



Published in final edited form as:

*Ophthalmol Retina*. 2021 November ; 5(11): 1061–1073. doi:10.1016/j.oret.2021.07.006.

## Cluster analysis and genotype-phenotype assessment of geographic atrophy in age-related macular degeneration: AREDS2 Report 25

Tiarnan D. L. Keenan<sup>1</sup>, Neal L. Oden<sup>2</sup>, Elvira Agrón<sup>1</sup>, Traci E. Clemons<sup>2</sup>, Alice Henning<sup>2</sup>, Lars G. Fritsche<sup>3</sup>, Wai T. Wong<sup>4</sup>, Emily Y. Chew<sup>1</sup>, AREDS2 Research Group<sup>5</sup>

<sup>1</sup>Division of Epidemiology and Clinical Applications, National Eye Institute, National Institutes of Health, Bethesda, MD, USA

<sup>2</sup>The Emmes Company LLC, Rockville, MD, USA

<sup>3</sup>University of Michigan, Ann Arbor, MI, USA

<sup>4</sup>Janssen Research and Development LLC, Raritan, NJ, USA

<sup>5</sup>Appendix for the listing of all the members of the AREDS2 Research Group

### Abstract

**Purpose**—To explore whether phenotypes in geographic atrophy (GA) secondary to age-related macular degeneration (AMD) can be separable into two or more partially distinct subtypes and if these have different genetic associations. This is important since the discovery of distinct GA subtypes associated with different genetic factors might require customized therapeutic approaches.

**Design**—Cluster analysis of participants within a controlled clinical trial, followed by assessment of phenotype-genotype associations.

**Participants**—AREDS2 participants with incident GA during study follow-up: 598 eyes of 598 participants (median age 75.7y).

**Methods**—Phenotypic features from reading center grading of fundus photographs were subjected to cluster analysis, by both k-means and hierarchical methods, in cross-sectional analyses (using 15 phenotypic features assessed principally at GA emergence) and longitudinal analyses (using 14 phenotypic features). In pre-specified hypothesis tests, identified clusters were compared by four pathway-based genetic risk scores (complement, extracellular matrix, lipid, and *ARMS2*). The analyses were repeated in reverse, i.e., clustering by genotype and comparison by phenotype.

---

**Corresponding author:** Tiarnan D. L. Keenan, BM BCh, PhD, NIH, Building 10, CRC, Room 10D45, 10 Center Dr, MSC 1204, Bethesda, MD 20892-1204, Telephone: 301 451 6330, Fax: 301 496 7295, tiarnan.keenan@nih.gov.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflict of interest:

No conflicting relationship exists for any author.

**Main outcome measures**—Characteristics and quality of cluster solutions, assessed by Calinski-Harabasz scores, unexplained variance, and consistency; genotype-phenotype associations, assessed by t test.

**Results**—In cross-sectional phenotypic analyses, k-means identified two clusters (labeled A, B), while hierarchical clustering identified four (C-F); A-E membership differed principally by GA configuration but in relatively few other ways. In longitudinal phenotypic analyses, k-means identified two clusters (G, H), which differed principally by smoking status but in relatively few other ways. These three sets of cluster divisions were not similar to each other ( $r = 0.20$ ). Despite adequate power, pairwise cluster comparison by the four genetic risk scores demonstrated no significant differences ( $p > 0.05$  for all). In clustering by genotype, k-means identified two clusters (I/J). These differed principally at *ARMS2*, but no significant genotype-phenotype associations were observed ( $p > 0.05$  for all).

**Conclusions**—Phenotypic clustering resulted in GA subtypes defined principally by GA configuration in cross-sectional analyses, but these were not replicated in longitudinal analyses. These negative findings, together with the absence of significant phenotype-genotype associations, indicate that GA phenotypes may vary continuously across a spectrum, rather than consisting of distinct subtypes that arise from separate genetic etiologies.

## Précis

Cluster analysis of individuals with geographic atrophy secondary to age-related macular degeneration demonstrated that geographic atrophy phenotypes vary continuously across a spectrum, rather than consisting of distinct subtypes that arise from separate genetic etiologies.

---

## Introduction

Geographic atrophy (GA) in age-related macular degeneration (AMD) is an important research priority, since no treatments with regulatory approval are available to restore vision to affected areas of the retina.<sup>1</sup> Progression to GA and rate of GA enlargement are both important endpoints in many clinical trials. Recent and ongoing clinical trials suggest that new treatments to decrease progression to GA and slow GA enlargement may become available in the near future.<sup>2–5</sup> Interest has also grown in potential genotype-phenotype correlations in GA<sup>6</sup>, including associations between genotype and enlargement rate<sup>7,8</sup>, as these associations can potentially reveal causative molecular mechanisms for therapeutic targeting.

In this context, it is important to consider whether GA in AMD may consist of two or more partially distinct subtypes, as suggested in a recent study.<sup>6</sup> This phenomenon would be particularly relevant if these GA subtypes progressed in distinct ways over time, especially with regards to altered speed of enlargement or differences in treatment response. For example, slowing GA enlargement might require a particular pharmacological intervention for one genetically driven subtype of GA but a different intervention for the other subtypes.

Cluster analysis is a form of unsupervised classification used to sort data into separate categories, when no aspects of the group structure are known *a priori*. It is particularly useful when the data are relatively complex and multidimensional. As argued in a recent editorial,

it is helpful not necessarily for finding clusters per se, but for identifying disease subgroups that may be useful in subsequent prospective studies for providing more accurate prognostic information or predicting differential responses to treatment.<sup>9</sup>

The Age-Related Eye Disease Study 2 (AREDS2) was a multicenter phase III randomized clinical trial (RCT) designed to assess the effects of nutritional supplementation on AMD progression.<sup>10</sup> The purpose of the current study was to perform cluster analysis on the cohort of AREDS2 participants with GA. Specifically, the main aim was to identify clusters of GA based on a broad range of phenotypic characteristics, then compare the clusters in a pre-specified way according to genetic characteristics. The secondary aim was the reverse, i.e., to identify clusters based on genetic characteristics and compare these according to the phenotypic characteristics. In this way, the objective was to determine whether GA may indeed consist of two or more subtypes, separated by phenotypic and/or genetic features, that remain partially distinct over time.

## Methods

### Study population and procedures

The AREDS2 study design has been described previously.<sup>10</sup> In brief, 4203 participants (aged 50 to 85 years) were recruited between 2006 and 2008 at 82 retinal specialty clinics in the United States. The inclusion criteria at enrollment were the presence of either bilateral large drusen or late AMD in 1 eye and large drusen in the fellow eye. Institutional review board approval was obtained at each site and written informed consent was obtained from all participants. The research was conducted under the tenets of the Declaration of Helsinki and complied with the Health Insurance Portability and Accountability Act.

The AREDS2 lasted five years. At baseline and annual study visits, comprehensive eye examinations were performed by certified study personnel using a standardized protocol, including measurement of the best-corrected visual acuity (BCVA) using the electronic Early Treatment Diabetic Retinopathy Study (ETDRS) visual acuity chart. Digital stereoscopic color fundus photographs were captured by certified study personnel using a standardized protocol. The fundus photographs were graded centrally by certified graders at the University of Wisconsin Fundus Photograph Reading Center. The details have been described previously.<sup>11</sup> GA was defined as a lesion equal to or larger than drusen circle I-2 (diameter 433  $\mu\text{m}$ , area 0.146  $\text{mm}^2$ , i.e., 1/4 disc diameter and 1/16 disc area) at its widest diameter. The configuration of GA was documented, using the definitions published by Sunness et al<sup>12</sup>, as either (1) small (single patch less than 1 disc area), (2) multifocal, (3) horseshoe or ring, (4) solid or unifocal, or (5) indeterminate. Planimetry tools were used to demarcate the area of GA within the AREDS grid in square millimeters. The color fundus photographs were also analyzed for drusen, calcified drusen, and the presence of neovascular AMD, as described previously.<sup>11</sup>

The reticular pseudodrusen (RPD) score was defined from deep learning-based automated grading of the color fundus photographs. The grading algorithm and its performance metrics have been described previously.<sup>13</sup> In brief, a deep learning algorithm was trained by exposing it to AREDS2 color fundus photographs, with accompanying grades for

RPD presence/absence derived from reading center grading of corresponding fundus autofluorescence images. The RPD score was a continuous variable (0.0–1.0) generated by the algorithm to describe its confidence in RPD presence. This metric was used because it is available for all AREDS2 eyes at all visits; by contrast, reading center RPD grading is available for only a small subset of eyes and visits.

The study population for these analyses was defined as follows, using methods similar to those described in previous analyses of GA in the AREDS2.<sup>7</sup> The eligible population comprised all eyes with incident GA that had at least one subsequent study visit; however, eyes with prior or simultaneous neovascular AMD were excluded. Only one eye per participant was selected: for participants with two eligible eyes, the eye that developed GA earlier was selected; if both eyes developed GA at the same visit, one eye was selected at random. Eyes with incident GA were selected in order to capture phenotypic variation at a relatively fixed point (early on in the natural history of GA), i.e., in order to decrease the possibility that clusters might be formed that could relate partially or predominantly to disease stage rather than genuine phenotypic variation. The same study population was used to create two datasets: one for cross-sectional analyses and one for longitudinal analyses.

### Phenotypic characteristics

Multiple demographic and phenotypic characteristics were used as the basis for cluster analysis of the study population: 15 characteristics for the cross-sectional analyses and 14 for the longitudinal analyses (Table 1). For the cross-sectional analyses, the characteristics comprised demographic ones, GA characteristics, AMD characteristics, and BCVA. The five demographic characteristics were defined at the AREDS2 baseline visit and the other characteristics (except GA enlargement rate) were defined at the first visit with GA, in order to capture phenotypic variation at a relatively fixed point. Although not strictly a cross-sectional characteristic, GA enlargement rate was included owing to its importance; it was calculated by regression of the square root of GA area over time, according to methods described previously.<sup>7</sup>

The 14 characteristics used for the longitudinal analyses were the same, except for (i) omission of square root of GA area, and (ii) use of BCVA rate (i.e., the slope from regression of BCVA over time, from time of first GA). For the following categorical variables, the characteristic was defined as the ‘maximum’ value during follow-up: GA central involvement, fellow eye involvement, drusen area, maximum drusen size, calcified drusen presence, and RPD score. For GA configuration, the most common configuration over time was selected for each eye; if an eye had multiple configurations with equal frequency, the last configuration was used.

The data were pre-processed as follows. Categorical variables with 3 levels (education and smoking) or 5 levels (GA configuration) were split up into constituent 2-level components (e.g., current smoking status y/n, former smoking status y/n, and never smoking status y/n). Drusen area (levels 1–7) and maximum drusen size (levels 2–5) were treated as continuous variables. All variables were standardized to have a mean of 0 and a variance of 1.

### Genetic characteristics: calculation of four pathway-based genetic risk scores

As part of the AREDS2, 1826 participants consented to genotype analysis. SNPs were analyzed using a custom Illumina HumanCoreExome array, as described previously.<sup>14</sup> The AMD Genetic Risk Score, a weighted risk score for late AMD based on 52 SNPs associated with altered risk of late AMD in a large genome-wide association study (GWAS), was calculated for each participant, using methods described previously.<sup>14</sup> In brief, it was computed as the sum of the 52 risk allele dosages for each SNP multiplied by their beta-coefficients (as reported in the GWAS<sup>14</sup>) in the log-odds scale.

In addition, four pathway-based genetic risk scores were calculated for each participant with genetic data available, using methods described previously.<sup>6</sup> These comprised: (i) complement pathway, (ii) lipid metabolism, (iii) extracellular matrix remodeling, and (iv) *ARMS2* genetic risk scores. The SNPs used to define these four scores were a subset of the 52 SNPs described above, and similar methods were used to calculate them (i.e., summed risk allele dosages with weighting by beta-coefficients). The corresponding SNP numbers are shown in Table 2, while the details to construct the genetic risk scores were previously published<sup>6</sup> and are given in Supplementary Table 1.

### Cluster analysis based on phenotypic characteristics

Principal components analysis was performed, separately for the cross-sectional dataset and longitudinal dataset, using the phenotypic variables described above. Scree plots were made, showing the proportion of variance explained in the datasets according to the number of principal components. K-means cluster analysis was performed, separately for the cross-sectional and longitudinal datasets, using the same phenotypic variables. K-means cluster analysis has been explained previously.<sup>6,9,15</sup> In this study, k-means cluster analysis was performed for multiple cluster numbers *k*, from 2 to 20. Following clustering, the optimal number of clusters and the clustering results were explored by calculating the Calinski-Harabasz scores (i.e., the ratio of between-cluster variance to within-cluster variance, where high scores indicate close-knit and separate clusters, though Calinski-Harabasz scores cannot be defined for a cluster number of 1)<sup>15,16</sup> and by plotting the proportion of unexplained sums of squares against the number of clusters.

In order to explore the potential consistency of cluster analysis results using different methods, agglomerative hierarchical cluster analysis was performed using the same dataset, based on Euclidean distance and separately for single, complete, and average linkage (i.e., three different approaches to judging cluster similarity). Agglomerative hierarchical clustering has been described previously.<sup>9,15</sup> The results are depicted on dendrograms: the participants are shown on the x-axis and each horizontal line signifies the fusion of two clusters, with height representing their dissimilarity. In this study, the optimal number of clusters and the clustering results were explored by plotting dendrograms and by calculating the Calinski-Harabasz scores and the cophenetic correlation coefficients.

### Hypothesis-based comparison of phenotypic clusters by genotype

In pre-specified and hypothesis-based tests, the phenotypic clusters were compared according to genotype, using the four pathway-based genetic risk scores. In this way,

the rationale was to perform cluster analysis (which is by definition exploratory) based purely on phenotypic characteristics, then compare the phenotypic clusters by genotype in a pre-specified and hypothesis-based way, in order to control for multiple testing. Specifically, pairwise comparisons were made between the cluster participants with genetic data available by two-tailed t test, with adjustment for multiple testing (by bootstrap, using PROC MULTTEST). MULTTEST was performed separately for each cluster pair under comparison. In MULTTEST, t-tests were performed for each of the four genetic risk scores and the smallest of the observed p-values recorded. Bootstrapping was then performed, where many pseudo-datasets are generated (by sampling with replacement, where the unit of bootstrapping is the individual and his/her genetic data), in order to approximate the distribution of smallest p-value recorded earlier. This distribution is then used to adjust the individual raw p-values. In addition to bootstrap adjustment for each pair of clusters, Bonferroni adjustment was also performed to account for the fact that several pairs of clusters were being compared.

The exploration and hypothesis testing stages of the study were kept distinct by separating the phenotypic and genetic characteristics into two datasets and coding the participant identities differently in the two datasets until the final hypothesis tests. Post hoc sample size/power calculations were performed in order to assess the Cohen effect size detectable with the sample sizes defined in the clusters, assuming power of 0.80.

Following clustering, in non-hypothesis-based analyses, the clusters were characterized and interpreted on the basis of their demographic and phenotypic characteristics and, separately, their genetic characteristics. This was performed by (i) classification and regression tree (CART) analysis, (ii) logistic regression by least absolute shrinkage and selection operator (LASSO), and (iii) analysis by Cohen's effect sizes. The results were assessed by confusion matrices and standard performance metrics.

### **Cluster analysis based on genetic characteristics**

In the second part of the study, cluster analysis was performed using the genetic characteristics as input (specifically the four pathway-based genetic risk scores). The study population was the same as that in the previous analyses but comprised only those participants with genetic data available. Similar methods were used: both k-means and hierarchical agglomerative cluster analysis were performed. This was followed by hypothesis-based comparison of the genetic clusters by phenotype, i.e., according to the same phenotypic characteristics described above (considered both cross-sectionally and longitudinally), through pairwise comparisons by two-tailed t test. Following this, the clusters were characterized and interpreted by CART, logistic regression, and Cohen's effect sizes. Again, post hoc sample size/power calculations were performed.

Finally, separately from cluster analysis, the Mantel test was performed to investigate potential correlations between the phenotypic characteristics and the genetic characteristics (specifically the 52-SNP genetic risk score and the four pathway-based genetic risk scores).<sup>17</sup>

Statistical analysis was performed using R (version 4.0.2). In the hypothesis-based testing, following adjustment for multiple testing,  $p$ -values  $< 0.05$  (in two-tailed tests) were defined as significant.

## Results

The cohort consisted of 598 eyes of 598 participants. Their demographic and phenotypic characteristics are shown in Table 3. Median follow-up in the longitudinal dataset, from time of first appearance of GA, was 2.6 years (IQR 1.5, 3.4). Of the 598 participants, 313 (52.3%) had genetic data available. Their genetic characteristics, according to the pathway-based genetic risk scores, are also shown in Table 3.

### Cluster analysis of the cross-sectional dataset, based on phenotypic characteristics

Principal components analysis was not successful in compressing the variance of the data into a small number of dimensions (Supplementary Figure 1). K-means cluster analysis was performed. The proportion of unexplained variance decreased slowly and gradually with increasing cluster number (Supplementary Figure 2). However, the optimal number of clusters, according to the Calinski-Harabasz scores, was two (Supplementary Figure 3). These two clusters were named A (367 participants) and B (231 participants; Table 4).

Agglomerative hierarchical cluster analysis was performed on the same dataset. The dendrograms are shown in Supplementary Figure 4. The Calinski-Harabasz scores did not reveal a cluster number that was clearly optimal (Supplementary Figure 5). Under complete linkage, a consistent cluster of 112 participants was observed with a cluster number of four or five. With four clusters, the numbers of participants were 469, 112, 12, and 5. These clusters were named C, D, E, and F, respectively (Table 4).

Under single and average linkage, even with large cluster numbers, no major subdivisions were observed. Membership of cluster A or B was not highly correlated with membership of cluster C or D; the Pearson correlation coefficients were 0.20 (A versus C) and  $-0.16$  (A versus D).

### Cluster analysis of the longitudinal dataset, based on phenotypic characteristics

Again, principal components analysis was not successful in compressing the variance of the data into a small number of dimensions (Supplementary Figure 6). On k-means cluster analysis, the proportion of unexplained variance decreased slowly and gradually with increasing cluster number (Supplementary Figure 7). However, the optimal number of clusters was two (Supplementary Figure 8). These two clusters were named G (310 participants) and H (288 participants; Table 5). Membership of cluster G or H was not highly correlated with membership of clusters A-D; the Pearson correlation coefficients were 0.01,  $-0.01$ , 0.02, and  $-0.01$  (G versus A-D, respectively).

Agglomerative hierarchical cluster analysis was performed on the same dataset. The dendrograms are shown in Supplementary Figure 9. The Calinski-Harabasz scores did not reveal a cluster number that was clearly optimal (Supplementary Figure 10). For any linkage

type, even with large cluster numbers, no major subdivisions were observed. Hence, no clusters from agglomerative hierarchical clustering were considered in subsequent analyses.

### Hypothesis-based comparison of phenotypic clusters by genotype

Pairwise comparisons between clusters participants with genetic data available were made for each of the four pathway-based genetic risk scores. The results are shown in Table 6. No significant differences were observed between the cluster participants of A versus B, C versus D, or G versus H, for any of the four genetic risk scores. Although the unadjusted p-value for the comparison of clusters A and B for the *ARMS2* risk score was 0.008, the adjusted p-value was not significant at 0.096. The *ARMS2* risk allele frequencies for A and B, respectively, were: 41.5% versus 25.7% with no risk alleles, 40.0% versus 48.7% with one, and 18.5% versus 25.7% with two.

Given the absence of significant findings, post hoc power analyses were performed (specifically for the comparison of C and D, containing the smallest cluster). Under the assumptions of power 0.80 and a two-tailed significance level of  $p=0.004$  (accounting for multiple testing), the smallest Cohen effect size detectable for this cluster pair was 0.56 (i.e., medium effect size). For completeness, the 52-SNP genetic risk score was also compared between the same clusters. In these three comparisons, no significant differences were observed at the nominal level of  $p=0.05$ .

### Characterization of phenotypic clusters

The characteristics of the cluster participants are shown in Tables 4 and 5. Notable differences between clusters A and B were observed for GA configuration and, to an extent, square root of GA area and GA enlargement rate. Specifically, 100.0% of cluster A members had the configuration small, compared to 0.4% for cluster B; 55.8% of cluster B members had multifocal GA, while 32.5% had solid/unifocal GA. Similarly, notable differences were observed between clusters C and D according to GA configuration. The large majority of cluster C members had the configurations small (66.7%) or multifocal (27.1%), while the large majority of cluster D members had solid/unifocal (52.7%) or small (45.5%). Additional potential differences were observed according to other characteristics. Notable differences between clusters G and H were observed for smoking status and, to an extent, sex and perhaps GA enlargement rate. Specifically, 100.0% of cluster G members had the smoking status former, compared to 0.0% for clusters H; 13.2% of cluster H members had the smoking status current, while 86.8% had never.

The clusters were further characterized by CART, logistic regression with LASSO, and Cohen's effect sizes, based on their phenotypic characteristics. The CART classification trees are shown in Supplementary Figure 11. According to CART, the clear deciding factor for cluster A versus B was GA configuration being small or not. Similarly, the deciding factor for cluster E was GA configuration being horseshoe/ring. The classification trees for membership of clusters C and D were more complex, but both involved GA configuration at the proximal nodes. The clear deciding factor for cluster G versus H was smoking status being former or not. The results of logistic regression with LASSO and Cohen's effect sizes



are shown in Supplementary Figures 12 and 13, respectively, and showed similar patterns to those from CART.

The clusters were also characterized by the same three methods based on their genetic characteristics. The results are shown in Supplementary Figures 14 and 15. Overall, from all three methods, the great complexity of the CART trees, the poor performance metrics, and the very low effect sizes were consistent with the findings from the pairwise comparisons described above, i.e., that the genetic data by genetic risk scores were not strongly related to the phenotypic clusters.

### **Cluster analysis of the dataset, based on genetic characteristics**

Principal components analysis was not successful in compressing the variance of the data into a small number of dimensions (Supplementary Figure 16). K-means cluster analysis was performed. The proportion of unexplained variance decreased slowly and gradually with increasing cluster number (Supplementary Figure 17). However, the optimal number of clusters was two (Supplementary Figure 18). These two clusters were named I (148 participants) and J (165 participants; Table 7).

Agglomerative hierarchical cluster analysis was performed on the same dataset. The dendrograms are shown in Supplementary Figure 19. The cophenetic correlation coefficient was highest for average linkage. The Calinski-Harabasz scores did not reveal a cluster number that was clearly optimal (Supplementary Figure 20) though, under average linkage, the score was higher for 9 clusters. With 9 clusters, the numbers of participants were 6, 106, 135, 19, 11, 16, 11, 7, and 2. The clusters with 106 and 135 participants were named K and L, respectively. The other seven clusters were named M-S; they are considered together in Table 7 but are shown individually in Supplementary Table 2.

Membership of cluster I was highly correlated with membership of cluster K (Pearson correlation coefficient 0.66). Similarly, membership of cluster J was highly correlated with membership of cluster L (coefficient 0.71).

### **Hypothesis-based comparison of genetic clusters by phenotype**

Pairwise comparisons between clusters participants were made for each of the phenotypic characteristics. The results are shown in Supplementary Table 3. No significant differences were observed between the cluster participants of I versus J, or K versus L, for any of the phenotypic characteristics.

Given the absence of significant findings, post hoc power analyses were performed (specifically for the comparison of K and L, containing the smallest cluster). Under the assumptions of power 0.80 and a two-tailed significance level of  $p=0.0007$  (accounting for multiple testing), the smallest Cohen effect size detectable for this cluster pair was 0.56 (i.e., medium effect size).

### **Characterization of genetic clusters**

The genetic clusters were characterized in non-hypothesis-based analyses (Table 7). In terms of genetic characteristics, the notable difference between clusters I and J was the *ARMS2*

risk score. The median *ARMS2* risk score for cluster I was 0.0 (IQR 0.0, 1.1), compared to 1.1 (IQR 1.1, 2.1) for cluster J. Hence, the large majority of cluster I participants (73.6%) had no *ARMS2* risk alleles and no participants had 2 risk alleles, while the large majority of cluster J participants had 1 (58.2%) or 2 (40.0%) risk alleles. Other potential differences were observed according to the phenotypic characteristics GA enlargement rate and perhaps GA configuration. Cluster I had a relatively low median GA enlargement rate and a relatively high proportion of participants with the GA configuration small. Similarly, the notable difference between clusters K and L was the *ARMS2* risk score. The median *ARMS2* risk score for cluster K was 0.0 (IQR 0.0, 0.0), compared to 1.1 (IQR 1.1, 2.1) for cluster L. Similar potential differences were observed according to GA enlargement rate and perhaps GA configuration as those described above for clusters I and J.

The clusters were further characterized by CART, logistic regression with LASSO, and Cohen's effect sizes, based on their genetic characteristics. The results are shown in Supplementary Figure 21–23. All three methods produced similar results: high *ARMS2*, low complement, high lipid, and low extracellular matrix genetic risk scores predicted J and L membership, while the opposite values predicted I and K membership. The CART classification trees were relatively simple and the performance metrics high, which suggested that, as expected, the genetic data could be used economically to predict genetic cluster membership. The methods suggested that, in order of importance for predicting cluster membership, the *ARMS2* risk score was most important; the second most important was the complement risk score, acting in an opposite direction.

The clusters were also characterized by the same three methods based on their phenotypic characteristics. The results are shown in Supplementary Figures 24 and 25. Overall, from all three methods, the great complexity of the CART trees, the poor performance metrics, and the very low effect sizes were consistent with the findings from the pairwise comparisons described above, i.e., that the phenotypic characteristics were not strongly related to the genetic clusters.

### Mantel test

Separately from cluster analysis, the Mantel test was performed to analyze potential correlations between the phenotypic Euclidean distances and the genetic Euclidean distances of the cohort (for the subset of 313 participants with genetic data available). The results are shown in Table 8. No significant relationship was observed between the phenotypic and genetic risk score distances.

### Sensitivity analyses

Sensitivity analyses were performed on a modified dataset that excluded the five non-white participants (i.e., n=593), as a more ethnically homogeneous study population. The results were very similar to the primary analyses. Similar cluster divisions and assignments were observed (again, principally according to GA configuration in the cross-sectional analyses, smoking status in the longitudinal analyses, and *ARMS2* genotype in the genetic analyses). As in the primary analyses, no significant differences were observed according to genetic

risk scores in the phenotypic cluster pairs, or according to phenotype in the genetic cluster pairs, and no significant genotype-phenotype relationship was found on Mantel testing.

## Discussion

### Main findings and implications

In this large study benefitting from prospective recruitment and standardized follow-up of patients in a clinical trial setting, cluster analysis was performed to classify patients with GA into subgroups according to their phenotypic or genetic characteristics. Multiple cluster analyses were performed using different methods in order to assess for consistency of the solutions. Clustering on phenotypic characteristics, assessed in a cross-sectional manner at time of GA emergence, identified two clusters by k-means methods and, separately, two main clusters by hierarchical methods. In both cases, GA configuration was the predominant factor. However, the cluster divisions were not consistent between k-means and hierarchical methods.

Clustering creates its divisions preferentially in places where participants differ according to more characteristics and/or differ more widely in these characteristics. It is therefore interesting to observe whether these GA clusters (that differed most strongly in GA configuration) differed also in other ways. For example, the two k-means clusters showed few differences in the other phenotypic characteristics, except for GA area and enlargement rate, in which differences would be expected.<sup>18</sup> However, a high level of heterogeneity was still observed between the clusters: the cluster divisions fell very short of any potential phenotypic 'signature', which would require wide and consistent differences across a large range of characteristics. For example, we did not observe one cluster clearly characterized by the combination of many soft drusen, calcified drusen, and central GA, and another cluster with many RPD, few soft drusen, and non-central GA, as suggested in a recent study.<sup>6</sup>

The clustering was repeated using the phenotypic characteristics considered longitudinally. This is important to assess the consistency of cluster solutions, e.g., to explore the possibility that any clusters based on cross-sectional phenotypic characteristics might relate partly to disease stage. Consistency was not observed. Indeed, smoking status (former versus not) was the predominant factor in cluster membership. Strangely, the current smokers were clustered together with those who had never smoked; however, this likely relates to the binary way in which the data were coded and the very small number of active smokers. Aside from smoking, few potential differences were observed in the other phenotypic characteristics, except for sex and GA enlargement rate, in which differences might be expected according to smoking status<sup>18</sup>; the cluster containing all of the former smokers contained more men and had slightly faster GA enlargement rate, while the cluster containing mostly never smokers contained more women and had slightly slower GA enlargement.

In phenotype-genotype analyses, despite adequate power to detect at least a medium effect size, pairwise cluster comparison demonstrated no significant differences. Hence, none of the GA subtypes suggested by the phenotypic cluster analysis was strongly related to a higher or lower genetic load at any one of the four biological pathways implicated in late

AMD risk. This suggests that, in any individual with GA, a physician is unlikely to infer the main genetic driver based on these phenotypic characteristics alone, i.e., that no clear phenotypic signature exists for GA driven by one particular biological pathway. In turn, this suggests that, should treatments for GA emerge that are heavily genotype-dependent, phenotype alone is unlikely to highlight which individuals may benefit most from treatment, and genetic testing would be required.

For completeness, the clustering and subsequent analyses were performed in both directions. Clustering on the four genetic risk scores identified two clusters, which differed principally by the *ARMS2* score and secondarily by the complement score, in the opposite direction. This could be consistent with the idea that two partially distinct subtypes of GA exist at the genetic level, one related to risk at *ARMS2* and the other to risk at *CFH* and other genes of the complement pathway. However, the distributions of the complement genetic risk scores in the two clusters overlapped substantially (Table 7). Repeating the clustering by genotype with the addition of multi-omics data would help explore the possibility of two partially distinct GA subtypes at the molecular level (driven principally by *ARMS2* or by complement).

In genotype-phenotype analyses, despite adequate power to detect at least a medium effect size, pairwise cluster comparison demonstrated no significant differences. This suggests that the two genetic subtypes of GA, i.e., with high or low genetic risk at *ARMS2*, do not differ strongly according to the phenotypic characteristics examined here. Some possible genotype-phenotype relationships were suggested weakly at the nominal level, related to GA configuration and enlargement rate. In the case of faster GA enlargement in individuals with *ARMS2* risk alleles, this has been observed more conclusively in previous studies of the same dataset.<sup>18</sup>

### Comparison with literature

We are aware of only one previous study that has examined GA by cluster analysis.<sup>6</sup> The authors pooled individuals with GA from three European studies and performed cluster analysis based on phenotype and genotype simultaneously. Specifically, six phenotypic characteristics were used (graded in a binary way), with the same four genetic risk scores used in the current study. Clustering identified three clusters. Post hoc analyses showed that the genetic characteristics were highly predictive of cluster membership, while the phenotypic characteristics were poorly predictive. The three clusters were characterized genetically by high complement, lipid, and *ARMS2* and low extracellular matrix risk scores for one, high extracellular matrix and *ARMS2* and low complement risk scores for another, and relatively low genetic risk scores overall for the third. RPD presence was observed more commonly in the second cluster and less commonly in the third cluster, while central GA was seen less commonly in the second and more commonly in the third.

However, this previous study was limited by aspects of its study population, phenotyping, and methodology.<sup>9</sup> The study was cross-sectional in nature, with the GA cases not necessarily captured at equivalent times in their natural history; this means that it was not possible to assess the stability of the cluster solution (i.e., whether the patients remained true to their clusters over time, in the context of dynamic phenotypic features like central

involvement and focality). Aside from the smaller size of the study population, one of the three centers contributed two thirds of the cases and the phenotypic grading differed by center, which can introduce bias.

Direct comparison of the results from the two studies is limited by the partially different approaches. However, the findings in the previous study of poor ability of phenotype to predict cluster membership are in keeping with the findings in the current study of no significant genotype-phenotype relationship. No meaningful comparison is possible for the results by phenotype alone of the current study, for several reasons. First, clustering by phenotype alone was not conducted in the previous study and phenotype was poorly predictive of cluster membership. Second, the spectrum of phenotypic characteristics was much wider in the current study; for example, GA configuration and smoking status were two important characteristics in the current study but were not included in the previous study. Third, the previous study considered cross-sectional data only; indeed, it is possible that the cross-sectional nature of the phenotypic assessment explained the poor contribution of phenotype to the clusters.

### Strengths and limitations

The strengths of this study include the large size and relative diversity of the GA cohort (with patients drawn from 82 retinal specialty clinics across the United States), together with its clinical trial setting. Importantly, the longitudinal nature of the dataset meant that the GA cases could be examined at a relatively similar stage in their natural history and permitted clustering by cross-sectional and longitudinal characteristics. This was important to decrease the possibility that any phenotypic clusters might relate to disease stage rather than genuine phenotypic variation.<sup>9</sup> Additional strengths include the large number and wide range of phenotypic characteristics examined, the use of pathway-based genetic risk scores, clustering by two different methods, and the assessment of cluster solutions by multiple metrics. Finally, the study benefitted from a pre-specified statistical plan, with explicit separation between the data exploration and hypothesis-testing stages.

The study may be partially limited in its generalizability to other populations, such as GA in Asian individuals, owing to its single country setting and the very high (almost exclusive) proportion of white participants. In addition, not all participants had genetic data available. The phenotypic characteristics examined were based principally on color fundus photography, rather than multimodal imaging, so some phenotypic heterogeneity in GA may not have been captured. RPD grading was performed by deep learning-based grading of the fundus photographs. However, the ground truth of the RPD algorithm's training was from reading center grading of fundus autofluorescence images (leading to high specificity of grading from fundus photographs alone, in a previous study<sup>13</sup>), and we considered it more important to have uniform grading available for all eyes rather than multimodal grading for a small subset only. Different results might be obtained with RPD grading by optical coherence tomography (OCT) or the inclusion of other OCT data, and future studies would be helpful in exploring potential GA subtypes and precursors related to RPE and outer retinal atrophy (RORA) versus isolated outer retinal atrophy (ORA).<sup>21</sup> Future studies may benefit from the addition of other characteristics, including multi-omics data.<sup>9</sup> Other

potential limitations include the absence of training/test set splits during clustering and a single definition of distance (Euclidean), though consistency of solutions was assessed in other ways, and the availability of ordinal rather than fully quantitative characteristics in some cases (e.g., total drusen area, which was unadjusted for GA area).

As a general limitation of cluster analysis, clustering is limited by the inability to compare the quality of cluster solutions for one versus multiple clusters. Since clustering algorithms are always successful in forming clusters, it is difficult to conclude definitively that GA exhibits continuous phenotypic variation across a spectrum, rather than variation clustered into two or more subtypes. For this reason, approaches such as assessing the consistency and quality of cluster solutions obtained by different clustering techniques, as used in this study, become particularly important.

## Conclusions

In conclusion, cross-sectional phenotypic cluster analyses revealed GA subtypes defined principally by GA configuration. However, these subdivisions were not replicated in longitudinal phenotypic analyses that are important for considering cluster stability over time. The inconsistencies in optimal cluster numbers and characteristics suggest that GA may show continuous phenotypic variation across a spectrum, rather than consisting of phenotypic subtypes that remain partially distinct over time, with separate genetic etiologies. Clustering by pathway-based genotype alone suggested two subtypes of GA that differed principally by *ARMS2* genotype. However, no significant genotype-phenotype associations were observed, either for these two subtypes or at the level of the whole dataset. This suggests that, for any eye with GA, physicians are unlikely to infer the main genetic driver of GA from these phenotypic characteristics alone.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Financial support:

This work was supported in part by the Intramural Research Program of the National Eye Institute, National Institutes of Health (NIH), Department of Health and Human Services, Bethesda, MD (AREDS2 contract HHS-N-260-2005-00007-C; ADB contract N01-EY-5-0007). Funds were generously contributed to these contracts by the following NIH institutes: Office of Dietary Supplements; National Center for Complementary and Alternative Medicine; National Institute on Aging; National Heart, Lung, and Blood Institute; National Institute of Neurological Disorders and Stroke. The funding organization participated in the design and conduct of the study, data collection, management, analysis, and interpretation, and preparation, review and approval of the manuscript.

## Abbreviations:

<b>AMD</b>	age-related macular degeneration
<b>AREDS2</b>	Age-Related Eye Disease Study 2
<b>BCVA</b>	best-corrected visual acuity
<b>CART</b>	classification and regression tree

<b>DHA</b>	docosahexaenoic acid
<b>ETDRS</b>	Early Treatment Diabetic Retinopathy Study
<b>EPA</b>	eicosapentaenoic acid
<b>GA</b>	geographic atrophy
<b>GWAS</b>	genome-wide association study
<b>LASSO</b>	least absolute shrinkage and selection operator
<b>RCT</b>	randomized controlled trial
<b>RPE</b>	retinal pigment epithelium
<b>RPD</b>	reticular pseudodrusen

## References

1. Ammar MJ, Hsu J, Chiang A, Ho AC, Regillo CD. Age-related macular degeneration therapy: a review. *Curr Opin Ophthalmol*. 2020;31(3):215–221. [PubMed: 32205470]
2. Liao DS, Grossi FV, El Mehdi D, et al. Complement C3 Inhibitor Pegcetacoplan for Geographic Atrophy Secondary to Age-Related Macular Degeneration: A Randomized Phase 2 Trial. *Ophthalmology*. 2020;127(2):186–195. [PubMed: 31474439]
3. Jaffe GJ, Westby K, Csaky KG, et al. C5 Inhibitor Avacincaptad Pegol for Geographic Atrophy Due to Age-Related Macular Degeneration: A Randomized Pivotal Phase 2/3 Trial. *Ophthalmology*. 2020.
4. Guymer RH, Wu Z, Hodgson LAB, et al. Subthreshold Nanosecond Laser Intervention in Age-Related Macular Degeneration: The LEAD Randomized Controlled Clinical Trial. *Ophthalmology*. 2019;126(6):829–838. [PubMed: 30244144]
5. Nebbioso M, Lambiase A, Cerini A, Limoli PG, La Cava M, Greco A. Therapeutic Approaches with Intravitreal Injections in Geographic Atrophy Secondary to Age-Related Macular Degeneration: Current Drugs and Potential Molecules. *Int J Mol Sci*. 2019;20(7).
6. Biarnes M, Colijn JM, Sousa J, et al. Genotype- and Phenotype-Based Subgroups in Geographic Atrophy Secondary to Age-Related Macular Degeneration: The EYE-RISK Consortium. *Ophthalmol Retina*. 2020.
7. Keenan TD, Agron E, Domalpally A, et al. Progression of Geographic Atrophy in Age-related Macular Degeneration: AREDS2 Report Number 16. *Ophthalmology*. 2018;125(12):1913–1928. [PubMed: 30060980]
8. Grassmann F, Harsch S, Brandl C, et al. Assessment of Novel Genome-Wide Significant Gene Loci and Lesion Growth in Geographic Atrophy Secondary to Age-Related Macular Degeneration. *JAMA Ophthalmol*. 2019.
9. Keenan TD. The Hitchhiker's Guide to Cluster Analysis: Multi Pertransibunt et Augebitur Scientia. *Ophthalmol Retina*. 2020;4(12):1125–1128. [PubMed: 33279007]
10. AREDS2 Research Group, Chew EY, Clemons T, et al. The Age-Related Eye Disease Study 2 (AREDS2): study design and baseline characteristics (AREDS2 report number 1). *Ophthalmology*. 2012;119(11):2282–2289. [PubMed: 22840421]
11. Danis RP, Domalpally A, Chew EY, et al. Methods and reproducibility of grading optimized digital color fundus photographs in the Age-Related Eye Disease Study 2 (AREDS2 Report Number 2). *Invest Ophthalmol Vis Sci*. 2013;54(7):4548–4554. [PubMed: 23620429]
12. Sunness JS, Bressler NM, Tian Y, Alexander J, Applegate CA. Measuring geographic atrophy in advanced age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 1999;40(8):1761–1769. [PubMed: 10393046]

13. Keenan TDL, Chen Q, Peng Y, et al. Deep Learning Automated Detection of Reticular Pseudodrusen from Fundus Autofluorescence Images or Color Fundus Photographs in AREDS2. *Ophthalmology*. 2020.
14. Fritsche LG, Igl W, Bailey JN, et al. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat Genet*. 2016;48(2):134–143. [PubMed: 26691988]
15. *Handbook of Cluster Analysis*. CRC Press. Taylor & Francis Group; 2015.
16. Calinski T, Harabasz JA dendrite method for cluster analysis. *Communications in Statistics*. 1974;3:1–27.
17. Mantel N The detection of disease clustering and a generalized regression approach. *Cancer Res*. 1967;27(2):209–220. [PubMed: 6018555]
18. Keenan TD, Agron E, Domalpally A, et al. Progression of Geographic Atrophy in Age-related Macular Degeneration: AREDS2 Report Number 16. *Ophthalmology*. 2018.
19. Hosoda Y, Miyake M, Yamashiro K, et al. Deep phenotype unsupervised machine learning revealed the significance of pachychoroid features in etiology and visual prognosis of age-related macular degeneration. *Sci Rep*. 2020;10(1):18423. [PubMed: 33116208]
20. Pool FM, Kiel C, Serrano L, Luthert PJ. Repository of proposed pathways and protein-protein interaction networks in age-related macular degeneration. *NPJ Aging Mech Dis*. 2020;6:2. [PubMed: 31934346]
21. Sadda SR, Guymer R, Holz FG, et al. Consensus Definition for Atrophy Associated with Age-Related Macular Degeneration on OCT: Classification of Atrophy Report 3. *Ophthalmology*. 2017.



**Table 1.**

Phenotypic characteristics used as input for cluster analysis.

Cross-sectional analyses	Longitudinal analyses
Age (years) *	Age *
Sex *	Sex *
White / non-white *	White / non-white *
Educational level (3 levels) *	Educational level (3 levels) *
Smoking status (current / former / never) *	Smoking status (current / former / never) *
Square root of GA area (mm)	-
GA central involvement (yes / no)	GA central involvement (yes / no) <sup>‡</sup>
GA configuration (small / multifocal / horseshoe or ring / solid or unifocal / indeterminate)	GA configuration (small / multifocal / horseshoe or ring / solid or unifocal / indeterminate) <sup>§</sup>
GA fellow eye involvement (yes / no)	GA fellow eye involvement (yes / no) <sup>‡</sup>
Square root GA enlargement rate (mm/year)	Square root GA enlargement rate (mm/year)
Total drusen area within AREDS grid (7 levels) <sup>‡</sup>	Total drusen area within AREDS grid (7 levels) <sup>‡‡</sup>
Maximum drusen size within AREDS grid (4 levels) <sup>‡</sup>	Maximum drusen size within AREDS grid (4 levels) <sup>‡‡</sup>
Calcified drusen presence (yes / no)	Calcified drusen presence (yes / no) <sup>‡</sup>
Reticular pseudodrusen score (0.0–1.0)	Reticular pseudodrusen score (0.0–1.0) <sup>‡</sup>
BCVA (ETDRS letter score)	BCVA rate (change in ETDRS letter score/year)

\* defined at AREDS2 baseline visit; all other characteristics in the cross-sectional analyses (apart from GA enlargement rate) were defined at first appearance of geographic atrophy

<sup>‡</sup> as described in AREDS2 Report 2 (Danis et al, IOVS 2013)

<sup>‡‡</sup> defined as the maximum value during follow-up

<sup>§</sup> defined as most common configuration during follow-up (see text)

Abbreviations: AREDS=Age-Related Eye Disease Study; BCVA=best-corrected visual acuity; ETDRS=Early Treatment Diabetic Retinopathy Study; GA=geographic atrophy

**Table 2.**

Genes used to derive each pathway-based genetic risk score.

Pathway-based genetic risk score	Genes
Complement pathway	<i>C3</i> (3 SNPs), <i>C9</i> (1 SNP), <i>CFH</i> (8 SNPs), <i>CFB/C2</i> (4 SNPs), <i>CFI</i> (2 SNPs), <i>TMEM97/VTN</i> (1 SNP)
Lipid metabolism	<i>ABCA1</i> (1 SNP), <i>APOE</i> (2 SNPs), <i>CETP</i> (2 SNPs), <i>LIPC</i> (2 SNPs)
Extracellular matrix remodeling	<i>ADAMTS9</i> (1 SNP), <i>COL4A3</i> (1 SNP), <i>COL8A1</i> (2 SNPs), <i>SYN3/TIMP3</i> (1 SNP), <i>VEGF-A</i> (1 SNP)
<i>ARMS2</i>	<i>ARMS2</i> (1 SNP)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3.**

Demographic, phenotypic, and genetic characteristics of the study population.

Variable		All Subjects (N=598)
Age (years): median (IQR)		75.7 (70.0, 79.8)
Male: n (%)		258 (43.1)
White: n (%)		593 (99.2)
Education: n (%)	High School or Less	211 (35.3)
	At least some College	279 (46.7)
	Post-graduate	108 (18.1)
Smoking status: n (%)	Current	38 (6.4)
	Former	310 (51.8)
	Never	250 (41.8)
Central GA: n (%)		192 (32.1)
Calcified Drusen: n (%)		260 (43.5)
GA Configuration: n (%)	Small (single patch <1DA)	368 (61.5)
	Multifocal	129 (21.6)
	Horseshoe, Ring	12 (2.0)
	Solid (center or not)	75 (12.5)
	Indeterminate	14 (2.3)
Drusen Area Within the ETDRS Grid: n (%)	< circle C1	1 (0.2)
	< circle C2	6 (1.0)
	< circle I2	8 (1.3)
	< circle O2	27 (4.5)
	< 1/2 DA	114 (19.1)
	< 1 DA	157 (26.3)
	>= 1 DA	285 (47.7)
Maximum Drusen Size: n (%)	<63 $\mu$ m (circle C0)	2 (0.3)
	<125 $\mu$ m (circle C1)	8 (1.3)
	<250 $\mu$ m (circle C2)	287 (48.0)
	>=250 $\mu$ m (circle C2)	301 (50.3)
GA in Fellow Eye: n (%)		153 (25.6)
RPD score: median (IQR)		0.37 (0.12, 0.78)
Square Root of GA area (mm): median (IQR)		0.9 (0.6, 1.3)
GA Enlargement from Regression of Square Root of GA area (mm/year): median (IQR)		0.23 (0.07, 0.46)
Visual acuity (ETDRS letters): median (IQR)		75.0 (65.0, 82.0)
Follow-up time from first appearance of GA (years): median (IQR)		2.6 (1.5, 3.4)
Variable		Subjects with genetic data (N=313)
52 SNP-based Genetic Risk Score: median (IQR)		15.5 (14.6, 16.4)
Complement GRS: median (IQR)		8.7 (8.1, 9.2)

Extracellular matrix GRS: median (IQR)		0.9 (0.8, 1.0)
Lipid metabolism GRS: median (IQR)		1.8 (1.6, 1.9)
ARMS2 GRS: median (IQR)		1.1 (0, 1.1)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 4.** Demographic, phenotypic, and genetic characteristics of the cross-sectional phenotypic clusters.

Variable	All participants (N=598)	K-means Cluster Analysis			Agglomerative Hierarchical Cluster Analysis					
		Cluster A (N=367)	Cluster B (N=231)	Cluster C (N=469)	Cluster D (N=112)	Cluster E (N=12)	Cluster F (N=5)			
Age (years): median (IQR)	75.7 (70.0, 79.8)	75.3 (69.4, 79.5)	76.3 (70.8, 80.2)	75.6 (69.6, 79.6)	76.7 (71.5, 81.0)	74.2 (66.6, 77.8)	73.8 (61.5, 76.0)			
Male: n (%)	258 (43.1)	161 (43.9)	97 (42.0)	207 (44.1)	44 (39.3)	5 (41.7)	2 (40.0)			
White: n (%)	593 (99.2)	363 (98.9)	230 (99.6)	469 (100.0)	112 (100.0)	12 (100.0)	0 (0.0)			
Education: n (%)	211 (35.3)	126 (34.3)	85 (36.8)	146 (31.1)	60 (53.6)	3 (25.0)	2 (40.0)			
At least some College	279 (46.7)	168 (45.8)	111 (48.1)	224 (47.8)	48 (42.9)	5 (41.7)	2 (40.0)			
Post-graduate	108 (18.1)	73 (19.9)	35 (15.2)	99 (21.1)	4 (3.6)	4 (33.3)	1 (20.0)			
Current	38 (6.4)	16 (4.4)	22 (9.5)	31 (6.6)	5 (4.5)	1 (8.3)	1 (20.0)			
Former	310 (51.8)	192 (52.3)	118 (51.1)	245 (52.2)	57 (50.9)	6 (50.0)	2 (40.0)			
Never	250 (41.8)	159 (43.3)	91 (39.4)	193 (41.2)	50 (44.6)	5 (41.7)	2 (40.0)			
Central GA: n (%)	192 (32.1)	108 (29.4)	84 (36.4)	137 (29.2)	52 (46.4)	2 (16.7)	1 (20.0)			
Calcified Drusen: n (%)	260 (43.5)	148 (40.3)	112 (48.5)	228 (48.6)	27 (24.1)	3 (25.0)	2 (40.0)			
GA Configuration: n (%)	368 (61.5)	367 (100.0)	1 (0.4)	313 (66.7)	51 (45.5)	0 (0.0)	4 (80.0)			
Small (single patch <1DA)	129 (21.6)	0 (0.0)	129 (55.8)	127 (27.1)	2 (1.8)	0 (0.0)	0 (0.0)			
Multifocal	12 (2.0)	0 (0.0)	12 (5.2)	0 (0.0)	0 (0.0)	12 (100.0)	0 (0.0)			
Horseshoe, Ring	75 (12.5)	0 (0.0)	75 (32.5)	15 (3.2)	59 (52.7)	0 (0.0)	1 (20.0)			
Solid (center or not)	14 (2.3)	0 (0.0)	14 (6.1)	14 (3.0)	0 (0.0)	0 (0.0)	0 (0.0)			
Indeterminate	1 (0.2)	1 (0.3)	0 (0.0)	0 (0.0)	1 (0.9)	0 (0.0)	0 (0.0)			
Drusen Area Within the ETDRS Grid: n (%)	6 (1.0)	4 (1.1)	2 (0.9)	0 (0.0)	6 (5.4)	0 (0.0)	0 (0.0)			
< circle C1	8 (1.3)	5 (1.4)	3 (1.3)	1 (0.2)	7 (6.3)	0 (0.0)	0 (0.0)			
< circle C2	27 (4.5)	20 (5.4)	7 (3.0)	15 (3.2)	11 (9.8)	1 (8.3)	0 (0.0)			
< circle O2	114 (19.1)	66 (18.0)	48 (20.8)	74 (15.8)	38 (33.9)	2 (16.7)	0 (0.0)			
< 1/2 DA	157 (26.3)	97 (26.4)	60 (26.0)	129 (27.5)	24 (21.4)	2 (16.7)	2 (40.0)			

	>= 1 DA	285 (47.7)	174 (47.4)	111 (48.1)	<b>250 (53.3)</b>	<b>25 (22.3)</b>	<b>7 (58.3)</b>	3 (60.0)
Maximum Drusen Size: n (%)	<63 µm (circle C0)	2 (0.3)	2 (0.5)	0 (0.0)	0 (0.0)	2 (1.8)	0 (0.0)	0 (0.0)
	<125 µm (circle C1)	8 (1.3)	6 (1.6)	2 (0.9)	1 (0.2)	7 (6.3)	0 (0.0)	0 (0.0)
	<250 µm (circle C2)	287 (48.0)	164 (44.7)	123 (53.2)	<b>201 (42.9)</b>	<b>77 (68.8)</b>	<b>6 (50.0)</b>	3 (60.0)
	>=250 µm (circle C2)	301 (50.3)	195 (53.1)	106 (45.9)	<b>267 (56.9)</b>	<b>26 (23.2)</b>	<b>6 (50.0)</b>	2 (40.0)
GA in Fellow Eye: n (%)		153 (25.6)	<b>69 (18.8)</b>	<b>84 (36.4)</b>	118 (25.2)	28 (25.0)	5 (41.7)	2 (40.0)
RPD score: median (IQR)		0.37 (0.12, 0.78)	0.29 (0.08, 0.74)	0.50 (0.17, 0.82)	0.37 (0.12, 0.77)	0.36 (0.07, 0.82)	0.46 (0.25, 0.73)	0.20 (0.08, 0.50)
Square Root of GA area (mm): median (IQR)		0.9 (0.6, 1.3)	<b>0.7 (0.6, 1.0)</b>	<b>1.3 (1.0, 1.9)</b>	<b>0.9 (0.6, 1.2)</b>	<b>1.1 (0.7, 2.0)</b>	<b>2.5 (1.8, 3.1)</b>	0.7 (0.6, 0.9)
GA Enlargement from Regression of Square Root of GA area (mm/year): median (IQR)		0.23 (0.07, 0.46)	<b>0.17 (0.05, 0.37)</b>	<b>0.34 (0.14, 0.55)</b>	<b>0.22 (0.07, 0.46)</b>	<b>0.26 (0.08, 0.45)</b>	<b>0.20 (-0.15, 0.44)</b>	0.11 (0.09, 0.18)
Visual acuity (ETDRS letters): median (IQR)		75.0 (65.0, 82.0)	75.0 (66.0, 82.0)	74.0 (60.0, 82.0)	75.0 (66.0, 83.0)	72.0 (58.0, 80.0)	69.5 (52.0, 82.5)	72.0 (59.0, 73.0)
<b>Variable</b>			<b>Subjects with genetic data (N=313)</b>	<b>Subjects with genetic data (N=313)</b>	<b>Subjects with genetic data (N=313)</b>			
52 SNP-based Genetic Risk Score: median (IQR)		15.5 (14.6, 16.4)	15.4 (14.6, 16.4)	15.8 (14.5, 16.6)	15.6 (14.6, 16.4)	15.4 (14.4, 16.5)	15.6 (14.4, 16.4)	16.4 (12.6, 16.5)
Complement GRS: median (IQR)		8.7 (8.1, 9.2)	8.7 (8.1, 9.2)	8.7 (7.9, 9.2)	8.8 (8.2, 9.2)	8.6 (8.0, 9.0)	8.3 (8.0, 9.2)	9.2 (7.7, 9.6)
Extracellular matrix GRS: median (IQR)		0.9 (0.8, 1.0)	0.9 (0.7, 1.0)	0.9 (0.8, 1.0)	0.9 (0.8, 1.0)	0.8 (0.7, 0.9)	0.9 (0.9, 1.1)	1.0 (0.8, 1.1)
Lipid metabolism GRS: median (IQR)		1.8 (1.6, 1.9)	1.7 (1.5, 1.9)	1.8 (1.6, 1.9)	1.8 (1.6, 1.9)	1.8 (1.6, 1.9)	1.8 (1.6, 1.8)	1.5 (1.1, 1.5)
ARMS2 GRS: median (IQR)		1.1 (0.1, 1.1)	1.1 (0.1, 1.1)	1.1 (0.2, 2.1)	1.1 (0.1, 1.1)	1.1 (0.2, 2.1)	1.1 (1.1, 1.1)	1.1 (0.1, 1.1)

Characteristics with apparent differences between clusters are highlighted in bold (for the larger clusters A-E)

**Table 5.**

Demographic, phenotypic, and genetic characteristics of the longitudinal phenotypic clusters.

Variable		Cluster G (N=310)	Cluster H (N=288)
Age (years): median (IQR)		75.7 (70.0, 79.7)	75.7 (70.0, 79.9)
Male: n (%)		<b>169 (54.5)</b>	<b>89 (30.9)</b>
White: n (%)		308 (99.4)	285 (99.0)
Education: n (%)	High School or Less	111 (35.8)	100 (34.7)
	At least some College	147 (47.4)	132 (45.8)
	Post-graduate	52 (16.8)	56 (19.4)
Smoking status: n (%)	Current	0 (0.0)	<b>38 (13.2)</b>
	Former	<b>310 (100.0)</b>	0 (0.0)
	Never	0 (0.0)	<b>250 (86.8)</b>
Central GA: n (%) <sup>*</sup>		183 (59.0)	167 (58.0)
Calcified Drusen: n (%) <sup>*</sup>		197 (63.5)	187 (64.9)
GA Configuration: n (%) <sup>†</sup>	Small (single patch <1DA)	128 (41.3)	119 (41.3)
	Multifocal	82 (26.5)	76 (26.4)
	Horseshoe, Ring	11 (3.5)	11 (3.8)
	Solid (center or not)	72 (23.2)	77 (26.7)
	Indeterminate	17 (5.5)	5 (1.7)
Drusen Area Within the ETDRS Grid: n (%) <sup>*</sup>	< circle C1	0 (0.0)	0 (0.0)
	< circle C2	0 (0.0)	0 (0.0)
	< circle I2	0 (0.0)	1 (0.3)
	< circle O2	3 (1.0)	10 (3.5)
	< 1/2 DA	33 (10.6)	27 (9.4)
	< 1 DA	84 (27.1)	58 (20.1)
	>= 1 DA	190 (61.3)	192 (66.7)
Maximum Drusen Size: n (%) <sup>*</sup>	<63 μm (circle C0)	0 (0.0)	0 (0.0)
	<125 μm (circle C1)	0 (0.0)	1 (0.3)
	<250 μm (circle C2)	95 (30.6)	82 (28.5)
	>=250 μm (circle C2)	215 (69.4)	205 (71.2)
GA in Fellow Eye: n (%) <sup>*</sup>		151 (48.7)	128 (44.4)
RPD score: median (IQR) <sup>*</sup>		0.64 (0.27, 0.87)	0.66 (0.28, 0.87)
GA Enlargement from Regression of Square Root of GA area (mm/year): median (IQR)		<b>0.25 (0.08, 0.51)</b>	<b>0.21 (0.06, 0.43)</b>
Slope from Regression of visual acuity (ETDRS letters/year): median (IQR)		-1.7 (-4.7, 0.5)	-1.4 (-3.7, 0.7)
Follow-up time from first appearance of GA (years): median (IQR)		2.6 (1.6, 3.3)	2.6 (1.3, 3.4)
<b>Variable</b>		<b>Subjects with genetic data (N=313)</b>	
52 SNP-based Genetic Risk Score: median (IQR)		15.7 (14.7, 16.4)	15.4 (14.5, 16.5)

Complement GRS: median (IQR)		8.7 (8.1, 9.3)	8.8 (8.1, 9.2)
Extracellular matrix GRS: median (IQR)		0.9 (0.8, 1.0)	0.9 (0.7, 1.0)
Lipid metabolism GRS: median (IQR)		1.7 (1.5, 1.9)	1.8 (1.6, 1.9)
<i>ARMS2</i> GRS: median (IQR)		1.1 (0, 1.1)	1.1 (0, 1.1)

\* defined as the maximum value during follow-up

† defined as most common configuration during follow-up

Characteristics with apparent differences between clusters are highlighted in bold

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Table 6.**

Results: p-values for pairwise comparisons of the phenotypic clusters, according to the four pathway-based genetic risk scores, by t test.

	Cross-sectional dataset				Longitudinal dataset	
	A vs B		C vs D		G vs H	
Pathway-based genetic risk score	Raw	Adjusted*	Raw	Adjusted*	Raw	Adjusted*
Complement	0.75	1.00	0.14	1.00	0.74	1.00
Lipid metabolism	0.54	1.00	0.81	1.00	0.43	1.00
Extracellular matrix	0.86	1.00	0.16	1.00	0.086	1.00
<i>ARMS2</i>	0.008	0.096	0.55	1.00	0.59	1.00

\* Adjusted for multiple testing: adjusted for the 4 genetic risk scores by MULTTEST bootstrap and adjusted for the 3 cluster groupings by multiplying by 3.

**Table 7.**

Demographic, phenotypic, and genetic characteristics of the genetic clusters.

Variable	K-means Cluster Analysis			Agglomerative Hierarchical Cluster Analysis		
	Cluster I (N=148)	Cluster J (N=165)	Cluster K (N=106)	Cluster L (N=135)	Neither K nor L (N=72)	
Age (years): median (IQR)	74.5 (69.5, 78.7)	74.9 (68.9, 79.4)	75.1 (70.0, 79.3)	75.1 (69.3, 79.5)	73.3 (66.4, 77.4)	
Male: n (%)	68 (45.9)	58 (35.2)	50 (47.2)	49 (36.3)	27 (37.5)	
White: n (%)	145 (98.0)	165 (100.0)	106 (100.0)	135 (100.0)	69 (95.8)	
Education: n (%)	51 (34.5)	56 (33.9)	34 (32.1)	48 (35.6)	25 (34.7)	
At least some College	72 (48.6)	76 (46.1)	54 (50.9)	63 (46.7)	31 (43.1)	
Post-graduate	25 (16.9)	33 (20.0)	18 (17.0)	24 (17.8)	16 (22.2)	
Current	6 (4.1)	10 (6.1)	4 (3.8)	8 (5.9)	4 (5.6)	
Former	77 (52.0)	90 (54.5)	52 (49.1)	75 (55.6)	40 (55.6)	
Never	65 (43.9)	65 (39.4)	50 (47.2)	52 (38.5)	28 (38.9)	
Central GA: n (%)	48 (32.4)	52 (31.5)	40 (37.7)	45 (33.3)	15 (20.8)	
Calcified Drusen: n (%)	60 (40.5)	72 (43.6)	42 (39.6)	57 (42.2)	33 (45.8)	
GA Configuration: n (%)	105 (70.9)	95 (57.6)	76 (71.7)	74 (54.8)	50 (69.4)	
Small (single patch <IDA)	26 (17.6)	38 (23.0)	18 (17.0)	35 (25.9)	11 (15.3)	
Multifocal	2 (1.4)	4 (2.4)	1 (0.9)	4 (3.0)	1 (1.4)	
Horseshoe, Ring	14 (9.5)	25 (15.2)	10 (9.4)	20 (14.8)	9 (12.5)	
Solid (center or not)	1 (0.7)	3 (1.8)	1 (0.9)	2 (1.5)	1 (1.4)	
Indeterminate	1 (0.7)	0 (0.0)	1 (0.9)	0 (0.0)	0 (0.0)	
< circle C1	2 (1.4)	0 (0.0)	2 (1.9)	0 (0.0)	0 (0.0)	
< circle C2	0 (0.0)	3 (1.8)	0 (0.0)	3 (2.2)	0 (0.0)	
< circle I2	8 (5.4)	5 (3.0)	5 (4.7)	5 (3.7)	3 (4.2)	
< circle O2	31 (20.9)	29 (17.6)	25 (23.6)	22 (16.3)	13 (18.1)	
< 1/2 DA	31 (20.9)	50 (30.3)	20 (18.9)	37 (27.4)	24 (33.3)	
< 1 DA	75 (50.7)	78 (47.3)	53 (50.0)	68 (50.4)	32 (44.4)	
>= 1 DA	1 (0.7)	0 (0.0)	1 (0.9)	0 (0.0)	0 (0.0)	
Maximum Drusen Size: n (%)	3 (2.0)	2 (1.2)	3 (2.8)	1 (0.7)	1 (1.4)	
<63 µm (circle C0)	66 (44.6)	79 (47.9)	45 (42.5)	69 (51.1)	31 (43.1)	
<125 µm (circle C1)						
<250 µm (circle C2)						

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Variable	K-means Cluster Analysis		Agglomerative Hierarchical Cluster Analysis		
	Cluster I (N=148)	Cluster J (N=165)	Cluster K (N=106)	Cluster L (N=135)	Neither K nor L (N=72)
>=250 µm (circle C2)	78 (52.7)	84 (50.9)	57 (53.8)	65 (48.1)	40 (55.6)
GA in Fellow Eye: n (%)	35 (23.6)	36 (21.8)	26 (24.5)	31 (23.0)	14 (19.4)
RPD score: median (IQR)	0.31 (0.08, 0.74)	0.34 (0.11, 0.77)	0.28 (0.06, 0.69)	0.37 (0.13, 0.80)	0.32 (0.11, 0.77)
Square Root of GA area (mm): median (IQR)	0.8 (0.6, 1.2)	0.9 (0.6, 1.3)	0.8 (0.6, 1.2)	0.9 (0.6, 1.3)	0.9 (0.6, 1.3)
GA Enlargement from Regression of Square Root of GA area (mm/year): median (IQR)	<b>0.18 (0.08, 0.37)</b>	<b>0.31 (0.09, 0.51)</b>	<b>0.19 (0.08, 0.37)</b>	<b>0.33 (0.08, 0.55)</b>	0.22 (0.10, 0.39)
Visual acuity (ETDRS letters): median (IQR)	77.0 (68.0, 83.0)	75.0 (65.0, 81.0)	76.0 (66.0, 83.0)	75.0 (64.0, 81.0)	75.0 (68.0, 82.0)
S2 SNP-based Genetic Risk Score: median (IQR)	15.1 (14.3, 16.1)	16.0 (15.0, 16.7)	15.1 (14.3, 15.8)	16.0 (15.0, 16.6)	15.9 (14.6, 16.5)
Complement GRS: median (IQR)	9.0 (8.3, 9.4)	8.5 (7.9, 9.0)	9.0 (8.3, 9.5)	8.5 (8.1, 9.0)	8.7 (8.0, 9.3)
Extracellular matrix GRS: median (IQR)	0.9 (0.8, 1.1)	0.8 (0.7, 0.9)	0.9 (0.8, 1.0)	0.8 (0.7, 0.9)	1.0 (0.8, 1.2)
Lipid metabolism GRS: median (IQR)	1.7 (1.5, 1.8)	1.8 (1.6, 1.9)	1.7 (1.5, 1.8)	1.8 (1.7, 1.9)	1.6 (1.3, 1.9)
ARMS2 GRS: median (IQR)	<b>0.0 (0, 1.1)</b>	<b>1.1 (1.1, 2.1)</b>	<b>0.0 (0, 0)</b>	<b>1.1 (1.1, 2.1)</b>	1.1 (0, 2.1)

Characteristics with apparent differences between clusters are highlighted in bold

**Table 8.**

Results of the Mantel test to examine potential relationships between the genetic and phenotypic characteristics.

Comparison	Mantel p-value	r
4 GRSs versus cross-sectional phenotype	0.66	-0.02
4 GRSs versus longitudinal phenotype	0.56	0.02
52-SNP GRS versus cross-sectional phenotype	0.28	0.03
52-SNP GRS versus longitudinal phenotype	0.07	0.06

Abbreviations: GRS=genetic risk score

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript