Automatic Drusen Quantification and Risk Assessment of Age-Related Macular Degeneration on Color Fundus Images

Mark J. J. P. van Grinsven,[1] Yara T. E. Lechanteur,[2] Johannes P. H. van de Ven,[2] Bram van Ginneken,[1] Carel B. Hoyng,[2] Thomas Theelen,[2] and Clara I. Sánchez[1]

[1]Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
[2]Department of Ophthalmology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

Correspondence: Thomas Theelen, Radboud University Nijmegen Medical Centre, Department of Ophthalmology (409), Philips van Leydenlaan 15, 6525 EX Nijmegen, The Netherlands; t.theelen@ohk.umcn.nl

MJJPvG and YTEL contributed equally to the work presented here and should therefore be regarded as equivalent and joint first authors. TT and CIS should be considered joint last authors.

Submitted: December 6, 2012
Accepted: April 1, 2013

Citation: van Grinsven MJJP, Lechanteur YTE, van de Ven JPH, et al. Automatic drusen quantification and risk assessment of age-related macular degeneration on color fundus images. *Invest Ophthalmol Vis Sci*. 2013;54:3019–3027. DOI:10.1167/iovs.12-11449

Purpose. To evaluate a machine learning algorithm that allows for computer-aided diagnosis (CAD) of nonadvanced age-related macular degeneration (AMD) by providing an accurate detection and quantification of drusen location, area, and size.

Methods. Color fundus photographs of 407 eyes without AMD or with early to moderate AMD were randomly selected from a large European multicenter database. A machine learning system was developed to automatically detect and quantify drusen on each image. Based on detected drusen, the CAD software provided a risk assessment to develop advanced AMD. Evaluation of the CAD system was performed using annotations made by two blinded human graders.

Results. Free-response receiver operating characteristics (FROC) analysis showed that the proposed system approaches the performance of human observers in detecting drusen. The estimated drusen area showed excellent agreement with both observers, with mean intraclass correlation coefficients (ICC) larger than 0.85. Maximum druse diameter agreement was lower, with a maximum ICC of 0.69, but comparable to the interobserver agreement (ICC = 0.79). For automatic AMD risk assessment, the system achieved areas under the receiver operating characteristic (ROC) curve of 0.948 and 0.954, reaching similar performance as human observers.

Conclusions. A machine learning system capable of separating high-risk from low-risk patients with nonadvanced AMD by providing accurate detection and quantification of drusen, was developed. The proposed method allows for quick and reliable diagnosis of AMD, opening the way for large dataset analysis within population studies and genotype-phenotype correlation analysis.

Keywords: age-related macular degeneration, drusen detection, risk assessment

Age-related macular degeneration (AMD) is the leading cause of irreversible vision loss in developed countries among individuals older than 50 years.[1] AMD is a gradual progressive disease that evolves from early and intermediate stages, with no or subtle visual changes, to an advanced stage, where the loss of central vision can occur. Patients with intermediate AMD are at higher risk of developing advanced AMD and thus suffering from severe visual loss, and they should undergo routine- and self-monitoring for a timely diagnosis.[2] Lifestyle changes such as cessation of smoking and prophylactic regimen like vitamin supplementation are recommended for patients at risk in order to slow progression of the disease.[3–6]

Deposits of extracellular material localized between the inner collagenous layer of Bruch's membrane and the basal lamina of the RPE, known as drusen, are considered the hallmark feature of AMD.[7] Macular drusen are important in the context of AMD grading, and certain drusen characteristics are associated with progressing toward end-stage AMD.[8–14] On fundus photography they appear as yellowish-white spots, and different drusen phenotypes can be distinguished. Hard drusen are defined as small (<63 μm) nodular lesions with well-defined borders. Soft drusen, on the other hand, tend to be larger, and are generally characterized by poorly demarcated boundaries.[7–9,12,15]

Identification and classification of eyes with AMD are performed mainly using color fundus images by manually determining the size and extension of drusen.[8,12,16–19] However, other imaging modalities, such as optical coherence tomography, are gaining traction as well.[20,21] Human observer classification is time-consuming and prone to interobserver variations.[22] Aside from speed, objectivity, and reproducibility, implementation of an automatic drusen detection and quantification system could prove useful in many ways. It may allow for a cost-efficient screening program for patients at risk and help identify and classify AMD patients in large cohort studies. Additionally, accurate quantitative measurements can help in large clinical studies for the evaluation and progression of drusen area, for example, in clinical trials concerning new therapeutic strategies for dry AMD, and it could help in applying inclusion criteria for large-scale clinical studies and genotype-phenotype correlation analysis.[23]

Table 1. Criteria for Grading AMD According to the CIRCL Grading Protocol

AMD Stage	Criteria
1. No AMD	No drusen or small, hard drusen only.
2. Early AMD	>10 small (<63 μm), hard drusen + pigmentary changes or 1–15 intermediate (63–124 μm) drusen.
3. Intermediate AMD	>15 intermediate (63–124 μm) drusen or any large (≥125 μm) drusen or GA not in the central circle of the ETDRS grid.
4. Advanced AMD (GA)	Presence of central GA.
5. Advanced AMD (CNV)	Evidence of active or previous CNV lesion.
6. CNV without signs for AMD	Chosen if CNV is present but no drusen of any size are present within the Field 2.
7. Cannot grade	Image is regarded as not gradable.

AMD, age-related macular degeneration; CIRCL, Cologne Image Reading Center and Laboratory; CNV, choroidal neovascularization; ETDRS, Early Treatment Diabetic Retinopathy Study; GA, geographic atrophy.

Druse size is measured as the diameter of the smallest enclosing circle of the druse.

Previously proposed methods automatically assessed the presence of drusen on color fundus photographs.[24,25] However, the presence of drusen alone is not directly correlated with the risk of progression to advanced AMD.[2] Other works focused on the automatic quantification of drusen without identifying patients at high risk or the AMD stage.[23,26–32] Here, we describe and evaluate a machine learning algorithm that automatically distinguishes between images from low-risk and those from high-risk AMD patients by providing an accurate quantification of drusen location, area, and size.

Methods

Study Dataset

A total of 407 images of different eyes with nonadvanced stages of AMD (i.e., stages 1, 2, and 3 according to the criteria shown in Table 1), with sufficient grading quality for human graders, was selected in consecutive fashion from the European Genetic Database (EUGENDA, http://www.eugenda.org), a large multicenter database for clinical and molecular analysis of AMD.[33,34] For each subject, images of both eyes were eligible for inclusion, but we did not select multiple images of the same eye. Images with presence of reticular pseudodrusen were excluded from analyses. Number of drusen, age, or ethnicity was not taken into account for the selection of data. Written informed consent was obtained before enrolling patients in EUGENDA. The study was performed according to the tenets set forth in the Declaration of Helsinki, and Investigational Review Board approval was obtained.

Digital nonstereoscopic color fundus photographs were acquired with a TRC 501X model digital fundus camera at 50° (Topcon Corp., Tokyo, Japan) or with a CR-DGi model nonmydriatic retinal camera at 45° (Canon, Inc., Tokyo, Japan), and pupil dilation was achieved with topical 1.0% tropicamide and 2.5% phenylephrine. All images were macula-centered. Image size varied from 1360 × 1024 to 3504 × 2336 pixels. Before analysis, images were resized in a preprocessing step to have a field of view with a standardized diameter of 630 pixels independently of the image resolution. The data were divided randomly into two sets: set A, consisting of 52 images, for the evaluation of automatic drusen quantification, and set B, consisting of 355 images, for the evaluation of automatic risk assessment. Images from the same patients were kept in the same set.

Observer Annotations

Resampled images were displayed on an LCD monitor similar to those used in ophthalmology practice and with the ability to zoom and pan. All visible drusen were manually outlined in set A by both observers using a specifically developed annotation tool. Whether confluent drusen were annotated as separate drusen or as one large drusenoid patch was left to the judgment of the observers. Two trained graders (JPHV, designated Observer 1, and YTEL, designated Observer 2) manually performed a risk assessment to develop late-stage AMD in all images of sets A and B. No AMD and early AMD were defined as low-risk stages, and intermediate AMD was considered high-risk stage. AMD was defined according to the standard protocol of the Cologne Image Reading Center and Laboratory (CIRCL) (Table 1),[33] an AMD classification that was adopted specifically for EUGENDA based on different international staging systems.[8,16–19] Both observers performed one session of AMD grading for each study eye in set B and performed one session of drusen annotation for each study eye in set A on a different day.

Machine Learning Algorithm

The proposed CAD system analyzed all color fundus images to automatically quantify the visible drusen and assigned each image a probability between 0 and 1, with 1 indicating a high risk of developing advanced AMD and 0 indicating low risk. To accomplish this, the system performed the following steps:

1. Druse candidate extraction: Each pixel in the image was assigned a probability that the pixel was part of a bright lesion structure, using a supervised pixel classification method. Supervised classification is a machine learning technique where manually labeled training examples are used to infer the classification rule.[30] Neighboring pixels with similar probability, not located close to the automatically detected optic disc,[30] were grouped into druse candidates.
2. Druse candidate segmentation: The boundary of each druse candidate was automatically delineated using intensity and contrast characteristics.
3. Druse candidate classification: Druse characteristics and a supervised lesion classification method were used to assign a probability to each segmented candidate which indicated the likelihood of being a true druse, creating a so-called drusen probability map.
4. AMD risk assessment: Based on the drusen probability map, a supervised image classification method assigned each image a probability to be at high risk of developing advanced AMD.

Figure 1 shows the steps of the CAD algorithm. The classification steps in the system were performed using statistical classifiers that could differentiate between different types of pixels, candidates, or images by using a training set of labeled examples and extracting numerical characteristics

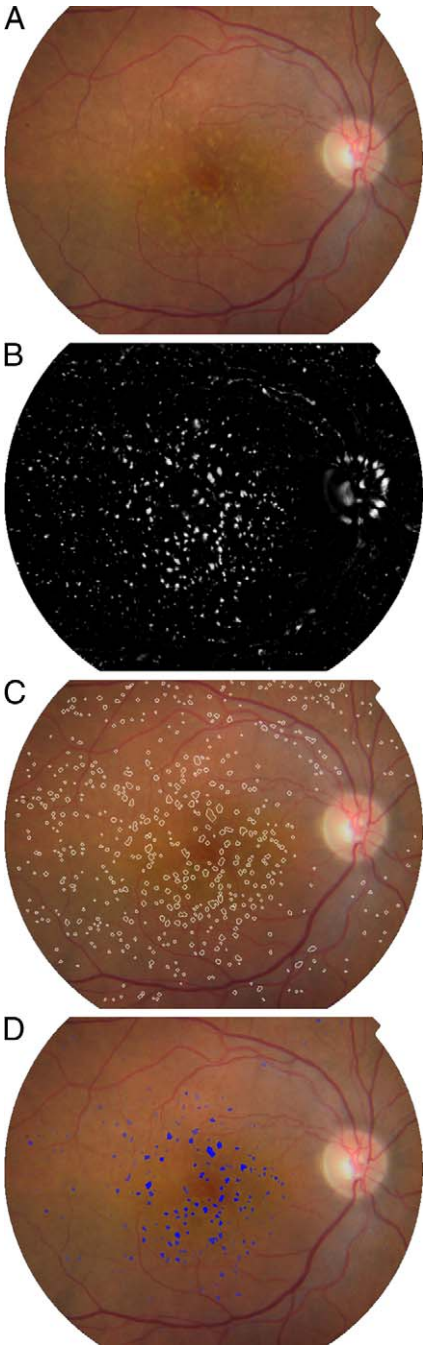

FIGURE 1. Example of outputs obtained in each step of the proposed CAD system. (**A**) Original color fundus image. (**B**) Each pixel was assigned a probability of being part of a bright structure after the druse candidate extraction step. A higher intensity indicates a higher probability. (**C**) The boundary of each druse candidate (shown overlaid on the original image) was delineated during the drusen candidate segmentation step. (**D**) Candidates were classified as true drusen in the drusen candidate classification step. The final detected drusen are shown overlaid on the original image. Brighter color represents a higher probability of being a true druse.

(features).[35] Several supervised classifiers were tested for each step, and the classifiers that performed best were chosen, namely a k-nearest neighbor (kNN) classifier[35] for step 1, a linear discriminant (LDA) classifier[35] for druse candidate classification (step 3), and a random forest (RF) classifier[36] for step 4. A more detailed description of the CAD system can be found in Appendix A. Given an image, the CAD system provides two outputs: (1) detection of all visible drusen in the image and quantification of the drusen area and maximum druse diameter; and (2) a probability indicating the likelihood that the patient was at high risk of developing advanced AMD based on the drusen probability map.

Data Analysis

To evaluate the proposed CAD system, two types of analyses were performed: (1) evaluation of automatic drusen quantification; and (2) evaluation of automatic AMD risk assessment. Due to the lack of a single gold standard, each evaluation was performed twice, taking Observer 1 as reference standard and comparing the CAD results with those obtained by Observer 2, and vice versa.

For drusen quantification, a fivefold cross-validation approach was performed to train and test the CAD system by using data from set A. Cross-validation analysis allows determination of the system performance in an unbiased manner.[37] Using the test folds, the lesion sensitivity (fraction of drusen marked in the reference standard that were detected as drusen by the CAD system) and the number of false positives per image were calculated after setting a threshold for the druse probabilities obtained in step 3. Varying this threshold, different lesion sensitivity-false positives pairs were calculated and summarized in a free receiver operating characteristic (FROC)[38] to evaluate the CAD performance on the detection of drusen. The observer performance compared to the reference standard, which corresponds with one lesion sensitivity-false positives pair, was also calculated and included in the obtained FROC curve.

Total drusen area and maximum druse diameter obtained by the CAD system were calculated using the distance between the fovea and the border of the optic disc as a reference distance of 3000 μm.[8,12] After thresholding the drusen probability map, total drusen area and maximum druse diameter were measured and compared to the observers' opinions, using intraclass correlation coefficient (ICC) analysis.[39] This threshold was set at the same false-positives rate as Observer 1, as this observer had fewer false positives than Observer 2. During the analysis, the performance of the CAD system and the observers of drusen quantification were evaluated inside and outside the Early Treatment Diabetic Retinopathy Study (ETDRS) grid,[40] which was manually set before the analysis.

For AMD risk assessment, a leave-one-out cross-validation approach[37] was performed to train and test the CAD system, using data from set B. The leave-one-out cross-validation allows measurement of the predictive performance measure of a statistical model by testing a single sample while training with the remaining samples. This is repeated such that each sample is used once as test data. Using the test folds, image sensitivity (fraction of images correctly classified by the CAD system in the high-risk stage) and image specificity (fraction of images correctly classified by the CAD system in the low-risk stage) were calculated after setting a threshold for the estimated risk obtained in step 4. Varying this threshold, different image sensitivity-image specificity pairs were calculated and summarized in a receiver operating characteristic[38] (ROC) to evaluate the CAD performance of distinguishing between low-risk and high-risk patients. The area (A_z) under the ROC was used as a measure of performance. The observer performance compared to the reference standard, which corresponds with one image sensitivity-image specificity pair, was also calculated and included in the obtained ROC curve. Overall agreement on risk assessment between the observers was calculated using κ statistics (version 17.0.0 software; SPSS, Chicago, IL).

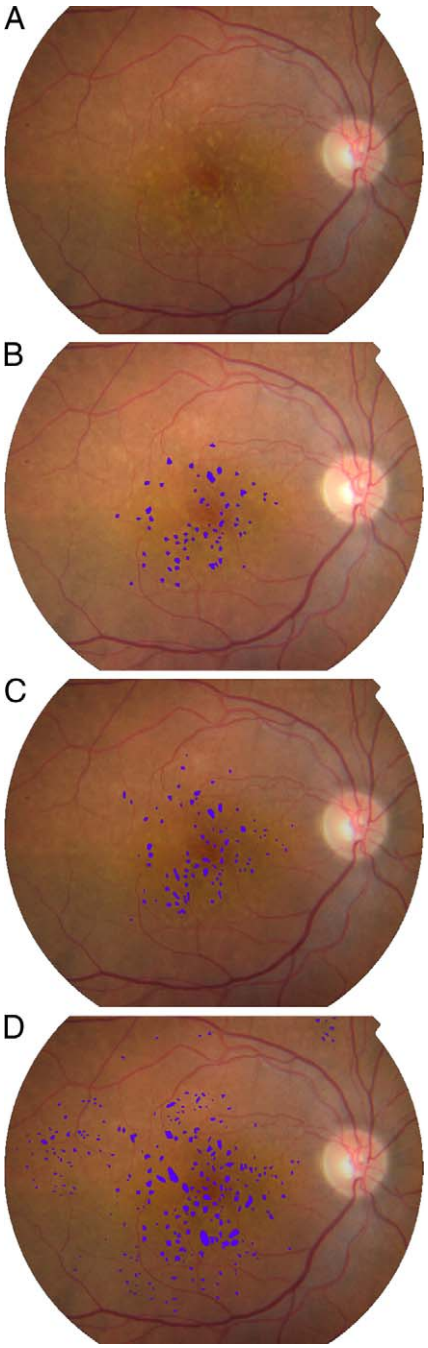

FIGURE 2. (**A**) Original color fundus image. (**B**) Drusen detected by the CAD system overlaid on the original image. (**C**) Drusen annotated by Observer 1 on the original image. (**D**) Drusen annotated by Observer 2 on the original image.

RESULTS

Of 407 images, 145 were captured with the Topcon camera, and 262 were captured with the Canon camera. Table 2 shows some statistics of the performed observer annotations for AMD risk assessment and drusen quantification.

Drusen Quantification

Figure 2 shows the drusen automatically detected by the CAD system from a sample image and shows the annotations of the

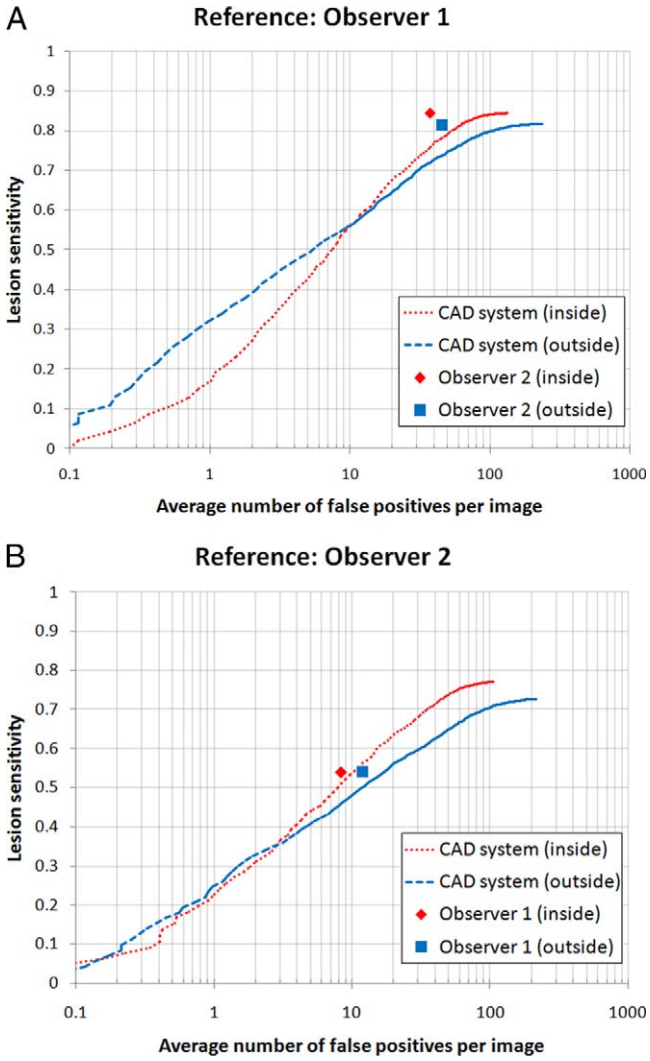

FIGURE 3. FROC curves for the CAD system inside and outside the ETDRS grid, considering Observer 1 (**A**) and Observer 2 (**B**) as reference standards. The corresponding observer performance compared to the reference standard is also plotted as a point in the graph.

observers. Figures 3A and 3B show the FROC curves for the CAD system inside and outside the ETDRS grid, using Observer 1 and Observer 2 as reference standard, respectively.

Tables 3 and 4 summarize the mean drusen area and maximum druse diameter obtained by the CAD system and the observers. The corresponding ICCs are shown in Figures 4A and 4B. For estimated drusen area, ICCs of 0.91 and 0.86 were obtained for the CAD system compared to Observer 1 and Observer 2, respectively, whereas the interobserver agreement reached an ICC equal to 0.87. The CAD system showed similar agreement with the observers independently of the camera used for the acquisition, reaching ICC values of 0.80 and 0.88 on images acquired with the Topcon digital fundus camera at 50° and the Canon nonmydriatic retinal camera at 45°, respectively. For the estimation of maximum druse diameter, defined as the diameter of the smallest enclosing circle of a druse, the agreement with the observers was lower with a maximum ICC of 0.69, whereas observers had an agreement with ICC of 0.79.

Investigative Ophthalmology & Visual Science

Table 2. Summary of the Manual Annotations Performed on Sets A and B by the Two Observers

	Set A		Set B	
Annotation	**Observer 1**	**Observer 2**	**Observer 1**	**Observer 2**
Risk assessment				
No AMD	17	20	216	218
Early AMD	13	9	64	64
Intermediate AMD	22	23	75	76
Drusen quantification				
Average number of drusen	130.4 ± 178.1	198.5 ± 243.1	-	-
Average size of drusen, μm^2	5,873 ± 10,027	5115 ± 8257	-	-

AMD, age-related macular degeneration.
No AMD and early AMD are defined as low-risk stages and intermediate AMD as high-risk stage.
Average number of drusen is the average number of annotated drusen per image. Average size of drusen is the average size of annotated drusen.

Table 3. Mean Area and Percentage Covered by Drusen Inside and Outside the ETDRS Grid and in the Total Image

Coverage	CAD	Observer 1	Observer 2
Inside grid			
Mean area, mm^2	0.43 ± 0.57	0.44 ± 0.68	0.56 ± 0.73
Area, %	1.52 ± 2.01	1.55 ± 2.40	1.98 ± 2.58
Outside grid			
Mean area, mm^2	0.35 ± 0.70	0.33 ± 0.72	0.46 ± 0.81
Area, %	0.36 ± 0.73	0.34 ± 0.75	0.49 ± 0.87
Total image			
Mean area, mm^2	0.78 ± 1.00	0.77 ± 1.07	1.01 ± 1.21
Area, %	0.67 ± 0.86	0.65 ± 0.96	0.89 ± 1.16

Table 4. Mean Maximum Druse Diameter (mm) and Pixels Inside and Outside the ETDRS Grid and in the Total Image

Coverage	CAD	Observer 1	Observer 2
Inside grid			
Maximum diameter, pix	11.54 ± 6.74	13.00 ± 12.53	11.79 ± 10.06
Maximum diameter, mm	0.21 ± 0.12	0.23 ± 0.23	0.21 ± 0.17
Outside grid			
Maximum diameter, pix	10.16 ± 6.30	7.15 ± 7.22	8.91 ± 10.06
Maximum diameter, mm	0.18 ± 0.12	0.13 ± 0.14	0.16 ± 0.15
Total image			
Maximum diameter, pix	13.88 ± 6.27	14.67 ± 12.33	13.74 ± 10.21
Maximum diameter, mm	0.25 ± 0.12	0.27 ± 0.23	0.25 ± 0.18

pix, pixel.

AMD Risk Assessment

Figures 5A and 5B show ROC curves using Observer 1 and Observer 2 as reference standards, obtaining A_z values of 0.948 and 0.954, respectively. Observer 2 reached an image sensitivity of 0.85 and an image specificity of 0.96, as shown in Figure 4, whereas Observer 1 obtained an image sensitivity of 0.84 and image specificity of 0.96 (Fig. 5). Table 5 shows the contingency table and kappa (κ) agreement between the observers and between the CAD system and the observers for AMD risk assessment. The threshold for the CAD system was set at the cutoff point that maximizes sensitivity + specificity.

Discussion

In this study, a supervised machine learning algorithm for automated AMD classification based on drusen identification and quantification, was developed. Our system was able to perform equally as experienced human graders with respect to AMD risk assessment, drusen localization, and determination of mean drusen area with a dataset considerably larger than those used in previous publications.[25–27,29,31,41–46]

Detecting drusen on color fundus images is a challenging task, as shown by the differences in observer annotations in Figure 2. These differences illustrate the need for a robust and accurate system for drusen detection. This would help in eliminating intra- and interobserver variability and the subjective character of manual drusen detection. For this reason, automated drusen detection on color fundus photographs has been a field of interest for the last couple of decades. However, many systems still require human adjustments or close supervision by experts and are therefore still amenable to subjective input.[26,31,32,44–46] Unsupervised automatic detection systems have been developed, but most have failed to

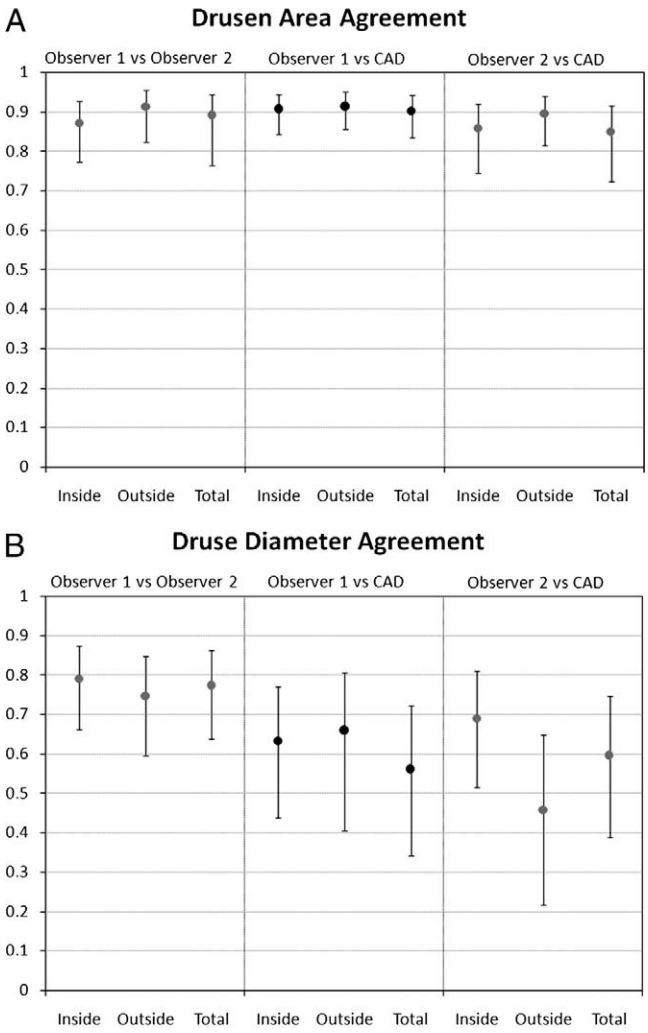

FIGURE 4. Intraclass correlation coefficient (ICC) values with *error bars* indicating 95% confidence interval for drusen area (**A**) and maximum druse diameter (**B**) between the CAD system and the observers. ICC values are calculated for the values obtained inside and outside the ETDRS grid, as well as for the total image.

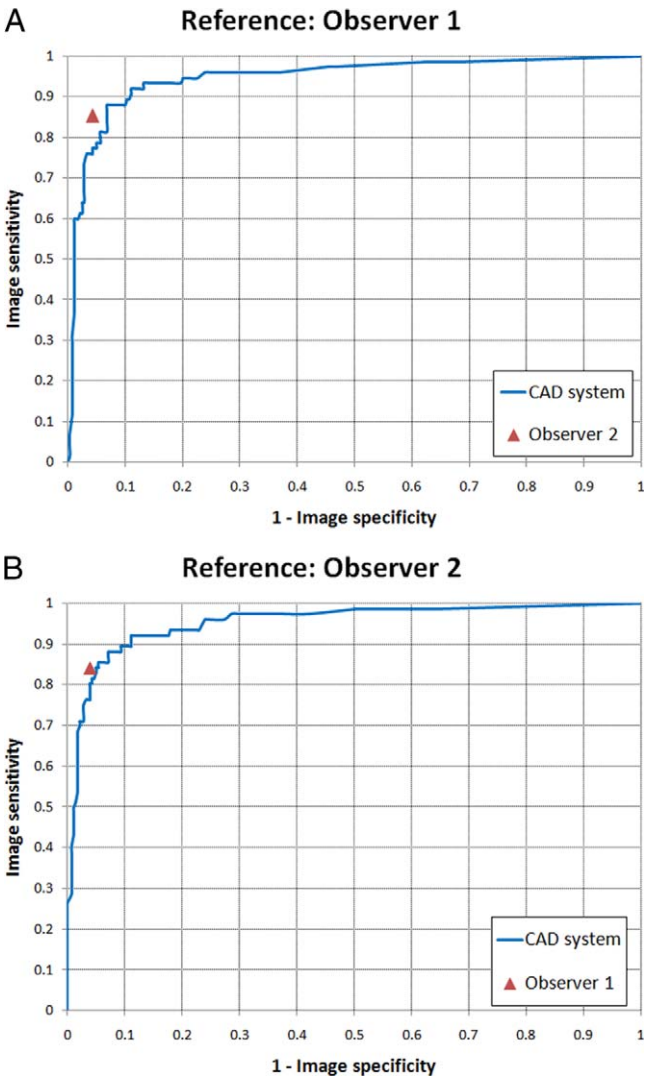

FIGURE 5. ROC curves for the CAD system considering Observer 1 (**A**) and Observer 2 (**B**) as reference standards, respectively. The corresponding observer performance compared to the reference standard is also plotted as a point in the graph.

achieve acceptable performances compared to human graders[41–43] or are only able to give categorized outcome values.[25,47]

In addition to detection, accurate localization and segmentation of drusen are very important to adequate quantification of drusen load in an image. In contrast to other methods,[27,29] where the performance analysis was carried out by pixel-to-pixel comparison, we performed FROC analysis,[38,48] stressing the importance of a correct localization and segmentation of individual lesions and providing higher statistical power than conventional ROC analysis for this task.[48]

With respect to quantification of the total drusen area, there is high agreement between the observers and the proposed CAD system (Fig. 4). The CAD system also showed similar agreement with the observers independently of the camera used for the acquisition. However, a more exhaustive analysis should be made in order to evaluate the effect of the image quality of manual and automatic drusen quantification. In a previously proposed drusen quantification method,[27] a slightly higher ICC value was reported (ICC = 0.92) than the ground truth based on the average grading of eight experts. However, in that study, images with the highest variability among observers were excluded from the study. This was the case for five images, resulting in the exclusion of more than 20% of the total dataset, which is likely to influence the outcome.

For the estimation of maximum druse diameter, agreement between the CAD system and the observers was lower, with a maximum ICC value of 0.69. However, the interobserver agreement on this measurement also decreased. These lower values might be explained by the fact that a correct druse diameter depends on accurate druse delineation, which may be hampered by several factors. For human observers, the main problem lies in the analysis and classification of complex morphological patterns that characterize drusen,[29] whereas the CAD system is impeded mostly by low image quality, poor contrast, or neighboring artifacts.

In this study, we also examined AMD risk assessment, and we showed that the CAD system performs as well as the human observers (Fig. 4, Table 4). Images incorrectly classified by the system in the low-risk stage corresponded mainly to cases of disagreement between the observers (40% of

Table 5. Contingency Table, κ Agreement, and 95% CI for AMD Risk Assessment Between Observer 1 and Observer 2

Assessment	Low Risk	High Risk
Observer 1 vs. Observer 2		
Low risk	268	11
High risk	12	64
κ	0.807	
95% CI	0.731–0.833	
CAD vs. Observer 1		
Low risk	261	10
High risk	19	65
κ	0.765	
95% CI	0.684–0.846	
CAD vs. Observer 2		
Low risk	259	10
High risk	20	66
κ	0.760	
95% CI	0.679–0.841	

CI, confidence interval.

misclassified images) or low quality images where the system was unable to localize low contrast drusen. Other studies have tried to identify AMD with automatic methods.[49,50] However, these were aimed primarily at identifying the presence or absence of disease instead of trying to separate high-risk from low-risk AMD patients, which is clinically more relevant. Zheng et al.[50] developed an algorithm for identification of AMD with a sensitivity of 99.4% and a specificity of 100%. However, the authors compared their CAD system only to a single human observer, which can lead to false high performance.

In contrast to previously published CAD systems,[25–27,29,31,32,42–46] our software performs drusen quantification independently of a fixed region of interest. With full image drusen detection, more information from the image is extracted which can be beneficial if our method would be deployed in clinical studies of AMD. For example, in studies of the cuticular drusen subtype of AMD, diagnosis is based on a typical pattern of innumerable small drusen on fluorescein angiography (FA), not only in the macular region but also in the peripheral retina.[34,51] It would be very valuable to evaluate this drusen pattern to see if regions identified on FA were also detected by the CAD system on color images.

In our system, misclassification of candidates as true drusen often occurred due to reflections of the internal limiting membrane or because of the presence of non–AMD-related abnormalities. Adding better features to characterize these regions or including them as samples in the learning process of the CAD system might solve this problem in the future. Depending on their number and size, false-positive drusen detection might lead to incorrect AMD risk assessment. However, in our study, this did not occur very often. It is possible that these false-positive drusen have a relatively low probability of being a true druse, which is accounted for during computation of AMD risk. In addition, we did not consider pigmentary changes for automatic AMD risk assessment. This could be unfortunate if we wanted to separate patients with early AMD from healthy controls. For detection of patients at high risk of developing late-stage AMD, this distinction is not relevant because no AMD and early AMD were both considered low risk. However, if we wanted to use the system for classifying groups of patients in different AMD stages in new studies, the need for a well-defined control group is high. We will investigate automatic detection of pigmentary changes in upcoming studies.

We are not aware of any implementation of AMD screening programs, but there have been studies evaluating cost effectiveness of such programs.[52,53] However, the proposed programs are based on self-testing, whereas screening based on evaluation of color fundus images would be preferable. Deployment of human graders in such a broad setting would be costly and time consuming, and implementation of an automatic detection system would circumvent these problems. The CAD system could, for example, be installed in opticians' offices and be implemented in routine evaluation of elderly people. High-risk individuals would be selected on site and referred for further ophthalmologic evaluation.

In conclusion, we have developed and evaluated a machine learning system for identification of high-risk AMD patients. Our system allows for accurate detection and quantification of drusen location, area, and size, with a performance equal to human observers under stringent testing conditions. Implementation of our system allows for quick and reliable diagnosis of AMD in screening as well as in research programs. Additionally, there is a need for detailed phenotyping of large datasets in order to gain more insight into risk factors and disease mechanisms involved in AMD.[34] With the use of an automatic detection system, identification of homogeneous AMD subgroups and genotype-phenotype correlations should be achievable in a broader context.[23]

Acknowledgments

Supported by MD-fonds, Oogfonds, Algemene Nederlandse Vereniging ter Voorkoming van Blindheid.

Disclosure: **M.J.J.P. van Grinsven**, None; **Y.T.E. Lechanteur**, None; **J.P.H. van de Ven**, None; **B. van Ginneken**, None; **C.B. Hoyng**, None; **T. Theelen**, None; **C.I. Sánchez**, None

Table 6. Features for Druse Candidate Classification

Feature	Number	Criteria
Shape	1–5	Area, perimeter, compactness, length, and width of the candidate.
Context	6,7	Average and standard deviation of vessel pixel probability at the candidate border.
	8	Distance to the closest candidate.
	9,10	Number and average pixel probability of neighboring candidates in a radius of 50 pixels.
Intensity	11–33	Features measuring the contrast of the candidate in the RGB channels.
	33–81	Mean and standard deviation of Gaussian filter bank outputs.
Color	82–105	Average and standard deviation inside and outside the candidate using the planes of the Luv color space and HSI color space.
Miscellaneous	106–109	Average, standard deviation, maximum and median pixel probability inside the candidate.

HSI, hue-saturation-intensity; Luv, luminescence-saturation-hue angle color space adopted by the International Commission on Illumination (CIE); RGB, red-green-blue.

Appendix A

Druse Candidate Extraction

In this step, pixels that are potentially bright lesion pixels are extracted by convolving the green channel of the color fundus image with a group of Gaussian filters. These filters are based on Gaussian derivatives up to second order at different scales of information.[30] A kNN classifier is then trained to classify every pixel in the image on the basis of the filter responses.[35] No preprocessing of the image is needed previous to the druse candidate extraction step, such as suppression of luteal pigmentation.[26,31,32] After classification, a pixel probability map is obtained that indicates the probability of each pixel to be part of a bright lesion. Neighboring pixels with similar probability were grouped into druse candidates. Algorithms that perform optic disc segmentation and vessel segmentation[30] were also applied in order to remove candidates that overlapped with these anatomical landmarks and to use in further processing.

Druse Candidate Segmentation

In order to find the border of the drusen candidates, dynamic programming[54] is applied around the local maxima of the calculated pixel probability map. During this process, the gradient magnitude of the Gaussian derivatives is used as cost function to guide the algorithm to the candidate borders.

Druse Candidate Classification

In order to determine whether a druse candidate is a true druse or not, a classification step using an LDA is performed.[35] For each druse candidate, a total of 109 features based on color, intensity, contextual information, and shape are extracted (Table 6).[30] These features exploit the different characteristics that the drusen show in color fundus images.

AMD Risk Assessment

To separate high-risk from low-risk patients, a weighted histogram of the calculated drusen probabilities in the image is created to encode drusen extension and size. The value b_n of the histogram bin n is defined as:

$$b_n = \sum_{i \in L_n} p_i$$

where p_i is the posterior probability of druse candidate i, and L_n is the group of candidates whose size is $\tau n \leq d_i < \tau(n+1)$, in which d_i is the size (μm) of candidate i. Terms τ and n control the bin size and the histogram resolution, respectively, and values were chosen as $n = 0, \ldots, 36$ and $\tau = 10$ μm. The last bin ($n = 36$) takes all candidates with sizes d_i larger than 360 μm into account. A random forest (RF) classifier[36] is then trained by using the histogram bins as features to distinguish high-risk patients.

References

1. Jager RD, Mieler WF, Miller JW. Age-related macular degeneration. *N Engl J Med*. 2008;358:2606–2617.
2. Seddon JM, Reynolds R, Yu Y, Daly MJ, Rosner B. Risk models for progression to advanced age-related macular degeneration using demographic, environmental, genetic, and ocular factors. *Ophthalmology*. 2011;118:2203–2211.
3. Age-Related Eye Disease Study Research Group. A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and E, beta carotene, and zinc for age-related macular degeneration and vision loss: AREDS report no. 8. *Arch Ophthalmol*. 2001;119:1417–1436.
4. Evans JR. Antioxidant vitamin and mineral supplements for slowing the progression of age-related macular degeneration. *Cochrane Database Syst Rev*. 2006;CD000254.
5. Krishnadev N, Meleth AD, Chew EY. Nutritional supplements for age-related macular degeneration. *Curr Opin Ophthalmol*. 2010;21:184–189.
6. Mares JA, Voland RP, Sondel SA, et al. Healthy lifestyles related to subsequent prevalence of age-related macular degeneration. *Arch Ophthalmol*. 2011;129:470–480.
7. Hageman GS, Luthert PJ, Chong VNH, Johnson LV, Anderson DH, Mullins RF. An integrated hypothesis that considers drusen as biomarkers of immune-mediated processes at the RPE-Bruch's membrane interface in aging and age-related macular degeneration. *Prog Retin Eye Res*. 2001;20:705–732.
8. Bird AC, Bressler NM, Bressler SB, et al. An international classification and grading system for age-related maculopathy and age-related macular degeneration. The International ARM Epidemiological Study Group. *Surv Ophthalmol*. 1995;39:367–374.
9. Bressler NM, Bressler SB, Seddon JM, Gragoudas ES, Jacobson LP. Drusen characteristics in patients with exudative versus non-exudative age-related macular degeneration. *Retina*. 1988;8:109–114.
10. Bressler SB, Maguire MG, Bressler NM, Fine SL. Relationship of drusen and abnormalities of the retinal pigment epithelium to the prognosis of neovascular macular degeneration. The Macular Photocoagulation Study Group. *Arch Ophthalmol*. 1990;108:1442–1447.
11. Holz FG, Wolfensberger TJ, Piguet B, et al. Bilateral macular drusen in age-related macular degeneration. Prognosis and risk factors. *Ophthalmology*. 1994;101:1522–1528.
12. Klein R, Davis MD, Magli YL, Segal P, Klein BE, Hubbard L. The Wisconsin age-related maculopathy grading system. *Ophthalmology*. 1991;98:1128–1134.
13. Klein R, Klein BE, Jensen SC, Meuer SM. The five-year incidence and progression of age-related maculopathy: the Beaver Dam Eye Study. *Ophthalmology*. 1997;104:7–21.
14. Smiddy WE, Fine SL. Prognosis of patients with bilateral macular drusen. *Ophthalmology*. 1984;91:271–277.
15. Bressler NM, Silva JC, Bressler SB, Fine SL, Green WR. Clinicopathologic correlation of drusen and retinal pigment epithelial abnormalities in age-related macular degeneration. *Retina*. 1994;14:130–142.
16. Age-Related Eye Disease Study Research Group. The Age-Related Eye Disease Study system for classifying age-related macular degeneration from stereoscopic color fundus photographs: the Age-Related Eye Disease Study report number 6. *Am J Ophthalmol*. 2001;132:668–681.
17. Klaver CC, Assink JJ, van Leeuwen R, et al. Incidence and progression rates of age-related maculopathy: the Rotterdam Study. *Invest Ophthalmol Vis Sci*. 2001;42:2237–2241.
18. Seddon JM, Sharma S, Adelman RA. Evaluation of the clinical age-related maculopathy staging system. *Ophthalmology*. 2006;113:260–266.
19. Smith W, Assink J, Klein R, et al. Risk factors for age-related macular degeneration: pooled findings from three continents. *Ophthalmology*. 2001;108:697–704.
20. Yehoshua Z, Wang F, Rosenfeld PJ, Penha FM, Feuer WJ, Gregori G. Natural history of drusen morphology in age-related macular degeneration using spectral domain optical coherence tomography. *Ophthalmology*. 2011;118:2434–2441.
21. Chiu SJ, Izatt JA, O'Connell RV, Winter KP, Toth CA, Farsiu S. Validated automatic segmentation of AMD pathology including drusen and geographic atrophy in SD-OCT images. *Invest Ophthalmol Vis Sci*. 2012;53:53–61.

Investigative Ophthalmology & Visual Science

22. Scholl HP, Peto T, Dandekar S, et al. Inter- and intra-observer variability in grading lesions of age-related maculopathy and macular degeneration. *Graefes Arch Clin Exp Ophthalmol*. 2003;241:39–47.
23. Quellec G, Russell SR, Seddon JM, et al. Automated discovery and quantification of image-based complex phenotypes: a twin study of drusen phenotypes in age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2011;52:9195–9206.
24. Niemeijer M, van Ginneken B, Russell SR, Suttorp-Schulten MS, Abramoff MD. Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis. *Invest Ophthalmol Vis Sci*. 2007;48:2260–2267.
25. Liang Z, Wong DW, Liu J, Chan KL, Wong TY. Towards automatic detection of age-related macular degeneration in retinal fundus images. *Conf Proc IEEE Eng Med Biol Soc. 2010*; 2010:4100–4103. doi:10.1109/IEMBS.2010.5627289.
26. Jain N, Farsiu S, Khanifar AA, et al. Quantitative comparison of drusen segmented on SD-OCT versus drusen delineated on color fundus photographs. *Invest Ophthalmol Vis Sci*. 2010; 51:4875–4883.
27. Mora AD, Vieira PM, Manivannan A, Fonseca JM. Automated drusen detection in retinal images using analytical modelling algorithms. *Biomed Eng Online*. 2011;10:59.
28. Ong BB, Lee N, Lee WP, et al. Optimisation of an automated drusen-quantifying software for the analysis of drusen distribution in patients with age-related macular degeneration. *Eye (Lond)*. 2013;27:554–560.
29. Rapantzikos K, Zervakis M, Balas K. Detection and segmentation of drusen deposits on human retina: potential in the diagnosis of age-related macular degeneration. *Med Image Anal*. 2003;7:95–108.
30. Sanchez CI, Niemeijer M, Isgum I, et al. Contextual computer-aided detection: improving bright lesion detection in retinal images and coronary calcification identification in CT scans. *Med Image Anal*. 2012;16:50–62.
31. Smith RT, Chan JK, Nagasaki T, Sparrow JR, Barbazetto I. A method of drusen measurement based on reconstruction of fundus background reflectance. *Br J Ophthalmol*. 2005;89: 87–91.
32. Smith RT, Sohrab MA, Pumariega NM, et al. Drusen analysis in a human-machine synergistic framework. *Arch Ophthalmol*. 2011;129:40–47.
33. Fauser S, Smailhodzic D, Caramoy A, et al. Evaluation of serum lipid concentrations and genetic variants at high-density lipoprotein metabolism loci and TIMP3 in age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2011;52:5525–5528.
34. van de Ven JP, Smailhodzic D, Boon CJ, et al. Association analysis of genetic and environmental risk factors in the cuticular drusen subtype of age-related macular degeneration. *Mol Vis*. 2012;18:2271–2278.
35. Jain AK, Duin RPW, Jianchang M. Statistical pattern recognition: a review. *IEEE T Pattern Anal*. 2000;22:4–37.
36. Ho TK. The random subspace method for constructing decision forests. *IEEE T Pattern Anal*. 1998;20:832–844.
37. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. *Stat Surv*. 2010;4:40–79.
38. He X, Frey E. ROC, LROC, FROC, AFROC: an alphabet soup. *J Am Coll Radiol*. 2009;6:652–655.
39. Koch GG. Intraclass correlation coefficient. In: Kotz S, Johnson NL, eds. *Encyclopedia of Statistical Sciences*. New York: John Wiley & Sons; 1982;213–217.
40. Grading diabetic retinopathy from stereoscopic color fundus photographs–an extension of the modified Airlie House classification. ETDRS report number 10. Early Treatment Diabetic Retinopathy Study Research Group. *Ophthalmology*. 1991;98:786–806.
41. Goldbaum MH, Katz NP, Nelson MR, Haff LR. The discrimination of similarly colored objects in computer images of the ocular fundus. *Invest Ophthalmol Vis Sci*. 1990;31:617–623.
42. Kirkpatrick JN, Spencer T, Manivannan A, Sharp PF, Forrester JV. Quantitative image analysis of macular drusen from fundus photographs and scanning laser ophthalmoscope images. *Eye (Lond)*. 1995;9(pt 1):48–55.
43. Morgan WH, Cooper RL, Constable IJ, Eikelboom RH. Automated extraction and quantification of macular drusen from fundal photographs. *Aust N Z J Ophthalmol*. 1994;22:7–12.
44. Peli E, Lahav M. Drusen measurement from fundus photographs using computer image analysis. *Ophthalmology*. 1986; 93:1575–1580.
45. Sebag M, Peli E, Lahav M. Image analysis of changes in drusen area. *Acta Ophthalmol*. 1991;69:603–610.
46. Shin DS, Javornik NB, Berger JW. Computer-assisted, interactive fundus image processing for macular drusen quantitation. *Ophthalmology*. 1999;106:1119–1125.
47. Brandon L, Hoover A. Drusen detection in a retinal image using multi-level analysis. In: Ellis RE, Peters TM, eds. *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2003*. Berlin: Springer; 2003:618–625.
48. Metz CE. Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems. *J Am Coll Radiol*. 2006;3:413–422.
49. Agurto C, Barriga ES, Murray V, et al. Automatic detection of diabetic retinopathy and age-related macular degeneration in digital fundus images. *Invest Ophthalmol Vis Sci*. 2011;52: 5862–5871.
50. Zheng Y, Hijazi MHA, Coenen F. Automated "disease/no disease" grading of age-related macular degeneration by an image mining approach. *Invest Ophthalmol Vis Sci*. 2012;53: 8310–8318.
51. van de Ven JP, Boon CJ, Smailhodzic D, et al. Short-term changes of Basal laminar drusen on spectral-domain optical coherence tomography. *Am J Ophthalmol*. 2012;154:560–567.
52. Karnon J, Czoski-Murray C, Smith K, et al. A preliminary model-based assessment of the cost-utility of a screening programme for early age-related macular degeneration. *Health Technol Assess*. 2008;12:iii–iv. ix-124.
53. Karnon J, Czoski-Murray C, Smith KJ, Brand C. A hybrid cohort individual sampling natural history model of age-related macular degeneration: assessing the cost-effectiveness of screening using probabilistic calibration. *Med Decis Making*. 2009;29:304–316.
54. Bellman R. On the theory of dynamic programming. *Proc Natl Acad Sci U S A*. 1952;38:716–719.

Investigative Ophthalmology & Visual Science