Geographic distribution of rare variants associated with age-related macular degeneration

Maartje J. Geerlings,1 Eveline Kersten,1 Joannes M.M. Groenewoud,2 Lars G. Fritsche,3 Carel B. Hoyng,1 Eiko K. de Jong,1 Anneke I. den Hollander1,4

(The first two authors contributed equally to this work.)

1Department of Ophthalmology, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands; 2Department for Health Evidence, Radboud University Medical Center, Nijmegen, the Netherlands; 3Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology Trondheim, Norway; 4Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands

Purpose: A recent genome-wide association study by the International Age-related Macular Degeneration Genomics Consortium (IAMDGC) identified seven rare variants that are individually associated with age-related macular degeneration (AMD), the most common cause of vision loss in the elderly. In literature, several of these rare variants have been reported with different frequencies and odds ratios across populations of Europe and North America. Here, we aim to describe the representation of these seven AMD-associated rare variants in different geographic regions based on 24 AMD studies.

Methods: We explored the occurrence of seven rare variants independently associated with AMD (CFH rs121913059 [p.Arg1210Cys], CFI rs141853578 [p.Gly119Arg], C3 rs147859257 [p.Lys155Gln], and C9 rs34882957 [p.Pro167Ser]) and three non-coding variants in or near the CFH gene (rs148553336, rs35292876, and rs191281603) in 24 AMD case-control studies. We studied the difference in distribution, interaction, and effect size for each of the rare variants based on the minor allele frequency within the different geographic regions.

Results: We demonstrate that two rare AMD-associated variants in the CFH gene (rs121913059 [p.Arg1210Cys] and rs35292876) deviate in frequency among different geographic regions (p=0.004 and p=0.001, respectively). The risk estimates of each of the seven rare variants were comparable across the geographic regions.

Conclusions: The results emphasize the importance of identifying population-specific rare variants, for example, by performing sequencing studies in case-control studies of various populations, because their identification may have implications for diagnostic screening and personalized treatment.

Genetic diversity is observed among populations of different ancestries. Allele frequencies can exhibit large diversity among populations due to forces such as genetic drift and natural selection. Although most common variants are shared worldwide, rare variants (minor allele frequency [MAF] <1%) have the tendency to cluster in specific populations. Population-specific rare variants tend to have a strong functional effect [1].

In age-related macular degeneration (AMD), large variability in rare variant frequency has been reported in case-control studies of various populations; for instance, for variant rs121913059 (p.Arg1210Cys) in complement factor H (CFH; HGNC 4883, OMIM 134370). CFH rs121913059 was first reported in a case-control study from the United States [2]. Several studies replicated the finding [3-5], but other Caucasian studies [6-9] and Asian studies [10,11] were unable to replicate the strong association (Table 1). Another example, variant rs141853578 (p.Gly119Arg) in complement factor I (CFI; HGNC 5394, OMIM 217030) first reported in a European cohort [7], was screened in a British [12] and American [13] cohort (odds ratio [OR] = 22.2; 8.5 and 2.6, respectively). However, although the variant was associated with AMD, the risk effect size was much weaker when compared to the first report.

In a recent genome-wide association study conducted by the International Age-related Macular Degeneration Genomics Consortium (IAMDGC) [3], seven rare variants were observed to independently confer risk for AMD. All seven rare variants are localized in or near genes encoding components of the complement system, namely, CFH, CFI, and complement components 3 and 9 (C3, HGNC 1318, OMIM 120700 and C9, HGNC: 1358, OMIM 120940).
## Table 1. Minor allele frequencies of the CFH rs121913059 (p.Arg1210Cys) variant among different geographical regions reported in literature.

<table>
<thead>
<tr>
<th>Descent</th>
<th>Previous reports</th>
<th>Carriers (n)</th>
<th>Total Cases (n)</th>
<th>Total Controls (n)</th>
<th>MAF Cases (%)‡</th>
<th>MAF Controls (%)‡</th>
<th>Odds-Ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Various populations, European descent</td>
<td>Fritsche 2016 [3]</td>
<td>108</td>
<td>16,144</td>
<td>17,832</td>
<td>0.319</td>
<td>0.014</td>
<td>20.3</td>
<td>8.9×10−24</td>
</tr>
<tr>
<td>Eastern USA</td>
<td>Raychaudhuri 2011 [2]*</td>
<td>34</td>
<td>2414</td>
<td>1120</td>
<td>0.684</td>
<td>0.045</td>
<td>NA</td>
<td>8.0 × 10−5</td>
</tr>
<tr>
<td></td>
<td>Zhan 2013 [5]</td>
<td>24</td>
<td>2268</td>
<td>2268</td>
<td>0.507</td>
<td>0.022</td>
<td>23.1</td>
<td>2.9 × 10−6</td>
</tr>
<tr>
<td></td>
<td>Helgason 2013 [8]</td>
<td>0</td>
<td>1143</td>
<td>51,435</td>
<td>0.000</td>
<td>0.000</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>European</td>
<td>Saksens 2016 [9]*</td>
<td>0</td>
<td>1589</td>
<td>1386</td>
<td>0.000</td>
<td>0.000</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Recalde 2016 [4]</td>
<td>5</td>
<td>259</td>
<td>330</td>
<td>0.965</td>
<td>0.000</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Asian</td>
<td>Shen 2012 [11]</td>
<td>0</td>
<td>258</td>
<td>426</td>
<td>0.000</td>
<td>0.000</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Miyake 2015 [10]</td>
<td>1</td>
<td>1364</td>
<td>1208</td>
<td>0.037</td>
<td>0.000</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Additional publications from: the Boston study [13,15] and EUGENDA study [6,7]. ‡ major allele C, minor allele T. NA=Not available or not reported.
The difference in association for rare variants among different AMD case-control studies may reflect the difference in distribution of such rare alleles across geographic regions. This observation raises the question whether these variants identified by the IAMDGC are represented in all case-control studies or whether the association is driven by one or more studies from a specific geographic region. Therefore, we sought to evaluate the representation of these seven AMD-associated rare variants in 24 AMD case-control studies of different geographic regions.

METHODS

Data for this study were provided by the IAMDGC. The genotypes are in part available via dbGaP under accession number phs001039.v1.p1. The original data set contained data from 40,633 individuals of European ancestry as described by Fritsche et al. [3].

For analyses of the current study, participants from the Utah case-control study were excluded due to their mixed regions of origin. In addition, the Jerusalem case-control study was excluded due to its small sample size compared to the other geographic regions. Final analyses were performed on 39,582 participants derived from 24 of 26 studies [3]. The included studies were grouped in five geographic regions: eastern USA, western Europe, Britain, western USA, and Australia (Table S1). Data were collected by all study groups in accordance with the tenets of the Declaration of Helsinki; participants provided informed consent, and study protocols were approved by local ethical committees [3]. Ancestry was determined based on the first two principal components using autosomal genotyped variants together with genotype information of the samples from the Human Genome Diversity Project [3]. The final data set included 39,582 successfully genotyped subjects, including 15,527 advanced AMD cases, 6,537 non-advanced AMD cases, and 17,518 control individuals of European ancestry.

The MAF in each region was calculated and compared independently of AMD status. For comparison of effect sizes and interaction analyses, individuals were assigned “AMD” when they exhibited signs of (1) advanced AMD defined as geographic atrophy and/or choroidal neovascularization in at least one eye, or (2) non-advanced AMD defined as pigmentary changes in the macula and/or more than five macular drusen with a diameter more than 63 μm. Individuals without any reported signs of AMD were assigned “No AMD.”

Genotype data of seven rare genetic variants were selected from array-based data generated by the IAMDGC [3]. Fritsche et al. [3] showed these seven rare variants were independently associated with AMD: CFH rs121913059 (p.Arg1210Cys), CFI rs141853578 (p.Gly119Arg), C3 rs147859257 (p.Lys155Gln), and C9 rs34882957 (p.Pro67Ser) and three non-coding variants in or near CFH (rs148553336, rs35292876, and rs191281603).

The software package SAS (Statistical Analysis System Institute, SAS Institute Inc., Cary, NC, V9.2) was used to compare MAFs between the different geographic regions in a logistic regression analysis with the Firth correction (S1 Supporting information) [14]. Furthermore, we estimated the mean allele frequency of each rare genetic variant in each of the geographic regions including a 95% confidence interval (details provided in S1 Supporting information). To study the potential difference in the effect size of each variant between the geographic regions, interaction analyses were performed using binary logistic regression models with SPSS statistics software (IBM SPSS Statistics, IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY, V22.0).

RESULTS

Demographic characteristics for each of the five geographic regions are shown in Supporting Information Table S1. The characteristics of the participants from the different regions were comparable, although the British study samples were slightly younger than the others, and the western European study samples included relatively more female participants compared to the remainder. These differences were comparable in the cases and the controls.

We analyzed the difference in distribution of the seven rare variants among case-control studies from eastern USA, western Europe, Britain, western USA, and Australia using logistic regression analysis with the Firth correction (Table 2; Figure 1) and observed a difference in the distribution of the variants CFH rs121913059 (p.Arg1210Cys, p=0.004) and CFH rs35292876 (p=0.001) across the different geographic regions. CFH rs121913059 was found at a higher frequency in eastern USA, especially compared to Britain and Australia (p=0.011 and p=0.003, respectively). CFH rs35292876 was found at a higher frequency in western Europe, compared to all other regions (ranging from p<0.001 in Britain to p=0.012 in eastern USA). The distribution of variants CFH rs121913059 (p.Arg1210Cys) and CFH rs35292876 was not confounded by the distribution of individuals with advanced and non-advanced AMD across the geographic regions (Table S2). The other five variants were found to have similar allele frequencies among all geographic regions.

The difference in distribution is also reflected by the estimated MAFs of each variant in the different geographic regions (Table S3). The allele frequency of CFH rs121913059 is nearly three times higher in eastern USA than in Britain.
Noteworthy is the near absence of this risk variant in control individuals without AMD, indicating that the difference in distribution appears to be driven solely by individuals with AMD (Table S4).

To determine whether the effect size was influenced by geographic region, we performed interaction analyses for each variant. We observed that the risk associated with each specific rare variant is independent of the geographic region (Table 2). The overall effect sizes of the rare variants are comparable to the effect sizes reported in the IAMDGC study [3].

**DISCUSSION**

The distribution of rare CFH variants rs121913059 (p.Arg1210Cys) and rs35292876 was significantly different between several of the studied geographic regions. This result confirms differences reported in previous studies for the CFH rs121913059 variant [2-11,13,15] (Table 1). CFH rs121913059 was first associated with AMD in a study from the USA [2]; however, the association was not consistently replicated in Dutch/German [7], Icelandic [8], Japanese [10], and Chinese [11] studies. In this study, we confirmed the hypothesis that rare variants can be differentially distributed among geographic regions. As expected, the risk estimates are comparable across the geographic regions. Rare variants tend to be recent, and therefore, their distribution tends to be restricted to a specific region. Interpretation of rare variants, therefore, may be focused to a geographic region or population [16,17].

In AMD, a difference in geographic distribution has already been described for common risk haplotypes of the CFH and ARMS2 (HGNC: 32685, OMIM 611313) genes, which are the most prominent common genetic AMD risk factors [3]. Although Asian populations report a lower frequency of CFH risk haplotypes, the opposite holds true for the ARMS2/HTRA1 risk haplotype which is more prevalent in Asians compared to Caucasian populations [18,19]. These patient- and population-specific variations have implications for genetic counseling and carrier screening in diagnostic and research settings.

In addition to single variant associations, a significant burden of rare variants in the CFH and CFI genes has been reported for AMD [3,15]. The disease burden in these genes is attributed to the cumulative effect of rare coding variants, some of which are identified in multiple studies, while others are restricted to a single population or even a single patient [20]. Carriers of specific rare genetic variants in the complement genes that increase complement activation may benefit more from complement-inhibiting therapy than those who do not carry such variants [20]. Personalized treatment aiming at complement-activating rare variants in clinical trials may be applicable only to specific populations where these variants are sufficiently common.

It is likely that additional rare variants, other than CFH Arg1210Cys and rs35292876, fluctuate in frequency among geographic regions. To identify these variants, additional large sequencing studies will need to be performed in populations and Australia. Noteworthy is the near absence of this risk variant in control individuals without AMD, indicating that the difference in distribution appears to be driven solely by individuals with AMD (Table S4).

### Table 2. Distribution and Interaction Analysis of Seven Rare AMD-associated Genetic Variants Across Five Geographical Regions.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Difference in distribution between geographical regions*</th>
<th>Interaction Analysis*</th>
<th>Overall effect size‡</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p-value</td>
<td>p-value</td>
<td>OR (95%CI)</td>
</tr>
<tr>
<td>CFH rs121913059</td>
<td>0.004</td>
<td>0.665</td>
<td>24.2 (8.9–65.6)</td>
</tr>
<tr>
<td>(p.Arg1210Cys)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFI rs141853578</td>
<td>0.707</td>
<td>0.563</td>
<td>3.7 (2.5–5.7)</td>
</tr>
<tr>
<td>(p.Gly119Arg)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3 rs147859257</td>
<td>0.665</td>
<td>0.680</td>
<td>2.8 (2.3–3.4)</td>
</tr>
<tr>
<td>(p.Lys155Gln)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C9 rs34882957</td>
<td>0.315</td>
<td>0.572</td>
<td>1.7 (1.5–2.0)</td>
</tr>
<tr>
<td>(p.Pro167Ser)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFH rs148553336</td>
<td>0.053</td>
<td>0.015</td>
<td>0.5 (0.4–0.6)</td>
</tr>
<tr>
<td>CFH rs35292876</td>
<td><strong>0.001</strong></td>
<td>0.709</td>
<td>2.3 (2.0–2.6)</td>
</tr>
<tr>
<td>CFH rs191281603</td>
<td>0.735</td>
<td>0.980</td>
<td>0.9 (0.7–1.1)</td>
</tr>
</tbody>
</table>

*Logistic Regression with Firth correction. Individual Wald Chi-Square from likelihood ratio test for each of the variants across the geographical regions. *Interaction Analysis: Effect sizes in entire study and interaction analysis to study potential differences in effect size between cohorts. ‡Overall effect size adjusted for geographical region. Bold values: p value considered significant after Bonferroni correction (p<0.007).
originating from diverse geographic regions. Until now, large sequencing initiatives are predominantly of North American or European origin, and sample sizes for non-European-descent population are limited [3,21]. Recruiting case-control studies from other geographic regions and ancestries could allow for identification of novel highly penetrant rare variants implicated in AMD pathogenesis. These variants may be located in known AMD pathways, such as the complement system, or novel pathways [22]. For example, Asian populations have a different genetic predisposition for AMD than Europeans. Only half of the common European loci could be replicated in East Asians [23]. Furthermore, the rare variants CFH rs121913059 (p.Arg1210Cys), CFI rs141853578 (p.Gly119Arg), and C3 rs147859257 (p.Lys155Gln) were not identified in Asian populations according to public database gnomAD [24]. It has been postulated that rare variants in

Figure 1. The two rare variants in CFH that are differently distributed variants among different geographic regions. Minor allele frequencies (in percentage) for CFH rs121913059 (A) and CFH rs35292876 (B). Variants mapped to geographic location (from left to right): western USA, eastern USA, Britain, western Europe, and Australia.
Asian populations may be found in other genes, for example, involved in cholesterol and lipid metabolism [25].

In conclusion, we demonstrated that rare AMD-associated variants $CFH$ rs121913059 and rs35292876 are differently distributed among different geographic regions. These results emphasize the importance of identifying population-specific rare variants in AMD.

APPENDIX 1.
Code for the software package SAS for a logistic regression analysis and Firth’s bias correction. To access the data, click or select the words “Appendix 1”

APPENDIX 2.
List of members of the Age-related Macular Degeneration Genomics Consortium. To access the data, click or select the words “Appendix 2”

APPENDIX 3.
Demographic characteristics of AMD cohorts grouped in five geographical regions. To access the data, click or select the words “Appendix 3”

APPENDIX 4.
Overall estimated mean MAF of $CFH$ rs121913059 (p.Arg1210Cys) and $CFH$ rs35292876 across five geographical regions in the complete study and subdivided for individuals with and without advanced AMD. Advanced AMD defined as geographic atrophy and/or choroidal neovascularization in at least one eye. Non-advanced $CFH$ defined as pigmentary changes in the macula and/or more than five macular drusen with a diameter >63 μm or individuals without any reported signs of AMD. To access the data, click or select the words “Appendix 4”

APPENDIX 5.
Overall estimated mean MAF of seven rare AMD-associated genetic variants across five geographical regions. To access the data, click or select the words “Appendix 5”

APPENDIX 6.
Minor allele frequencies (%) of seven rare AMD-associated genetic variants across five geographical regions stratified by AMD status. To access the data, click or select the words “Appendix 6”

ACKNOWLEDGMENTS

We would like to acknowledge the contribution of the International AMD Genomics Consortium (IAMDG) that is supported by a grant from NIH (R01EY022310). Full list of IAMDG members is provided in S2 Appendix. Genotyping was supported by a contract (HHSN2682012000081) to the Center for Inherited Disease Research. The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP/2007–2013) / ERC Grant Agreement n. 310644 (MACULA). This project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 634479.

REFERENCES


