasts regions with intact RPE, allowing semi-automated, objective quantification.[7]

Clinically, the RPE atrophy in *ABCA4*-related retinopathy presents in variable transition stages towards complete atrophy ranging from questionably (QDAF) to definitely (DDAF) decreased AF, making the delineation more challanging.[8] A recent study revealed distinctly higher inter-reader agreement in the grading of DDAF in comparison to QDAF, qualifying the use of DDAF progression as suitable surrogate marker.[7] Larger prospective multicenter studies such as ProgSTAR aim to evaluate the potential of DDAF as anatomic endpoint for future therapeutic trials.[9]

In view of variable phenotypic presentation and slower progression rates of *ABCA4*-related retinopathy compared to other retinal diseases,[10] the simple measurement of RPE atrophy lesion size visualized by AF might be uneconomic in terms of required trial

duration and sample size or even insufficient in order to correctly determine treatment effects.[11] Shape descriptive parameters of DDAF such as lesion perimeter and/or circularity recently have been proposed for the identification of eyes with rapid disease progression.[12] However, the longitudinal determination of shape descriptive parameters as prognostic biomarker for disease progression of *ABCA4*-related retinopathy is still pending.

Herein, we refined the analysis of shape-descriptive factors using AF imaging. We aimed to evaluate the relevance of demographic, functional, genetic and these lesion geometry factors as prognostic biomarkers to determine RPE atrophy progression secondary to *ABCA4*-related retinopathy. Based on the obtained results, sample size calculations were performed to facilitate the design of future therapeutic trials.

METHODS

This study was performed at the Department of Ophthalmology at the University of Bonn, Germany. The study was in adherence with the declaration of Helsinki. Institutional Review Board (IRB) approval of the University of Bonn, and patients' written informed consent were obtained.

Patient selection and classification

Inclusion criteria were defined as (1) presence of at least one mutated allele in *ABCA4* (NM_000350.2) in Sanger sequencing with multiplex ligation-dependent probe amplification analysis or next-generation sequencing,[9] (2) a compatible phenotype with flecks at the level of the RPE consistent with *ABCA4*-related Stargardt disease,[9] (3) presence of RPE atrophy, and (4) serial examinations with an interval of at least 6 months. As described before,[8,13] assessment of RPE atrophy in ABCA4-related retinopathy is challenging due to different transition stages using AF imaging. We therefore restricted the definition of RPE atrophy to the end-stage (i.e. demarcated lesions with ≥90% darkness of the optic disc under short-wavelength excitation light, DDAF) that previously showed highest inter-reader agreement.[7] The size of DDAF has to be at least 0.05 mm² (each single atrophic area in cases of multifocality), and the entire lesion must be completely visualized on the AF image at each visit in order to be advanced for analysis.[13,14] If both eyes met the inclusion criteria, both were included. Insufficient pupil dilation, additional retinal pathology, previous retinal treatment or other ocular comorbidities substantially affecting image quality led to exclusion from the study.

Due to variability in clinical manifestation, eyes were classified according to age of self-reported symptom onset into early-onset (≤ 10 years), intermediate-onset (11-44 years) and late-onset (≥ 45 years) on the one hand,[10] as well as full-field electroretinogram (ERG) using a Ganzfeld stimulator and Burian–Allan bipolar corneal electrodes (Toennies Multiliner Vision 1.70, Hochberg, Germany) on the other hand.[15] According to Lois and colleagues, Group 1 contained eyes with normal scotopic and photopic responses, group 2 eyes with normal scotopic responses but reduced (over 2 standard deviations) photopic B-wave and 30-Hz flicker amplitudes, and group 3 eyes with impairment of both rod- and cone-driven responses.

Imaging

All patients underwent a complete ophthalmologic examination at each visit. Prior to multimodal imaging, pupils were dilated to at least 7mm using 0.5% tropicamide and 2.5% phenylephrine. The standardized imaging protocol consisted of fundus photography (Visucam, Carl Zeiss Meditec, Jena, Germany), and conventional short-wavelength AF imaging (488 nm excitation and 500-700 nm emission) using a confocal scanning laser ophthalmoscope (cSLO; Spectralis HRA, Heidelberg Engineering, Heidelberg, Germany). Using the manufacturer's software, up to 100 images (centered to the fovea, 30° x 30°, 1536 x 1536 pixels) were automatically aligned and averaged in order to optimize the signal-to-noise ratio.

Image Processing

Measurements of RPE atrophy lesions (i.e. DDAF) were performed by two independent experienced readers (P.L.M. and I.P.C.), masked to the results of each other, using the semi-automated RegionFinder software (Heidelberg Engineering, Heidelberg, Germany,

version 2.6.3) which has been previously established for retinal degenerations.[7,10] If the gradings between both readers differed > 0.15 mm², an additional reader (J.B.) performed the measurement.[10] In these cases, his measures along with the closer of the two primary reader measures were finally advanced for further analysis. As centrifugal and centripetal progression kinetics were previously shown to differ in age-related macular degeneration (AMD) patients,[16] the images were further graded according to presence of foveal involving RPE atrophy (i.e. presence or no presence of DDAF at the foveal center).

The graded annotated images were exported and transferred to ImageJ (Bethesda, Maryland, USA) to measure the shape-descriptive factors using a custom-built plug-in as previously described (Figure, Supplemental Digital Content 1, which illustrates the measurements of the shape-descriptive factors).[12] In brief, *'Focality'* described the number of atrophic foci with an area ≥ 0.05 mm² in each eye, '*Area'* showed the cumulative size of RPE atrophy, 'S*quare Root Area'* was calculated as the square root of the '*Area'* to reduce the dependence of growth rates on the baseline *area*,[12] the cumulative circumference of all foci was described by the '*Perimeter'*, '*Circularity'* (i.e. the roundness of the lesion) was calculated by $\frac{4\times\pi\times Area}{Perimeter^2}$, and the maximum and minimum perpendicular distance between parallel tangents touching opposite sides of the lesion was represented by '$Feret_{max}$' and '$Feret_{min}$', respectively. Scaling-factors for the pixel-to-µm conversion were obtained from the HEYEX software (Heidelberg Engineering, Germany).

Statistical analysis

Statistical analyses were performed using the software environment R (version 3.2.3, The R Foundation for Statistical Computing, Vienna, Austria). Differences in demographic, functional, and genetic data in dependence of the ERG-based groups, age-of-onset groups

or foveal status groups were tested for statistical significance using an analysis of variance (for interval scaled data) or a chi-squared test (for frequencies).Differences of lesion size and geometry at baseline were tested using the Kruskal-Wallis test due to the non-normal distribution. To quantify the differences of subgroups on RPE atrophy progression, linear mixed-effects models were used.

In order to probe whether square-root transformation reduces the association between lesion area and future DDAF progression two mixed-effects models were analyzed considering eyes nested in patients as random effects. In the first model, DDAF progression rate (in mm^2/year) served as dependent variable and lesion area (in mm^2) as independent variable. In the second model, square-root transformed DDAF progression rate (in mm/year) served as dependent variable and lesion area (in mm^2) again as independent variable. P-values were based on the Wald tests.

Multiple potential predictors for future (square-root transformed) DDAF progression rates were evaluated using linear mixed-effects models analogous to Pfau and coworkers.[12] Significant predictors were selected though backward elimination of non-significant effects using P-values for the fixed effects calculated from F tests based on Sattethwaite's approximation. The predictive accuracy of the final model was probed using patient-wise leave-one-out cross-validation (i.e. repeated fitting of the model to n-1 patients and evaluation of the model on the one remaining patient). The cross-validated coefficient-of-determination (R^2) served as measure of model accuracy.

The course of visual acuity over time was analyzed as proposed by Lambertus et al. and Giolo et al. using Kaplan-Meier curves with Turnbull’s estimator.[10,17].

A simulation study for power calculation was performed for possible future interventional trials aiming at slowing RPE atrophy progression. Only patients with two eyes within the data were analyzed for the relevant parameters, so that one eye could serve as control while the other one was treated.[10] The simulations were done using Mediana, an implementation of the Clinical Scenario Evaluation framework in the R language.[18] Clinical trial simulation with 5000 runs each, a significance level of 0.05, a dropout rate of 15 %, and a trial durations of 1-4 years were performed. The statistical powers for the sample sizes were then extracted as well as the threshold for a power of 80 % was determined.

For evaluation of inter-rater reliability, intraclass correlation coefficients (ICC, two-way random, absolute agreement), 95% coefficients of repeatability (CoR) and coefficients of variation (CV) were determined for lesion size and shape-descriptive factors. Bland-Altman plots were generated for visualization of limits of agreement, and Spearman's rank correlation coefficients (ρ) were calculated between the absolute differences and the mean values to determine whether measurement variability increases with lesion size.

RESULTS

Cohort characteristics

Seventy eyes (38 patients), of which 46 (66%) were female and 24 (34%) male, with RPE atrophy secondary to *ABCA4*-related retinopathy (age [years], 45.51±16.70; range 14 – 78) were included in the study (Table 1 and Table, Supplemental Digital Content 2, which shows the individual baseline and progression data of all included eyes).

Twenty-one (30.0%), 20 (28.6%), 5 (7.1%), and 24 (34.3%) eyes revealed a visual acuity of ≤ 0.1 logMAR (≥ 20/25 Snellen equivalent, no impairment), 0.2-0.5 logMAR (20/32-20/60, mild impairment), 0.6-0.9 logMAR (20/80-20/160, moderate impairment), and ≥ 1.0 logMAR (≤ 20/200, severe impairment), respectively. Consistently, the majority of our cohort (41 of 70 eyes, 59%) showed foveal non-involvement (i.e., no foveal involving RPE atrophy). These eyes revealed a significantly higher age of onset compared to eyes with foveal involving DDAF lesions. Disease onset was identified as early-onset in 10 (14%) eyes, intermediate-onset in 37 (53%) eyes and late-onset in 23 (33%) eyes. The disease duration did not differ between these groups. In contrast, ERG-based group 3 eyes (n = 10, 14%) revealed significantly longer disease duration and earlier age of onset compared to ERG-based group 2 (n = 45, 64%) and group 1 (n = 15, 21%) eyes. Accordingly, the latter revealed a distinctly higher proportion of foveal non-involving DDAF (60% and 67% in group 1 and 2 versus 20% in group 3) and associated better visual acuity. Both eyes of each patient always met the same group criteria.

At least two mutations in *ABCA4* could be detected in 92%of included patients (n=35, n_{eyes}=65). The remaining rare cases with only one detected *ABCA4*-variant were

consistently negative for mutations in *RDS/PRPH2* (Table, Supplemental Digital Content 2, which shows the individual baseline and progression data of all included eyes). The single most common *ABCA4*-variants in our cohort were truncating mutations (n=35, 50%, always with non-truncating variant *in-trans*) and the N1868I-mutation (n=19, 27%). The latter revealed an association with older age of onset (intermediate onset, n=10, 53%; late onset, n=9, 47%), foveal non-involvement (n=14, 74%) as well as less functional impairment (group 1, n=7, 37%; group 2, n=12, 63%). The generally common mild G1961E-variant was rarely found in eyes with DDAF lesions (n=6, 9%) and uniquely in group 1 eyes (i.e., eyes with preserved full-field scotopic and photopic function) with foveal involvement.

Baseline characteristics of lesion geometry

A detailed compilation of baseline characteristics can be found in Table 1. Briefly, the lesion size ('*Area'* and '*Square Root Area')* as well as most shape-descriptive factors (*Perimeter*, *Circularity*, $Feret_{max}$, $Feret_{min}$ and *Focality)* differed significantly between ERG-based groups, with least and most widespread RPE atrophy in group 1 and group 3, respectively. Of note, *Circularity* decreases with a higher number of DDAF spots. The highest values could be found in Group 1 eyes that always showed a unifocal DDAF lesion. Subgroups classified for age of onset, foveal status, and/or genotype did not show significant different DDAF lesion size or shape-descriptive parameters in our cohort at baseline. However a trend to more widespread DDAF lesions could be found with later age of onset.

Progression data

Overall, 191 visits have been graded. The follow-up time ranged from 10 to 100 months with a mean (± standard error of the estimate) of 35.52±4.32 months (Table 2). The mean

follow-up time was not significant different between the different subgroups classified dependent on full-field ERG, age of onset, foveal status, and/or genotype (i.e., presence of the truncating, G1961E or N1868I mutations). Meanwhile, distinct changes could be observed that will be presented in the following (Figure 1).

Functional impairment over time

The biggest loss in BCVA was found in eyes with foveal non-involvement and eyes assigned to late age of onset with a change of [mean ± standard error of the estimate] 0.07±0.01 and 0.07±0.02 LogMAR/year, respectively (Table 2). In 9 of 40 (22.5%) eyes with initial foveal non-involvement, the DDAF lesion progressed into the foveal center during the individual observation time. This low proportion may be traced back to the significantly slower centripetal than centrifugal spread of RPE atrophy leading to long-lasting preservation of foveal integrity (Figure 2), and the different observation time:, The time to event from baseline to the visit when foveal involvement was first noticed in these patients ranged from 2.26 to 7.62 years with a median of 3.22 years. However, 70 % of the entirety of patients with initial foveal non-involving DDAF had a shorter observation time.

Using a regression model, time-to-event analysis of patients between age-of onset and the endpoints ≥ 0.2 logMAR (≤ 20/32 Snellen equivalent, mild impairment), ≥ 0.6 logMAR (≤ 20/80, moderate impairment) and ≥ 1.0 logMAR (≤ 20/200, severe impairment) could be performed (Figure 3). The 90% cumulative fraction (95% confidence interval) of eyes for reaching these endpoints was 4.76 (3.18, 6.34), 10.36 (8.71, 12.01) and 9.86 (8.63, 11.09) years after baseline visit, respectively. Eyes with initial preserved visual acuity (≤ 0.1 logMAR, ≥20/25 Snellen equivalent) reached the endpoint ‘mild impairment’ (90%

cumulative fraction) at 6.74 (95% confidence interval: 4.94, 8.54) years after baseline – significantly later compared to the entirety of eyes described above.

Progression of RPE atrophy and change in lesion-shape

The progression rates (mean ± standard error) were 0.89±0.13 mm^2/year for *Area*, 0.20±0.02 mm/year for *Square Root Area,* 1.57±0.35 mm/year for *Perimeter*, 0.34±0.05 mm/year for $Feret_{max}$, 0.28±0.04 mm/year for $Feret_{min}$. Hereby, factors describing primarily the lesion shape remained rather constant with a change of -0.00±0.01 /year for *Circularity*, and 0.04±0.08 /year for *Focality* (Table 2). There was a good correlation between both eyes of individual patients (Pearson Correlation for *Square Root Area* measures, ρ = 0.660; Figure, Supplemental Digital Content 3, which illustrates the statistical analysis of the progression rates.).

A mixed-effects model analysis (with eyes nested in patients as random effects) revealed, that the individual DDAF lesion size (*Area*, in mm^2) had a small, yet significant effect on future non-transformed (i.e. mm^2/year) progression rates ([effect estimate ± SE] 0.10 ± 0.01 mm^2/year per mm^2, $p<0.001$, Figure 1). Application of the square root transformation to *Area* measurements for the progression rate determination eliminated the statistical significance of this relationship (Figure, Supplemental Digital Content 3, which illustrates the statistical analysis of the progression rates).

Despite this square root transformation, ERG-based groups 2 and 3 revealed significantly faster progression of lesion size compared to ERG-based group 1. Classification depending on age of onset and foveal status did not reveal significant differences in lesion size progression rates between subgroups. However, there was a trend to faster progression in eyes with intermediate and late age of onset as well as foveal

non-involvement. The changes of the shape-descriptive factors during the observation time did not differ significantly within the different group classifications (including ERG-based classification). However, a trend similar to the lesion size progression was observable. Despite these longitudinal changes, *Circularity* stayed stable over time in all subgroups (Table 2).

As a common phenotypic feature, eyes with the highest progression rates revealed multifocal DDAF lesions on a heterogeneous AF background, as found most often in ERG group 2/3 and intermediate/late age of onset eyes (Figure 4 A and B). In contrast, eyes with a monofocal DDAF lesion on a homogeneous background represented the slowest progression rates (Figure 4 C and D).

Prediction of progression rates

For predictive modeling of future progression rates of *Square Root Area*, all visits were evaluable. This corresponds to a median [IQR] follow-up of 2.1 years [1.1; 3.0]. Hereby, multiple putative predictors were initially considered including demographic characteristics (sex, age), disease subgroup characteristics (age of onset, full-field ERG subgroup, presence of a G1961E mutation, presence of a N1868I mutation, presence of a truncating mutation) as well as the multiple lesion-shape-descriptive factors (*Area*, *Perimeter*, *Circularity*, $Feret_{max}$, $Feret_{min}$, *Focality*, and status of the fovea). Backward elimination of non-significant effects eliminated all variables except for the number of foci, the full-field ERG group and the age of onset category as predictors for future *Square Root Area* progression. The leave-one-out cross-validated coefficient-of-determination (R^2) was 0.257, indicating that these predictors allow explaining 25.7 % of the variability in future

DDAF progression rates (Table, Supplemental Digital Content 4, which shows the potential predictors of square-root transformed progression rates).

Inter-rater reliability

The image grading of all unique visits by two independent readers (PLM and IPC) and following automated evaluations of *Area, Square Root Area, Perimeter*, *Circularity*, $Feret_{max}$, $Feret_{min}$, and *Focality* revealed excellent inter-rater reliability (Table, Supplemental Digital Content 5, which shows the values of the inter-rater agreement): The measurements had excellent ICC and low CV as well as CoR (i.e., with 0.95 probability, the absolute differences between two ratings would lie below this value). Also, the Bland-Altman plots did not showed systematic inter-rater discrepancies (Figure, Supplemental Digital Content 6, which demonstrates Bland-Altman plots to visualize the inter-rater agreement). However, the inter-reader variability significantly increased with larger values for measurements of *Area (*$\rho = 0.518$, $P < 0.001$*), Perimeter* ($\rho = 0.226$, $P = 0.002$), *Circularity* ($\rho = 0.375$, $P < 0.001$), and *Focality* ($\rho = 0.308$, $P < 0.001$) according to Spearman's rank correlation coefficient (ρ) for absolute differences and mean values. *Square Root Area* ($\rho = 0.116$, $P = 0.109$), $Feret_{max}$ ($\rho = -0.058$, $P = 0.429$), and $Feret_{min}$ ($\rho = 0.067$, $P = 0.360$) did not show this correlation. In 41 visits (21%), the third reader was required because *Area* grading between both readers differed > 0.15 mm².

Sample size calculation

The observed progression kinetics allowed sample size calculations for future therapeutic trials with an expected reduction to 30%, 40%, 50%, 60% and 70% progression rates of *Square Root-Area* depending on the desired power (Figure 5 A, B, C, D) and trial duration (Figure 5 E, F, G, H). As the progression rates were assumed to be

distributed in a multimodal fashion, we differentiated four different scenarios based on above presented results concerning the predictive value of the baseline data (Appendix, Supplemental Digital Content 7, which illustrates the statistical analysis process from the progression rates to the sample size calculations): I) Evaluating without prior patient selection; II)and III) evaluation with patient selection according to high progression rates (i.e., ERG-based group 2 or 3, intermediate or late age of onset) with II) multifocality of DDAF lesions with more than 6 foci, III) multifocality with 2-5 foci; or IV) patients other than II) and III). The necessary sample size significantly reduced with an extension of the trial duration (with most effect from one to two years) and proposed patient selection (II-IV) giving a more homogenous progression rate with minimized interindividual deviations. To obtain a statistical power of ≥ 0.8 in case I, at least 16, 10, 8, 6 and 5 patients with biallelic DDAF lesions would be required (Figure 5 A). In case II, at least 7, 5, 4, 4 and 4 (Figure 5 B), for III) at least 10, 6, 5, 4 and 4 (Figure 5 C), and for IV) at least 8, 5, 4, 4 and 4 patients would be sufficient (Figure 5 D).

DISCUSSION

We systematically and longitudinally investigated different shape descriptive parameters of RPE atrophy secondary to *ABCA4*-related retinopathy as prognostic biomarkers for disease progression of *ABCA4*-related retinopathy. The presented findings provided evidence for the prognostic relevance of clinical data and shape-descriptive factors at baseline explaining up to 25.7% of the variance in future progression rates of previously unknown patients. With adequate patient selection, future interventional trials could be realized in a more effective manner, as a sufficient statistical power may be achieved with a relatively low sample size.

Validated clinical biomarkers are inevitable for future therapeutic trials. Such markers should be readily captured, reflect the current disease severity and ideally be predictive for long-term progression based on short-term changes. Best corrected visual acuity loss does not constitute a useful endpoint in clinical trials for *ABCA4*-related retinopathy due to its high interindividual variability and phenomena such as foveal non-involvement.[19] Most recent studies, therefore, rely on RPE atrophy assessed by easily accessible AF imaging as morphologic endpoint. Therefore, the presented models focused to predict the progression of complete RPE atrophy defined as DDAF (≥90% darkness of the optic disc). RPE atrophy manifesting as questionable decreased AF (QDAF, 50-90% darkness of the optic disc) was not assessed in this study. Recent studies described QDAF as a transition state (i.e., negative growth rates are possible) and revealed distinctly higher inter-reader variability in the grading of QDAF in comparison to DDAF lesions, making the use of QDAF progression as surrogate marker impractical.[7] The extent of lesions characterized by incomplete RPE atrophy, however, might be expected to correlate to the extent of

complete RPE atrophy. To date, it is difficult to extrapolate from preclinical experimental therapies what component and which stages of human disease may successfully be treated with different approaches. In this context, the use of novel imaging tools like adaptive optics or quantitative autofluorescence might visualize earlier pathognomonic alterations and further reduce necessary sample size.[6,20] However, the potential as clinical endpoint cannot finally be evaluated as longitudinal data are still pending.

The systematic longitudinal evaluation of caliper diameters ($Feret_{max}$ and $Feret_{min}$), cumulative circumference (*Perimeter*), roundness (*Circularity*), and *Focality* of RPE atrophy is novel for *ABCA4*-related retinopathy. As DDAF grading is already part of the eligibility assessment of currently ongoing studies, the evaluation of lesion geometry could be performed fully-automated, allowing for a more accurate prediction of progression rates compared to currently used selection criteria. In this context, it should be considered that *Perimeter* measurements (and hence *Circularity* calculations) depend on contrast of the DDAF lesion. Apart from the fact that the grading with the RegionFinder software is predominantly independent of the image quality,[13] all annotations were equally smoothed before annotation in order to circumvent this limitation in our study.[12] To define limits of agreement in future trials, reading center standard operating procedures will become essential. The caliper diameters seem to reflect the planar extent of severely diseased RPE cells.[12] A further extent of $Feret_{max}$ including also hypo- and hyperautofluorescent flecks would resemble the ‘leading disease front’ as described by Cideciyan and collegues.[21] Interestingly, the described centrifugal expansion rate for the ‘leading disease front’ of 2 degrees per year is markedly higher compared to the here observed increase of the DDAF $Feret_{max}$ diamerter (0.5 degrees [in terms of radius] per year).[21] This is very

much compatible with the typical phenotype of ABCA4-related retinopathy with a markedly greater extent of flecks compared to DDAF.

Our data indicated a distinct relationship between progression rates and baseline area of RPE atrophy, similar to previous findings concerning geographic atrophy secondary to AMD.[12] For the first time in *ABCA4*-related retinopathy, we could show that this quasi-exponential progression kinetics of the lesion size was eliminated with square root transformation. Accordingly, square root transformation abolishes the need to adjust for lesion size and/or limit the inclusion to a specific range of lesion sizes, which might facilitate the conduct of future clinical trials.

Explanations for this relationship might derive from a cascading process at the lesion boundary that drives the RPE atrophy progression in retinal diseases. As a model disease for RPE atrophy early in disease progression, findings in *ABCA4*-related retinopathy might also apply to multifactorial diseases like AMD and vice-versa. Histopathologic studies of *ABCA4*-related retinopathy revealed that lipofuscin accumulation in RPE leads to a change in the cellular homeostasis, integrity and functionality: Intracellular components get displaced or lost and new derivates are formed, visible as flecks at the posterior pole in funduscopy as well as fundus autofluorescence imaging under different excitation wavelength.[22] Impairment of functional integrity is indicated by affected photoreceptor integrity forming deposits associated with progressive loss of layers in OCT images, reduced retinal thickness and increased cone spacing. [4,21] Gradually, RPE undergoes cell death leading to a inhomogeneous pattern of bright and dark flecks being the preliminary stage to atrophic lesions.[22] According, to these findings in *ABCA4*-related retinopathy, histopathologic studies in AMD were able to confirm that the percentage of abnormal RPE

cells with “sloughed,” “rounding”, “shedding,” and “intraretinal” morphologies increases toward the atrophic lesion.[23] In this context, the spreading of heterogeneous background around the atrophic lesions as found in eyes with distinct functional impairment (group 2 and 3) represents the severity of altered RPE cells around the RPE atrophy and is thus associated with the progression rate. Accordingly, a consistent finding in eyes with fast progressing RPE atrophy was widespread hyper- and hypoautofluorescent flecks and multifocal presentation of RPE atrophy (Figure 4 A and B).

Taking these baseline characteristics into account, more homogenous study cohorts can be defined. Based on our presented filters II-IV, we could demonstrate that the sample size and trial duration to allow for significant results of an expected fixed treatment effect could be significantly reduced (Figure 5). These filters comprise parameters that are easily available (demographic characteristics) and/or can be determined fully-automated (lesion geometry). Thereby, upcoming therapeutic trials might be feasible more cost-and time-effective compared with currently used selection criteria. Interestingly, extension of the study length from 1 to 2 years reduces the required sample size by more than a factor of two (Figure 5). This would imply that longer studies may provide a benefit in terms of absolute number of required study visits across all participants and could potentially provide a cost-benefit. The marked reduction in required sample size for longer study may be attributable to the greater absolute difference between the placebo and treatment arm over time and lower variability in progression rates for longer observational periods.

Eyes with foveal non-involvement exhibited a specific pattern of DDAF surrounding the central fovea. The proportion of this specific disease manifestation was untypical high Irrespective of the differentiation between the typical pattern of foveal sparing (i.e., RPE

atrophy surrounding ≥270° of the fovea) and general foveal non-involvement,[16] foveal sparing had been described to usually be found quite infrequent in a general Stargardt disease population.[24] However, the restriction to clearly demarcated DDAF lesions in this study might explain the high proportion of patients with late age-of-onset and hence foveal non-involvement.[10] Specific environmental, metabolic or genetic factors might account for the long-term preservation of foveal integrity. Of note, the mild N1868I variant was highly correlated with foveal non-involvement in our study. The higher progression rate in this subgroup despite smaller baseline lesion size might derive from the higher proportion of higher age-of-onset as well as *Focality* values (both were shown to be significant prognostic variables) compared to eyes with foveal involving RPE atrophy. The additional growth at the posterior border of RPE atrophy seems unlikely to account for these findings, as the centripetal spread was visually slower compared to the centrifugal progression, similar to directional kinetics of RPE atrophy in AMD (Figure 2).[16] Initial foveal non-involvement was lost in 22.5% over the observation time. The median time intervall to this loss (from baseline) was higher than the observation time of the majority of patients with initially foveal non-involvement, as the follow-up time was individually different. Hence, the true percentage of loss of foveal non-involvement might be expected to be higher. Nevertheless. slightly lower values have been found in a previous study, probably due to shorter follow-up time.[10] Interestingly, patients with AMD have been described to have higher conversion rates along with higher progression rates.[25] Giving the slower centripetal spread and the associated foveal preservation in the majority over time, a survival analysis revealed that the visual impairment in the cohort with foveal non-involvement might be expected notably later in life (Figure 3).[10]

Concerning limitations of our study, small study cohorts are common as *ABCA4*-related retinopathy ranks among the rare retinal diseases. The resulting low statistical power could have precluded further correlations of investigated parameters with progression rates. Given that we found an explanation of 25.7% of the variance in future progression rates of previously unknown patients, possible prognostic factors (i.e. possible genotype correlations) for the majority of variance could not be detected. Further insight might derive from larger multi-center natural history study on Stargardt disease/*ABCA4*-related retinopathy. Besides, the use of short-wavelength AF has to focus increasing concerns due to the theoretical phototoxicity mediated by the lipofuscin accumulation. To date, this effect has only been observed in cell cultures and *Abca4*$^{-/-}$ mice.[26] Nevertheless, reduced light energy by reduced illuminance (RAFI) or longer excitation wavelength (green and near-infrared fundus autofluorescence) was proposed and has been shown to prevent light damage in cell culture and distinctly increases patient comfort during the examination.[27] Of note, the radiant exposure of commercial AF devices as used in this study was shown to lie below the standard limits defined by ISO 15004-2 (regulatory standard for Light Hazard protection), and falls into the category ‘ophthalmic instruments for which no potential light hazard exists’, as no human harm has been shown. Further, the expected macular pigment shadowing under short-wavelength condition has been shown to be insignificant in ABCA4-related retinopathy,[7] due to the overall high intensity of short-wavelength AF.[6,20] Wavelength-dependent light absorption of optical media opacities (e.g. cataract) could also be neglected as exclusively patients with clear ocular media were included.

As the choroid and the RPE cell monolayer represent a functional complex, choroidal changes resolved using OCT imaging might effect RPE atrophy progression. Previous studies showed most severe choroidal alterations in widespread *ABCA4*-related

retinopathy that revealed higher progression rate in our cohort.[28,29] Diffusible factors from the RPE sustaining the choroidal structure were hypothesized. Accordingly, reduced secretion of vascular endothelial growth factor from the RPE leads to choriocapillaris atrophy with subsequent atrophy of the RPE and photoreceptors in the mouse model.[30] A multimodal comparison of the actual effect of OCT layer alteration on RPE atrophy progression rate is worth analysis in the future. Machine learning approaches might allow for an even more precise prediction of progression rates in this context.

The findings confirm the relevance of demographic, functional, genetic, and shape-descriptive factors as prognostic variables for RPE atrophy progression secondary to *ABCA4*-related retinopathy. Including these parameters in patient selection would allow for a more accurate prediction of progression rates over time and a reduction of the sample size for future interventional trials compared with currently used selection criteria. Substantial remaining variations in RPE atrophy progression rates imply the existence of further, yet unknown prognostic factors.

REFERENCES

1. Birtel J, Eisenberger T, Gliem M, et al. Clinical and genetic characteristics of 251 consecutive patients with macular and cone/cone-rod dystrophy. *Sci Rep*. 2018;8(1):4824.

2. Sparrow JR, Gregory-Roberts E, Yamamoto K, et al. The bisretinoids of retinal pigment epithelium. *Prog Retin Eye Res*. 2012;31(2):121-135.

3. Delori FC, Staurenghi G, Arend O, Dorey CK, Goger DG, Weiter JJ. In vivo measurement of lipofuscin in Stargardt's disease--Fundus flavimaculatus. *Invest Ophthalmol Vis Sci*. 1995;36(11):2327-2331.

4. Dysli C, Müller PL, Birtel J, Holz FG, Herrmann P. Spectrally Resolved Fundus Autofluorescence in ABCA4 -Related Retinopathy. *Investig Opthalmology Vis Sci*. 2019;60(1):274-281.

5. Hoyng CB, Lambertus S, Bax NM. Stargardt Disease. In: Querques G, Souied EH, eds. *Macular Dystrophies*. Springer New York; 2016:25-30.

6. Müller PL, Gliem M, Mangold E, et al. Monoallelic ABCA4 Mutations Appear Insufficient to Cause Retinopathy: A Quantitative Autofluorescence Study. *Invest Ophthalmol Vis Sci*. 2015;56(13):8179-8186.

7. Müller PL, Pfau M, Mauschitz MM, et al. Comparison of Green Versus Blue Fundus Autofluorescence in ABCA4 -Related Retinopathy. *Transl Vis Sci Technol*. 2018;7(5):13.

8. Lambertus S, Bax NM, Fakin A, et al. Highly sensitive measurements of disease progression in rare disorders: Developing and validating a multimodal model of retinal degeneration in Stargardt disease. Wedrich A, ed. *PLoS One*. 2017;12(3):e0174020.

9. 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):1232-1241.

10. Lambertus S, Lindner M, Bax NM, et al. Progression of Late-Onset Stargardt Disease. *Investig Opthalmology Vis Sci*. 2016;57(13):5186–5191.

11. Csaky KG, Richman EA, Ferris FL. Report from the NEI/FDA Ophthalmic Clinical Trial Design and Endpoints Symposium. *Invest Ophthalmol Vis Sci*. 2008;49(2):479-489.

12. Pfau M, Lindner M, Goerdt L, et al. Prognostic Value of Shape-Descriptive Factors for the Progression of Geographic Atrophy Secondary to Age-Related Macular Degeneration. *Retina*. 2019;39(8):1527-1540.

13. Kuehlewein L, Hariri AH, Ho A, et al. Comparison of Manual and Semiautomated Fundus Autofluorescence Analysis of Macular Atrophy in Stargardt Disease Phenotype. *Retina*. 2016;36(6):1216-1221.

14. Schmitz-Valckenberg S, Brinkmann CK, Alten F, et al. Semiautomated Image Processing Method for Identification and Quantification of Geographic Atrophy in Age-Related Macular Degeneration. *Investig Opthalmology Vis Sci*. 2011;52(10):7640-7646.

15. Lois N, Holder GE, Bunce C, Fitzke FW, Bird AC. Phenotypic subtypes of Stargardt macular dystrophy-fundus flavimaculatus. *Arch Ophthalmol*. 2001;119(3):359-369.

16. Lindner M, Böker A, Mauschitz MM, et al. Directional Kinetics of Geographic Atrophy Progression in Age-Related Macular Degeneration with Foveal Sparing. *Ophthalmology*. 2015;122(7):1356-1365.

17. Giolo SR. *Turnbull's Nonparametric Estimator for Interval-Censored Data'*.; 2004.

18. Paux G, Dmitrienko A. Mediana: Clinical Trial Simulations. R package version 1.0.8. 2019.

19. Rotenstreich Y, Fishman GA, Anderson RJ. Visual acuity loss and clinical observations in a large series of patients with Stargardt disease. *Ophthalmology*. 2003;110(6):1151-1158.

20. Cideciyan A V, 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.

21. Cideciyan A V., 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.

22. Gomes NL, Greenstein VC, Carlson JN, et al. A comparison of fundus autofluorescence and retinal structure in patients with Stargardt disease. *Invest Ophthalmol Vis Sci*. 2009;50(8):3953-3959.

23. Zanzottera EC, Ach T, Huisingh C, Messinger JD, Spaide RF, Curcio CA. Visualizing retinal pigment epithelium phenotypes in the transition to geographic atrophy in age-related macular degeneration. *Retina*. 2016;36 Suppl 1:S12-S25.

24. van Huet RAC, Bax NM, Westeneng-Van Haaften SC, et al. Foveal Sparing in Stargardt Disease. *Investig Opthalmology Vis Sci*. 2014;55(11):7467–7478.

25. Lindner M, Lambertus S, Mauschitz MM, et al. Differential Disease Progression in Atrophic Age-Related Macular Degeneration and Late-Onset Stargardt Disease. *Investig Opthalmology Vis Sci*. 2017;58(2):1001–1007.

26. Teussink MM, Lambertus S, de Mul FF, et al. Lipofuscin-associated photo-oxidative stress during fundus autofluorescence imaging. Lewin AS, ed. *PLoS One*. 2017;12(2):e0172635.

27. Cideciyan A V, Swider M, Aleman TS, et al. Reduced-illuminance autofluorescence imaging in ABCA4-associated retinal degenerations. *J Opt Soc Am A*.

2007;24(5):1457–1467.

28. Müller PL, Fimmers R, Gliem M, Holz FG, Charbel Issa P. Choroidal Alterations in ABCA4-Related Retinopathy. *Retina*. 2017;37(2):359-367.

29. Müller PL, Pfau M, Möller PT, et al. Choroidal Flow Signal in Late-Onset Stargardt Disease and Age-Related Macular Degeneration: An OCT-Angiography Study. *Invest Ophthalmol Vis Sci*. 2018;59(4):AMD122-AMD131.

30. Kurihara T, Westenskow PD, Bravo S, Aguilar E, Friedlander M. Targeted deletion of Vegfa in adult mice induces vision loss. *J Clin Invest*. 2012;122(11):4213-4217.

FIGURE LEGENDS

Figure 1: Progression of RPE atrophy.

Fundus autofluorescence images without (top row) and with (bottom row) delineation of the patient with the longest observation time in our cohort (100 months) are exemplarily demonstrated. In the first 16 months, a unifocal parafoveal atrophic lesion of the retinal pigment epithelium (RPE) was present. After 39 months, a second atrophic spot developed. With further growth of lesion size, both spots coalesced 27 months later leaving a wedge towards the preserved foveal center. The annual progression rate increased with rising RPE atrophy lesion size, up to the point when RPE atrophy reached the foveal center by t_0+85 months.

Figure 2: Presentation of foveal non-involvement over time

Fundus autofluorescence images without (top row) and with (bottom row) delineation of an eye with foveal non-involvement visualized the different directional kinetic with significantly faster centrifugal than centripetal spread of RPE atrophy leading to long lasting preservation of foveal integrity. Values indicate *Area* (top) and *Square Root Area* (bottom).

Figure 3: Course of visual impairment.

Time-to-event curve showing the cumulative fraction of eyes reaching the following clinical endpoints: mild visual impairment (≥0.2 LogMAR, ≤20/32 Snellen equivalent, brown solid line), moderate visual impairment (≥0.6 LogMAR, ≤20/80 Snellen equivalent, pink dashed line), and severe visual impairment (≥1.0 LogMAR, ≤20/200 Snellen equivalent, blue

dotted line). Analysis was done for all eyes in the dataset (A) and eyes having no visual impairment (≤0.1 LogMAR, ≥20/25 Snellen equivalent) at the beginning of the study (B).

Figure 4: Characteristics of eyes with extreme progressing rates.

Fundus autofluorescence images without (top row) and with (bottom row) delineation of two eyes are demonstrated that revealed high (A and B, first row) and low (C and D, second row) progression rates of RPE atrophy. Values indicate *Area* (top) and *Square Root Area* (bottom). Both fast progressing eyes were assigned to ERG group 2 and intermediate (A) or late (B) age of onset. Multiple spots of RPE atrophy with rough borders, and a heterogeneous background with a multitude of hypoautofluorescent flecks were typical findings in eyes with fast progressing lesion sizes. Both slow progressing eyes were assigned to ERG group 1 and revealed the mild G1961E (C) or N1868I (D) variant. Unifocal RPE atrophy with smooth borders, and a quite homogeneous background with hyperautofluorescent flecks were typical findings in eyes with slow progressing lesion sizes.

Figure 5: Statistical power as a function of sample size for all patients.

Sample size calculations are shown for an expected treatment effect of 70 %, 60 %, 50 %, 40 %, and 30 % reduction of *Square Root Area* progression in patients with bilateral DDAF lesions for a trial duration of 2 years (A-D). Additionally, the minimal sample size required for a statistical power greater or equal to 80 % for the aforementioned expected treatment effects as a function of treatment duration are shown (E-H). Symbols indicate data points predicted by the clinical trial simulation, the respective lines show a fit of the corresponding

data points. The minimal sample sizes for E-H were calculated based on the fits to achieve a more fine-grained resolution. A and E show the simulation results for all patients, B and F for patients with intermediate or late age of onset and retina-wide photopic or photopic and scotopic functional impairment and 6 or more foci, C and G for patients with the aforementioned characteristics but with 2-5 foci and D and H for patients not within these filters. The statistical analysis process of the sample size calculation can be found in the Appendix, Supplemental Digital Content 7.

LIST OF SUPPLEMENTAL DIGITAL CONTENT

1. Supplemental Digital Content 1. Figure which illustrates the measurements of the shape-descriptive factors. pdf

2. Supplemental Digital Content 2. Table which shows the individual baseline and progression data of all included eyes. pdf

3. Supplemental Digital Content 3. Figure which illustrates the statistical analysis of the progression rates. pdf

4. Supplemental Digital Content 4. Table which shows the potential predictors of square-root transformed progression rates. pdf

5. Supplemental Digital Content 5. Table which shows the values of the inter-rater agreement. pdf

6. Supplemental Digital Content 6. Figure which demonstrates Bland-Altman plots to visualize the inter-rater agreement. pdf

7. Supplemental Digital Content 7. Appendix which illustrates the statistical analysis process from the progression rates to the sample size calculations. pdf