Insights into Rare and Common Genetic Variation
From a Large Study of Age-Related Macular Degeneration

Lars G. Fritsche[1†], Wilmar Igl[2†], Jessica N. Cooke Bailey[3†], Felix Grassmann[4†], Sebanti Sengupta[1†], Jennifer L. Bragg-Gresham[1,5], Kathryn P. Burdon[6], Scott J. Hebbring[7], Cindy Wen[8], Mathias Gorski[2], Ivana K. Kim[9], David Cho[10], Donald Zack[11,12,13,14,15], Eric Souied[16], Hendrik P. N. Scholl[11,17], Elisa Bala[18], Kristine E. Lee[19], David J. Hunter[20,21], Rebecca J. Sardell[22], Paul Mitchell[23], Joanna E. Merriam[24], Valentina Cipriani[25,26], Joshua D. Hoffman[27], Tina Schick[28], Yara T. E. Lechanteur[29], Robyn H. Guymer[30], Matthew P. Johnson[31], Yingda Jiang[32], Chloe M. Stanton[33], Gabriëlle H. S. Buitendijk[34,35], Xiaowei Zhan[1,36,37], Alan M. Kwong[1], Alexis Boleda[38], Matthew Brooks[39], Linn Gieser[38], Rinki Ratnapriya[38], Kari E. Branham[39], Johanna R. Foerster[1], John R. Heckenlively[39], Mohammad I. Othman[39], Brendan J. Vote[6], Helena Hai Liang[30], Emmanuelle Souzeau[40], Ian L. McAllister[41], Timothy Isaacs[41], Janette Hall[40], Stewart Lake[40], David A. Mackey[6,30,41], Ian J. Constable[41], Jamie E. Craig[40], Terrie E. Kitchner[7], Zhenglin Yang[42,43], Zhiguang Su[44], Hongrong Luo[8,44], Daniel Chen[8], Hong Ouyang[8], Ken Flagg[8], Danni Lin[8], Guanping Mao[8], Henry Ferreyra[8], Klaus Stark[2], Claudia N. von Strachwitz[45], Armin Wolf[46], Caroline Brandl[2,4,47], Guenther Rudolph[46], Matthias Olden[2], Margaux A. Morrison[48], Denise J. Morgan[48], Matthew Schu[49,50,51,52,53], Jeeyun Ahn[54], Giuliana Silvestri[55], Evangelia E. Tsironi[56], Kyu Hyung Park[57], Lindsay A. Farrer[49,50,51,52,53], Anton Orlin[58], Alexander Brucker[59], Mingyao Li[60], Christine Curcio[61], Saddek Mohand-Saïd[62,63,64,65], José-Alain Sahel[62,63,64,65,66,67,68], Isabelle Audo[62,63,64,69], Mustapha Benchaboune[65], Angela J. Cree[70], Christina A. Rennie[71], Srinivas V. Goverdhan[70], Michelle Grunin[72], Shira Hagbi-Levi[72], Peter Campochiaro[11,13], Nicholas Katsanis[73,74,75], Frank G. Holz[17], Frédéric Blond[62,63,64], Hélène Blanché[76], Jean-François Deleuze[76,77], Robert P. Igo Jr.[3], Barbara Truitt[3], Neal S. Peachey[18,78], Stacy M. Meuer[19], Chelsea E. Myers[19], Emily L. Moore[19], Ronald Klein[19], Michael A. Hauser[79,80,81], Eric A. Postel[79], Monique D. Courtenay[22], Stephen G. Schwartz[82], Jaclyn L. Kovach[82], William K. Scott[22], Gerald Liew[23], Ava G. Tƒan[23], Bamini Gopinath[23], John C. Merriam[24], R. Theodore Smith[24,84], Jane C. Khan[41,83,85] , Humma Shahid[85,86], Anthony T. Moore[25,26,87], J. Allie McGrath[27], Reneé Laux[3], Milam A. Brantley Jr.[88], Anita Agarwal[88], Lebriz Ersoy[28], Albert Caramoy[28], Thomas Langmann[28], Nicole T. M. Saksens[29], Eiko K. de Jong[29], Carel B. Hoyng[29], Melinda S. Cain[30], Andrea J. Richardson[30], Tammy M. Martin[89], John Blangero[31], Daniel E. Weeks[32,90], Bal Dhillon[91], Cornelia M. van Duijn[35], Kimberly F. Doheny[92], Jane Romm[92], Caroline C. W. Klaver[34,35], Caroline Hayward[33], Michael B. Gorin[93,94], Michael L. Klein[89], Paul N. Baird[30], Anneke I. den Hollander[29,95], Sascha Fauser[28], John R. W. Yates[25,26,85], Rando Allikmets[24,96], Jie Jin Wang[23], Debra A. Schaumberg[20,97,98], Barbara E. K. Klein[19], Stephanie A. Hagstrom[78], Itay Chowers[72], Andrew J. Lotery[70], Thierry Léveillard[62,63,64], Kang Zhang[8,44], Murray H. Brilliant[7], Alex W. Hewitt[6,30,41], Anand Swaroop[38],

Emily Y. Chew[99], Margaret A. Pericak-Vance[22‡], Margaret DeAngelis[48‡], Dwight Stambolian[10‡], Jonathan L. Haines[3,100‡], Sudha K. Iyengar[3‡*], Bernhard H. F. Weber[4‡], Gonçalo R. Abecasis[1‡*], Iris M. Heid[2‡*]

[1]Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA
[2]Department of Genetic Epidemiology, University of Regensburg, Germany
[3]Department of Epidemiology and Biostatistics, Case Western Reserve University School of Medicine, 2103 Cornell Rd Cleveland, OH 44106
[4]Institute of Human Genetics, University of Regensburg, Germany
[5]Kidney Epidemiology and Cost Center, Department of Biostatistics, Department of Internal Medicine - Nephrology, University of Michigan, Ann Arbor, MI 48109, USA
[6]School of Medicine, Menzies Research Institute Tasmania, University of Tasmania, Hobart, Tasmania, Australia
[7]Center for Human Genetics, Marshfield Clinic Research Foundation, 1000 N. Oak Ave. Marshfield, WI 54449, USA
[8]Department of Ophthalmology, University of California San Diego and VA San Diego Health System, La Jolla, California 92093, USA
[9]Retina Service, Massachusetts Eye and Ear, Department of Ophthalmology Harvard Medical School, Boston, MA, USA
[10]Department of Ophthalmology, Perelman School of Medicine, University of Pennsylvania
[11]Department of Ophthalmology, Wilmer Eye Institute - Johns Hopkins University School of Medicine - 400 North Broadway, Smith Building - Baltimore, Maryland 21287, USA
[12]Department of Molecular Biology and Genetics - Johns Hopkins University School of Medicine - 400 North Broadway, Smith Building - Baltimore, Maryland 21287, USA
[13]Department of Neuroscience - Johns Hopkins University School of Medicine - 400 North Broadway, Smith Building - Baltimore, Maryland 21287, USA
[14]Institute of Genetic Medicine - Johns Hopkins University School of Medicine - 400 North Broadway, Smith Building - Baltimore, Maryland 21287, USA
[15]Institue de la Vision, Université Pierre et Marie Curie, 17 rue Moreau, Paris, France
[16]Hôpital Intercommunal de Créteil, Hôpital Henri Mondor - Université Paris Est Créteil, France
[17]University of Bonn - Department of Ophthalmology - Ernst-Abbe-Str. 2 - D-53127 Bonn – Germany
[18]Louis Stokes Cleveland VA Medical Center, 10701 East Boulevard, Cleveland OH 44106, USA
[19]Department of Ophthalmology and Visual Sciences, University of Wisconsin, Madison, WI

[20]Department of Epidemiology, Harvard School of Public Health
[21]Department of Nutrition, Harvard School of Public Health
[22]John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, Florida, United States
[23]Centre for Vision Research, Department of Ophthalmology and Westmead Millennium Institute for Medical Research, University of Sydney, Sydney, Australia
[24]Department of Ophthalmology Columbia University, New York, NY
[25]UCL Institute of Ophthalmology, University College London, London, EC1V 9EL UK
[26]Moorfields Eye Hospital, London, EC1V 2PD UK
[27]Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, Tennessee, United States
[28]University Hospital of Cologne, Department of Ophthalmology, Kerpener Str. 62, 50924 Cologne, Germany
[29]Department of Ophthalmology, Radboud University Medical Centre, Nijmegen, the Netherlands
[30]Centre for Eye Research Australia, University of Melbourne, Royal Victorian Eye and Ear Hospital, East Melbourne, Victoria, 3000, Australia
[31]South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX 78520
[32]Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA
[33]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, EH4 2XU Scotland, UK
[34]Department of Ophthalmology, Erasmus Medical Center, Rotterdam, the Netherlands PO box 2040, 3000CA
[35]Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands PO box 2040, 3000CA
[36]Quantitative Biomedical Research Center, Department of Clinical Science, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390, USA
[37]Center for the Genetics of Host Defense, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390-8505, USA
[38]Neurobiology Neurodegeneration & Repair Laboratory (N-NRL), National Eye Institute, National Institutes of Health, Bethesda, MD 20892, USA
[39]Department of Ophthalmology and Visual Sciences, University of Michigan, Kellogg Eye Center, Ann Arbor, MI 48105, USA
[40]Department of Ophthalmology, Flinders Medical Centre, Flinders University, Adelaide, South Australia, Australia

[41]Centre for Ophthalmology and Visual Science, Lions Eye Institute, University of Western Australia, Perth, Western Australia, Australia
[42]Sichuan Provincial Key Laboratory for Human Disease Gene Study, Hospital of the University of Electronic Science and Technology of China and Sichuan Provincial People's Hospital, Chengdu, China
[43]Sichuan Translational Medicine Hospital, Chinese Academy of Sciences, Chengdu, China
[44]Molecular Medicine Research Center, State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Sichuan 610041, China
[45]EyeCentre Southwest Stuttgart
[46]University Eye Clinic, Ludwig-Maximilians-University Munich
[47]Department of Ophthalmology, University Hospital Regensburg, Regensburg, Germany
[48]Department of Ophthalmology and Visual Sciences, University of Utah, Salt Lake City, UT, USA
[49]Department of Medicine (Biomedical Genetics), Boston University Schools of Medicine and Public Health, Boston, MA, USA
[50]Department of Ophthalmology, Boston University Schools of Medicine and Public Health, Boston, MA, USA
[51]Department of Neurology, Boston University Schools of Medicine and Public Health, Boston, MA, USA
[52]Department of Epidemiology, Boston University Schools of Medicine and Public Health, Boston, MA, USA
[53]Department of Biostatistics, Boston University Schools of Medicine and Public Health, Boston, MA, USA
[54]Department of Ophthalmology, Seoul Metropolitan Government Seoul National University Boramae Medical Center, Seoul, Republic of Korea
[55]Centre for Experimental Medicine, Queen's University, Belfast, UK
[56]Department of Ophthalmology, University of Thessaly, School of Medicine, Larissa, Greece
[57]Department of Ophthalmology, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
[58]Department of Ophthalmology, Weill Cornell Medical College, New York, NY, USA
[59]Scheie Eye Institute, Department of Ophthalmology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104
[60]Department of Biostatistics and Epidemiology University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104
[61]Department of Ophthalmology, The University of Alabama at Birmingham 1670 University Boulevard VH 360 Birmingham, AL 35294-0019
[62]INSERM, U968, Paris, F-75012, France

[63]UPMC Univ Paris 06, UMR_S 968, Institut de la Vision, Department of Genetics, Paris, F-75012, France
[64]CNRS, UMR_7210, Paris, F-75012, France
[65]Centre Hospitalier National d'Ophtalmologie des Quinze-Vingts, INSERM-DHOS CIC 503, Paris, F-75012, France
[66]Fondation Ophtalmologique Adolphe de Rothschild, Paris, F-75019, France
[67]Institute of Ophthalmology, University College of London, London, WC1E 6BT, UK
[68]Académie des Sciences–Institut de France, Paris, F-75006 France
[69]Department of Molecular Genetics, Institute of Ophthalmology, London, UK
[70]Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, UK
[71]University Hospital Southampton, Southampton, UK
[72]Department of Ophthalmology, Hadassah Hebrew University Medical Center, Jerusalem. Israel
[73]Center for Human Disease Modeling, Duke University USA
[74]Department of Cell Biology, Duke University USA
[75]Department of Pediatrics, Duke University USA
[76]CEPH Fondation Jean Dausset 27 rue Juliette Dodu 75010 Paris, France
[77]CEA – IG – Centre National de Génotypage 2 rue Gaston Crémieux 91057 Evry Cédex, France
[78]Cole Eye Institute, Cleveland Clinic, 9500 Euclid Avenue, Cleveland OH 44195
[79]Department of Ophthalmology, Duke University Medical Center, Durham NC
[80]Department of Medicine, Duke University Medical Center, Durham NC
[81]Duke Molecular Physiology Institute, Duke University Medical Center, Durham NC
[82]Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, 3880 Tamiami Trail North, Naples, FL 34103
[83]Department of Ophthalmology, Royal Perth Hospital, Perth, Western Australia 6001, Australia
[84]Department of Ophthalmology, NYU School of Medicine, New York, NY
[85]Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, CB2 0QQ, UK
[86]Department of Ophthalmology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, CB2 0QQ, UK
[87]Department of Ophthalmology UCSF Medical School, USA
[88]Department of Ophthalmology and Visual Sciences, Vanderbilt University, Nashville, Tennessee, United States
[89]Casey Eye Institute, Oregon Health & Science University, Portland OR 97239, USA

[90]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA
[91]School of Clinical Sciences University of Edinburgh, EH16 4SB, Scotland, UK
[92]Center for Inherited Disease Research (CIDR) Institute of Genetic Medicine Johns Hopkins University School of Medicine Baltimore, MD
[93]Department of Ophthalmology, David Geffen School of Medicine—UCLA, Stein Eye Institute, Los Angeles, CA 90095, USA
[94]Department of Human Genetics, David Geffen School of Medicine—UCLA, Los Angeles, CA 90095, USA
[95]Department of Human Genetics, Radboud University Medical Centre, Nijmegen, the Netherlands
[96]Department of Pathology & Cell Biology, Columbia University, New York, NY
[97]Center for Translational Medicine, Moran Eye Center, University of Utah School of Medicine, Salt Lake City UT
[98]Division of Preventive Medicine, Brigham & Women's Hospital, Harvard Medical School
[99]Division of Epidemiology and Clinical Applications, Clinical Trials Branch, National Eye Institute, National Institutes of Health, Bethesda, MD 20892, USA
[100]Institute for Computational Biology, Case Western Reserve University School of Medicine, 2103 Cornell Rd Cleveland, OH 44106

[†]First authors, [‡]Senior authors
*Correspondence to: iris.heid@klinik.uni-regensburg.de (I.M.H.), goncalo@umich.edu (G.R.A.), ski@case.edu (S.K.I.),

ABSTRACT

Advanced age-related macular degeneration (AMD) is the leading cause of blindness in the elderly with limited therapeutic options. Here, we report on a study of >12 million variants including 163,714 directly genotyped, most rare, protein-altering variant. Analyzing 16,144 patients and 17,832 controls, we identify 52 independently associated common and rare variants ($P < 5x10^{-8}$) distributed across 34 loci. While wet and dry AMD subtypes exhibit predominantly shared genetics, we identify the first signal specific to wet AMD, near *MMP9* (difference-$P = 4.1x10^{-10}$). Very rare coding variants (frequency < 0.1%) in *CFH*, *CFI*, and *TIMP3* suggest causal roles for these genes, as does a splice variant in *SLC16A8*. Our results support the hypothesis that rare coding variants can pinpoint causal genes within known genetic loci and illustrate that applying the approach systematically to detect new loci requires extremely large sample sizes.

Advanced age-related macular degeneration (AMD) is an ocular neurodegenerative disease and the leading cause of vision loss among the elderly with prevalence estimated at 5% for those above 75 years of age[1,2]. The disease is characterized by reduced function of the retinal pigment epithelium (RPE) and loss of photoreceptors in the macula. Advanced AMD is classified as wet (choroidal neovascularization, CNV, when accompanied by angiogenesis) or dry AMD (geographic atrophy, GA, when angiogenesis is absent). These advanced stages of disease are typically preceded by a gradual accumulation of acellular debris in the form of drusen and by pigmentary abnormalities in the macula[3]. Advanced AMD is estimated to affect 9.6 million patients currently worldwide and early AMD stages more than 154.6 million[4]. At present, our understanding of disease biology and therapies remains limited[5].

Genetic variants, whether associated with small or large changes in disease risk, can help uncover disease mechanisms and provide entry points into therapy. Analysis of common variation have uncovered numerous risk loci for a multitude of complex diseases (see Web Resources) including 21 loci for AMD[6-12]. However, for most disease loci, translation into biological insights remains a major challenge, since the functional consequences of associated common variants are typically subtle[13] and therefore open to inconsistent interpretations.

With advances in sequencing technology, it is expected that genetic analyses will gradually extend to rare variation, which often has more obvious functional consequences[14,15] and thus can accelerate translation of genetic findings into biological understanding[14,16]. For example, identifying multiple disease-associated coding variants in the same gene would provide strong evidence that disrupting gene function leads to disease[17] particularly when these are naturally occurring knock-out alleles. Studies that implicate specific rare variants in complex diseases are few and limited in their generalizability, as they either rely on special populations[8,18,19], on targeted examinations of a few genes[7,9-11,20,21], or on genome-wide assessments of relatively modest numbers of individuals[22-25]. In contrast, systematic analyses of common variation are now available in hundreds of thousands of phenotyped individuals[26,27]. Thus, there remains considerable uncertainty about the relative role of rare variants in complex disease and the best strategies to identify highly informative rare variants. Importantly, the optimal sample sizes and study designs for such studies remain poorly understood[16].

Here, we set out to systematically examine common and rare variation of AMD in the International AMD Genomics Consortium (IAMDGC) incorporating both a genome-wide approach as well as enrichment from a targeted approach. The preceding largest study of AMD examined ~2.4 million variants including ~18,000 imputed or genotyped protein-altering variants using meta-analysis[6]. Customizing a chip for *de novo* centralized genotyping, we analyze >12 million variants including 163,714 directly typed protein-altering variants in

43,566 unrelated subjects of predominantly European ancestry. Our study thus constitutes a detailed simultaneous assessment of common and rare variation in a complex disease and a large sample, setting expectations for other well-powered studies combining common and rare variant information.

RESULTS

The study data and genomic heritability

We gathered advanced AMD cases with GA and/or CNV, intermediate AMD cases, and control subjects across 26 studies (**Supplementary Table 1**). While recruitment and ascertainment strategies varied (**Supplementary Table 2**), DNA samples were collected and genotyped centrally. Making maximal use of genotyping technologies, we utilized a chip with (i) the usual genome-wide variant content, (ii) exome content comparable to the exome chip (adding protein-altering variants from across all exons), and a specific customization to add (iii) protein-altering variants detected by our prior sequencing of known AMD loci (see **Methods**) and (iv) previously observed and predicted variation in *TIMP3* and *ABCA4*, two genes implicated in monogenic retinal dystrophies. After quality control, we retained 439,350 directly typed variants including a grid of 264,655 common variants (frequency among controls >1%) distributed across autosomes, sex chromosomes, and mitochondria, primarily (93%) non-coding, and 163,714 directly genotyped protein-altering variants (including 8,290 from known AMD loci), mostly rare (88% with frequency among controls ≤1%). Imputation to the 1000 Genomes reference panel enabled examining a total of 12,023,830 variants (**Supplementary Table 3A**). Our final data set included a total of 43,566 subjects consisting of 16,144 advanced AMD patients and 17,832 control subjects of European ancestry for our primary analysis, as well as 6,657 Europeans with intermediate disease and 2,933 subjects with Non-European ancestry (**Supplementary Table 3B, Supplementary Figure 1**).

Altogether, our genotyped markers accounted for 46.7%[28] of variability in advanced AMD risk in the European ancestry subjects (95% confidence interval [CI] 44.5% to 48.8%). Regarding AMD subtypes, estimates for CNV (h^2 = 44.3%, CI 42.2% to 46.5%) and GA (h^2 = 52.3%, CI 47.2% to 57.4%) were similar; a bivariate analysis[29] showed a high genetic correlation of 0.85 (CI 0.78 to 0.92) between disease subtypes.

Thirty-Four Susceptibility Loci for AMD

We first conducted a genome-wide single variant analysis of the >12 million genotyped or imputed variants (applying genomic control correction, λ=1.13) comparing the 16,144 advanced AMD patients and 17,832 control subjects of European ancestry (full results online; see **Web resources**). We obtained >7000 genome-wide significant variants ($P \leq 5x10^{-8}$, **Supplementary Figure 2**). To identify independently associated variants, we

adopted sequential forward selection (**Supplementary Figure 3**), resulting in 52 independently associated variants that reach genome-wide significance (**Supplementary Table 4, Supplementary File 1**). These are distributed across 34 locus regions (**Figure 1A**), each extending across the identified and correlated variants, $r^2 \geq 0.5$, ±500kb (**Supplementary Table 5**). While each of these 52 variants points to a genomic element contributing to AMD biology, variants in the 34 different loci reside relatively far from each other and likely contribute to disease by regulating or modifying the function of different genes. The 34 loci include 16 loci that reached genome-wide significance for the first time (novel loci, **Table 1**) and include genes with compelling biology like extra-cellular matrix genes (*COL4A3*, *MMP19*, *MMP9)*, an ABC transporter linked to HDL cholesterol (*ABCA1)*, and a key activator in immune function (*PILRB)*. Also included are 18 of the 21 AMD loci that reached genome-wide significance previously[6,9] (known loci, **Table 1**), between-study heterogeneity was low, particularly for the new loci (**Supplementary Note 1**).

Most associated variants are common (45 out of 52) with fully conditioned odds ratios (OR) from 1.1 to 2.9 (**Figure 1B, Supplementary Table 4**) with two interacting variants (**Supplementary Note 2**). We also observed seven rare variants with frequencies between 0.01% and 1% and ORs between 1.5 and 47.6 (**Figure 1B, Supplementary Table 4**). All of these variants were also rare in Non-European ancestries (**Supplementary Table 6,** extended association results on Non-European in **Supplementary File 2**). All seven rare variants are located in/near complement genes: four non-synonymous (*CFH*:R1210C, *CFI*:G119R, *C9*:P167S, *C3*:K155Q) and previously found in targeted analyses of complement genes[7-11]; three others (CFH: rs148553336, rs191281603, rs35292876) described here for the first time, including two with the rare allele decreasing disease risk by ~2.5 to ~3.3-fold and one increasing risk 1.6 fold. To ensure validity of our results, we verified associations of lead variants in sensitivity analyses that relied on alternate association tests, adjusted for age, gender, or ten ancestry principal components, or were restricted to population-based controls or controls ≥ 50 years of age (data not shown). Altogether, our genome-wide single variant analysis nearly doubled the number of AMD loci and has identified several novel rare variants in *CFH*.

Prioritizing variants within 52 association signals

It is often challenging to translate common variant association signals into mechanistic understanding of biology; two key challenges are (i) a large number of variants with similar signals because of linkage disequilibrium and (ii) their often subtle functional consequences. Without narrowing down the lists of candidate variants, follow-up functional experiments are complicated. In our large data set, we were able to prioritize among nearby variants: we computed each variant's ability to explain the observed signal and derived, for each of the 52

signals, the smallest set of variants that included the causal variant with 95% probability [30,31]. The 52 credible sets each included from 1 to >100 variants (total of 1,345 variants, **Supplementary File 3**). For 27 out of the 52 sets, the sets were small with ≤10 variants (19 signals with ≤5 variants, **Supplementary Table 7**); seven sets included only one variant -- – demonstrating the potential for fine-mapping association signals when dense genotype data is systematically analyzed in large samples**.** Among the 205 variants with >5% probability of causing the statistical signal, we observe 11 protein-altering (all non-synonymous) variants (versus 2 expected assuming 1% protein-altering variants overall, P for enrichment = 8.7×10^{-6}, **Supplementary Table 8**). These variants provide a focused starting point for future functional analyses, although we recognize that the analysis has limitations [for example, when causal variants are not genotyped nor well-imputed, or when the signal is due to a combination of multiple variants, see **Supplementary Figure 4** for a counter example]. We also note that other variants in each locus (potentially including variants in linkage disequilibrium with lead variants and/or other variants nearby) could also contribute to disease risk.

Rare Variant Association Signals

Analysis of rare variants that potentially alter peptide sequences (non-synonymous), truncate proteins (premature stop), or affect RNA splicing (splice site) can help to identify causal mechanisms – particularly when multiple such associated variants reside in the same gene[16,32]. We examined the cumulative effect of these protein-altering variants with a frequency ≤1% in each of our ancestry groups. Genome-wide, no signal was detected with P ≤ 0.05/17,044 = 2.9×10^{-6} outside the 34 AMD loci (**Figure 1C**). Within the 34 loci, we found 14 genes with significant disease burden (P < 0.05/703 genes = 7.1×10^{-5}, **Supplementary Table 9**). To eliminate settings where a rare variant burden finding is a linkage disequilibrium shadow of a nearby stronger common variant, we evaluated each burden signal upon its independence from already identified variants in the locus (from **Supplementary Table 4**). Four of the 14 genes preserved a significant (P < 0.05/703 = 7.1×10^{-5}) rare variant burden when conditioning on already identified variants in the locus (*CFH*, *CFI*, *TIMP3*, *SLC16A8*; conditioned P = 1.2×10^{-6}, 1.0×10^{-8}, 9.0×10^{-8}, or 3.1×10^{-6}, respectively, **Table 2**). Sensitivity analyses provide similar (excluding previously sequenced subjects) and extended results (prioritizing variants with high predicted functionality, **Supplementary Note 3)**. Several interesting patterns emerge, many of which we owe to our chip design.

First, three of the four rare variant burden signals (*CFH*, *CFI*, *TIMP3*) are due to very rare variants, each with frequency <0.1%, all genotyped (**Supplementary File 4**). Many human genetic studies have used frequency thresholds of 1% to 5% as a working definition of “rare”, but our data suggests that trait associated variants with clear function may often be

much rarer – likely necessitating very large sample sizes for analysis. In two genes (*CFH, CFI*), the rare burden was detected because we enriched arrays with variants from previous sequencing of known AMD loci in cases and controls[10] (54 of 80 variants). The burden findings in *CFH* (new, **Supplementary Note 4**) and *CFI*[9] together with variants *CFH:*R1210C and *CFI*:G119R[7,9], corroborate a causal role for these genes in AMD etiology.

The third signal (*TIMP3*) was in a gene previously associated with Sorsby's fundus dystrophy, a rare disease with early onset at <45 years of age but with clinical presentation strikingly similar to AMD[33,34]. Because the majority of Sorsby's alleles disrupt cysteine-cysteine bonds in *TIMP3*, we arrayed all possible cysteine disrupting sites together with other previously described Sorsby's risk alleles [33,34]. The nine rarest *TIMP3* variants were cumulatively associated with >30-fold increased risk of disease. *TIMP3* resides in an established AMD locus[5,35] targeted in previous sequencing efforts[32,35], that were too small to demonstrate an excess of rare variation on this scale (1 variant in 17,832 controls versus 29 variants in 16,144 cases). Interestingly, although Sorsby-associated *TIMP3* variants typically occur in exon 5, four of the unpaired cysteine residues we observed map to other exons – perhaps because unpaired cysteines in different locations impair protein folding in different ways, contributing to variation in disease severity or age of onset: disease onset for our 29 cases with *TIMP3* variants was ≥50 years of age (average 64.5 years). AMD cases with these rare *TIMP3* risk alleles still exhibited much higher counts of AMD risk alleles across the genome than controls, suggesting that *TIMP3* is not a monogenic cause of AMD but contributes to disease together with alleles at the other risk loci. Our finding illustrates a locus where complex and monogenic disorders arise from variation in the same gene, similar to *MC4R* and *POMC* in obesity[36] or *UMOD* in kidney function[37]. In a similar approach, we analyzed 146 rare protein-altering variants in *ABCA4*, a gene underlying Stargardt disease[38], but found no association (P=0.97).

The rare variant burden signal in *SLC16A8* was primarily driven by a putative splice variant (c.214+1G>C, rs77968014, minor allele frequency among controls, CAF = 0.81%, OR = 1.5, imputed with R^2=0.87, **Supplementary File 4**). This is thus not a true burden from multiple rare variants, but a single variant emerging as significant due to the reduced multiple testing from gene-wide testing (single variant association P = 9.1×10^{-6}, conditioned on rs8135665 P = 1.3×10^{-6}). This variant is interesting as it is predicted to disrupt processing of the encoded transcript (as +1 G variant, Human Splicing Finder 3.0); however, functional analyses in relevant tissue would be required to substantiate the direct implication for gene function and AMD. S*LC16A8* encodes a cell membrane transporter, involved in transport of pyruvate, lactate and related compounds across cell membranes[39]. This class of proteins mediates the acidity level in the outer retinal segments, and *SLC16A8* gene knock-out animals have changes in visual function and scotopic electroretinograms, but not overt retinal

pathology [40]. Interestingly, a progressive loss of *SLC16A8* expression in eyes affected with GA was reported with increasing severity of disease[41]. In summary, our chip design and our large data set enabled us not only to detect interesting features of AMD genetics, but also to provide guidance for future investigations on rare variants.

From Disease Loci to Biological Insights

In addition to fine-mapping and searches for protein-altering rare variants, many analyses can further narrow the list of candidate genes in our identified loci. We annotated the 368 genes closest to our 52 association signals (index variant and proxies, r^2 ≥0.5, ±100kb, **Supplementary File 5**), noting among these the genes those that contained associated credible set variants (**Supplementary File 3**) or a rare variant burden (**Table 2**) – these are the highest priority candidates, consistent with previous analysis of putative cis-regulatory variants[42]. We further checked whether genes were expressed in retina (82.6% of genes) or RPE/choroid (86.4%, **Supplementary File 6**). We sought relevant eye phenotypes in genetically modified mice (observed in 32 of the 368 queried genes, **Supplementary File 7**). We tagged genes in biological pathways enriched across loci, such as the alternative complement pathway, HDL transport, and extracellular matrix organization and assembly (**Supplementary Table 10**) – highlighting genes that connect multiple pathways (*COL4A3*/*COL4A4*, *ABCA1*, *MMP9*, and *VTN*). We also highlighted genes that were approved or experimental drug targets (31 of the 368 queried, **Supplementary File 8**). Finally, we prioritized genes where at least one of the credible set variants (**Supplementary File 3**) was protein-altering or located in a putative functional region (promoter, 3'/5' UTR).

All this information is summarized in the gene priority score table (**Supplementary File 9, Supplementary Note 5**), which uses a simple customizable scoring scheme to assign priority: the scheme using equal weights for each column assigns highest scores per novel locus (**Figure 2A, Supplementary Table 11**) to genes such as master regulators of immune function (*PILRB)*, matrix metalloproteinase genes (*MMP9, MMP19)*, a gene involved in lipid transport (*ABCA1)*, a gene playing a role in lipid peroxidation and inflammation (*GPX4*), an inhibitor of the complement cascade (*VTN*), another collagen gene known to cause Alport's syndrome (*COL4A3*), a gene causing a developmental monogenic disorder, the Noonan syndrome (*PTPN11*), and a retinol dehydrogenase involved in the regeneration of cone and rod photoreceptor segments previously associated with autosomal recessive night-blindness (*RDH5)*. All of these are expressed in relevant tissues, several of these show relevant mouse phenotypes *(MMP9, MMP19, COL4A3, PTPN11, GPX4,* and *RDH5*), and six of these are current drug targets (*ABCA1*, *MMP19*, *RDH5, PTPN11, VTN, GPX4*). In the known AMD loci, the highest scores per locus included the usual suspects (*CFH*, *CFI*, *CFB*, *C3*, and *APOE*) as well as *TIMP3* and *SLC16A8* (**Figure 2B**). This summary of evidence may help prioritize

genes for follow-up functional experiments. It should be noted that much of the information was collected specifically for the genes in the identified loci (for example, by reviewing literature for animal models for each respective gene), rather than systematically annotating all genes genome-wide, so that this summary of evidence is not amenable to formal statistical enrichment analysis.

Commonalities and differences between advanced AMD subtypes

Previously identified risk variants all contribute to the two advanced AMD subtypes, CNV and GA. We compared association signals between our 10,749 cases with CNV and 3,235 cases with GA. Four of the 34 lead variants show significant difference (P_{diff} < 0.05/34 = 0.00147) between disease subtypes (in the loci *ARMS2/HTRA1*, *CETP*, *MMP9*, *SYN3/TIMP3,* **Figure 3A, Supplementary Table 12**). Variant rs42450006 upstream of *MMP9* was the only one that was specific to one subtype, being exclusively associated with CNV (frequency in controls = 14.1%; OR_{CNV} = 0.78 vs. OR_{GA} = 1.04; P_{diff} = $4.1x10^{-10}$), but not with GA (P_{GA}=0.39). The signal was markedly stronger in an analysis restricted to CNV (**Supplementary Note 6**). The *MMP9* signal for neovascular disease fits well with prior evidence: upregulation of *MMP9* appears to induce neovascularization[43]; a feedback loop between *VEGF* signaling and *MMP9* has been proposed in the RPE[44]. VEGF currently provides an effective therapy for patients with neovascular AMD, but the struggle to keep vision continues. Beyond confirming a shared genetic predisposition of the two subtypes, our data identifies – for the first time – one variant that is specific to one subtype.

Commonalities and differences between advanced AMD and earlier disease stages

We evaluated our association signals in 6,657 individuals with intermediate AMD, defined as having more than five macular drusen greater than 63µm and/or pigmentary changes in the RPE. Examining all genotyped variants[28], we found a correlation of rho = 0.78, indicating substantial overlap between genetic determinants of advanced effects and and intermediate AMD (95% CI 0.69 to 0.87). Among our 34 index variants, 24 showed nominally significant association ($P_{intermediate}$ ≤ 0.05) with intermediate AMD (2 expected, $P_{binomial}$ = $4.8x10^{-24}$); all had ORs in the same direction but smaller in magnitude (**Figure 3B**, **Supplementary Table 13**). The other 10 variants showed no association with intermediate AMD ($P_{intermediate}$ > 0.05), despite sufficient power (**Supplementary Table 14**). Interestingly, these 10 variants point to 7 extra-cellular matrix genes (*COL15A1*, *COL8A1*, *MMP9*, *PCOLCE*, *MMP19*, *CTRB1*/2, *ITGA7,* **Supplementary Table 15**), based on which one may hypothesize that the extra-cellular matrix points to a disease subtype without early stage manifestation or with extremely rapid progression. If confirmed, a group of rapidly progressing patients or without

early symptoms might eventually derive maximum benefit from genetic diagnosis and future preventive therapies.

An Accounting of AMD Genetics

To account for progress made here in understanding AMD genetics, we estimated the proportion of disease risk explained by our 52 independent variants and compared it to our initial estimates of heritability obtained by examining all genotyped variants. We computed a weighted risk score of the 52 variants[45] and modeled a realistic genetic risk score distribution (see **Materials and Methods**). Individuals in the highest decile of genetic risk have a 44-fold increased risk of developing advanced AMD compared to the lowest decile; of these, 22.7% are predicted to have AMD in an elderly general population above 75 years of age with ~5% disease prevalence (**Figure 4A, Supplementary Table 16**). Altogether, the 52 variants explain 27.2% of disease variability (**Figure 4B,** also highlighting results based on other prevalence assumptions), including a 1.4% contribution from rare variants. The 52 identified variants thus explain more than half of the genomic heritability (estimated as 46.7%, see first results chapter). The balance might be attributed to additional variation not studied here, or to genetic interaction with environmental factors such as smoking, diet or sunlight exposure, or to chance.

DISCUSSION

We set out to improve our understanding of rare and common genetic variation for macular degeneration biology, so as to guide the development of therapeutic interventions and facilitate early diagnosis, monitoring and prevention of disease. AMD is an ideal role model to study complex disease genetics: it was the focus of the first successful genome-wide study of common variants [46], and a total of 21 disease susceptibility loci with a broad range of effect sizes have been identified altogether [6-12]. Here, we systematically examine rare variation (through direct genotyping) and common variation (through genotyping and imputation) for AMD in a study designed to discover >80% of associated protein-altering variants with an allele frequency of >0.1% and >3-fold increased disease risk (or >0.5% frequency and >1.8-fold increased disease risk). Our study provides a comprehensive simultaneous assessment of common and rare variation enabling us to understand the relative roles of rare and common variants and the scientific insights to be gained from rare variation.

Rare protein-altering variants are an especially attractive target for genetic studies because most of these variants are expected to damage gene function. Furthermore, observing that many rare variants in a gene are, together, associated with a change in disease risk strongly suggests that the gene is causally implicated in disease biology and –

further – suggests the consequences of mimicking or blocking gene action using a drug. Our study demonstrates that when rare variants are systematically assessed in genome-wide assessments of large numbers of cases and controls, significant signals can be assigned to single rare variants as well as to rare variant burden in specific genes.

Our study also demonstrates the challenges of these analyses. For three of the genes where we identified a rare variant burden, the accumulated evidence was spread across very rare variants with frequencies <0.1% in controls. Most of these variants derived from our enrichment of the chip with protein-altering variants in known AMD loci based on our own sequencing including AMD patients. This emphasizes the value of a hybrid approach with direct targeted sequencing in large sample sizes including patients to detect very rare variants and genotyping these variants in even larger sample sizes for association analysis. Another conclusion is about required sample sizes: although such rare variants are expected to exist in nearly all genes, no rare variant burden was observed in most of the 34 loci we studied. For these loci, identifying causal mechanisms through the study of rare protein-altering variants will require even larger sample sizes to identify variants missed by our customized exome arrays. While our findings of rare variant burden are predominantly from targeted enrichment, the knowledge about effect sizes and frequencies of contributing variants illustrates that applying the approach genome-wide to detect new loci requires extremely large sample sizes. In our view, a recent estimate that sequencing of 25,000 cases will be needed to identify genes where rare variants have a substantial impact on disease risk is likely to be a starting point for rare variant analysis, rather than an ultimate target, particularly given the fact that effect sizes for AMD risk alleles appear to be larger than for many other complex traits [16].

In addition to corroborating previous reports of rare variants that disrupt genes in the complement pathway and lead to large increases in disease risk, our study also includes two unexpected rare variant findings. First, we show that a putative splice variant in *SLC16A8* can greatly increase the risk of age-related macular degeneration – providing strong evidence that the gene is directly involved in disease biology. *SLC16A8* is a lactate transporter expressed[39] specifically by the RPE, and a deficit of lactate transport toward the choroid vasculature results in acidification of the retina and photoreceptor dysfunction as reported for *SLC16A8* knock-out mice[40]. Second, we show a >30-fold excess of rare *TIMP3* mutations among putative cases of macular degeneration. *TIMP3* is an especially attractive candidate that has been the subject of previous, underpowered, genetic association studies.

While it has been hypothesized that studies of rare and low frequency genetic variants will greatly increase the proportion of genetic risk that can be explained, our results don't support this. Our study and others successfully identify many low frequency disease risk alleles, and these provide clues about disease biology, but our results also show that

common variants make a much larger contribution to variability in disease risk. Common variants suggest a large number of interesting leads and pathways for future analysis (**Supplementary Table 11, Figure 2A**), including attractive candidates such as immune regulators (*PILRB)*, genes implicated in mouse ocular phenotypes (*MMP9, MMP19, COL4A3, PTPN11, GPX4,* and *RDH5)*, and proven drug targets (*ABCA1*, MMP19, *RDH5, PTPN11, VTN, GPX4*). In a literature search, we identified no previous candidate gene association studies targeting our novel loci, although several model organism, cellular, and functional studies evaluated potential links between genes in these loci and AMD (highlights of this search in **Supplementary Table 11**) and a few loci were nominally associated and proposed as candidates in prior genome-wide searches [47,48]. As richer functional annotations of the genome[49] become available in diverse cell types, systematic assessment of overlap between these and our loci should clarify disease biology.

Our study also suggests additional important observations. While our results show that the majority of genetic risk is shared between GA and CNV, we also identify – for the first time – a variant that is specific to one advanced AMD subtype: a genetic variant near *MMP9* is specific to CNV, a candidate gene also supported by prior gene expression analyses in the Bruch's membrane of patients with neovascular disease[50]. Future efforts extending to longitudinal data might help improve the dissection of pure CNV and pure GA and their genetic make-up even further, but longitudinal data has still to be extended to yield sufficient sample sizes. If substantiated, the fact that nearly all disease associated variants modulate risk of both CNV and GA has potentially significant therapeutic consequences. It implies that individuals at high risk of CNV are also at high risk of GA. This suggests that therapeutic strategies which mitigate CNV but not GA will only provide temporary relief to patients – who are likely to remain at high risk of developing GA and may still require future interventions to prevent it.

Therefore, our findings have several important implications for future studies of rare variation in human complex traits. First, they clearly emphasize the need for very large sample sizes in population studies: the functionally most interesting variants we identify have frequencies in the range of 0.01 – 1.0% and, despite their strong impact on disease risk, could only be implicated using 10,000s of individuals. Second, they illustrate the value of hybrid approaches, where sequencing is used to detect interesting variants and custom arrays and imputation are used to examine these variants in very large samples. Since all the large effect rare variants we identify reside in or near GWAS loci, as with most complex trait associated rare variants [7-11,20,21,23,51], focused studies around GWAS loci may continue to be a cost-effective compromise. Third, our analysis of cysteine variants in *TIMP3* illustrates not only the potential for targeted variant discovery but the critical need to understand the consequences of rare variants when analyzing them together. While very large samples will

be needed, our results also show that the effort to extend genetic studies to rare variants is worthwhile as these variants can pinpoint causal genes and advance our understanding of disease biology.

References

1. Smith, W. *et al.* Risk factors for age-related macular degeneration: Pooled findings from three continents. *Ophthalmology* **108**, 697-704 (2001).
2. Chakravarthy, U., Evans, J. & Rosenfeld, P.J. Age related macular degeneration. *BMJ* **340**, c981 (2010).
3. Ferris, F.L. *et al.* A simplified severity scale for age-related macular degeneration: AREDS Report No. 18. *Arch Ophthalmol* **123**, 1570-4 (2005).
4. Wong, W.L. *et al.* Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *Lancet Glob Health* **2**, e106-16 (2014).
5. Fritsche, L.G. *et al.* Age-related macular degeneration: genetics and biology coming together. *Annu Rev Genomics Hum Genet* **15**, 151-71 (2014).
6. Fritsche, L.G. *et al.* Seven new loci associated with age-related macular degeneration. *Nat Genet* **45**, 433-9, 439e1-2 (2013).
7. Raychaudhuri, S. *et al.* A rare penetrant mutation in CFH confers high risk of age-related macular degeneration. *Nat Genet* **43**, 1232-6 (2011).
8. Helgason, H. *et al.* A rare nonsynonymous sequence variant in C3 is associated with high risk of age-related macular degeneration. *Nat Genet* **45**, 1371-4 (2013).
9. Seddon, J.M. *et al.* Rare variants in CFI, C3 and C9 are associated with high risk of advanced age-related macular degeneration. *Nat Genet* **45**, 1366-70 (2013).
10. Zhan, X. *et al.* Identification of a rare coding variant in complement 3 associated with age-related macular degeneration. *Nat Genet* **45**, 1375-9 (2013).
11. van de Ven, J.P. *et al.* A functional variant in the CFI gene confers a high risk of age-related macular degeneration. *Nat Genet* **45**, 813-7 (2013).
12. Arakawa, S. *et al.* Genome-wide association study identifies two susceptibility loci for exudative age-related macular degeneration in the Japanese population. *Nat Genet* **43**, 1001-4 (2011).
13. Gibson, G. Rare and common variants: twenty arguments. *Nat Rev Genet* **13**, 135-45 (2011).
14. Do, R., Kathiresan, S. & Abecasis, G.R. Exome sequencing and complex disease: practical aspects of rare variant association studies. *Hum Mol Genet* **21**, R1-9 (2012).
15. Nelson, M.R. *et al.* An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* **337**, 100-4 (2012).
16. Zuk, O. *et al.* Searching for missing heritability: designing rare variant association studies. *Proc Natl Acad Sci U S A* **111**, E455-64 (2014).
17. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546-58 (2013).
18. Styrkarsdottir, U. *et al.* Severe osteoarthritis of the hand associates with common variants within the ALDH1A2 gene and with rare variants at 1p31. *Nat Genet* **46**, 498-502 (2014).
19. Styrkarsdottir, U. *et al.* Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. *Nature* **497**, 517-20 (2013).
20. Rivas, M.A. *et al.* Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet* **43**, 1066-73 (2011).
21. Flannick, J. *et al.* Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat Genet* **46**, 357-63 (2014).
22. Cruchaga, C. *et al.* Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature* **505**, 550-4 (2014).
23. Do, R. *et al.* Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature* **518**, 102-6 (2015).
24. Lange, L.A. *et al.* Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. *Am J Hum Genet* **94**, 233-45 (2014).
25. Walters, R.G. *et al.* A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* **463**, 671-5 (2010).
26. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).
27. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187-96 (2015).
28. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
29. Lee, S.H., Yang, J., Goddard, M.E., Visscher, P.M. & Wray, N.R. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* **28**, 2540-2 (2012).

30. Wellcome Trust Case Control, C. *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat Genet* **44**, 1294-301 (2012).
31. Wen, X. Bayesian model selection in complex linear systems, as illustrated in genetic association studies. *Biometrics* **70**, 73-83 (2014).
32. Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J.A. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* **324**, 387-9 (2009).
33. Sorsby, A. & Mason, M.E. A fundus dystrophy with unusual features. *Br J Ophthalmol* **33**, 67-97 (1949).
34. Weber, B.H., Vogt, G., Wolz, W., Ives, E.J. & Ewing, C.C. Sorsby's fundus dystrophy is genetically linked to chromosome 22q13-qter. *Nat Genet* **7**, 158-61 (1994).
35. Abecasis, G.R. *et al.* Age-related macular degeneration: a high-resolution genome scan for susceptibility loci in a population enriched for late-stage disease. *Am J Hum Genet* **74**, 482-94 (2004).
36. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* **42**, 937-48 (2010).
37. Kottgen, A. *et al.* New loci associated with kidney function and chronic kidney disease. *Nat Genet* **42**, 376-84 (2010).
38. Allikmets, R. *et al.* Mutation of the Stargardt disease gene (ABCR) in age-related macular degeneration. *Science* **277**, 1805-7 (1997).
39. Halestrap, A.P. The SLC16 gene family - structure, role and regulation in health and disease. *Mol Aspects Med* **34**, 337-49 (2013).
40. Daniele, L.L., Sauer, B., Gallagher, S.M., Pugh, E.N., Jr. & Philp, N.J. Altered visual function in monocarboxylate transporter 3 (Slc16a8) knockout mice. *Am J Physiol Cell Physiol* **295**, C451-7 (2008).
41. Shoshan, V., MacLennan, D.H. & Wood, D.S. A proton gradient controls a calcium-release channel in sarcoplasmic reticulum. *Proc Natl Acad Sci U S A* **78**, 4828-32 (1981).
42. Stranger, B.E. *et al.* Patterns of cis regulatory variation in diverse human populations. *PLoS Genet* **8**, e1002639 (2012).
43. Lambert, C. *et al.* Gene expression pattern of cells from inflamed and normal areas of osteoarthritis synovial membrane. *Arthritis Rheumatol* **66**, 960-8 (2014).
44. Hollborn, M. *et al.* Positive feedback regulation between MMP-9 and VEGF in human RPE cells. *Invest Ophthalmol Vis Sci* **48**, 4360-7 (2007).
45. Rudnicka, A.R. *et al.* Age and gender variations in age-related macular degeneration prevalence in populations of European ancestry: a meta-analysis. *Ophthalmology* **119**, 571-80 (2012).
46. Klein, R.J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385-9 (2005).
47. Chen, W. *et al.* Genetic variants near TIMP3 and high-density lipoprotein-associated loci influence susceptibility to age-related macular degeneration. *Proc Natl Acad Sci U S A* **107**, 7401-6 (2010).
48. Logue, M.W. *et al.* A search for age-related macular degeneration risk variants in Alzheimer disease genes and pathways. *Neurobiol Aging* **35**, 1510.e7-18 (2014).
49. The Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
50. Hussain, A.A., Lee, Y., Zhang, J.J. & Marshall, J. Disturbed matrix metalloproteinase activity of Bruch's membrane in age-related macular degeneration. *Invest Ophthalmol Vis Sci* **52**, 4459-66 (2011).
51. Johansen, C.T. *et al.* Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* **42**, 684-7 (2010).
52. Price, A.L. *et al.* Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet* **86**, 832-8 (2010).

Acknowledgments:

Data permitted for sharing by respective Institutional Review Boards, and/summary statistics reported in the paper will be archived in the database of Genotypes and Phenotypes (dbGaP; http://www.ncbi.nlm.nih.gov/gap).

The UWA, LEI & Flinders group acknowledges financial support for participant recruitment and sample processing provided by the National Health and Medical Research Council

(NHMRC) of Australia (#1023911), the Ophthalmic Research Institute of Australia, the BrightFocus Foundation and a Ramaciotti Establishment Grant. CERA receives Operational Infrastructure Support from the Victorian Government. KPB, JEC and AWH are supported by NHMRC Fellowships. The authors acknowledge the support of B. Usher-Ridge, L. Palmer, L Ma and DL Lim in patient recruitment and data collection.

The Pittsburgh group acknowledges funding to MBG from NIH/NEI R01 EY09859, Research to Prevent Blindness (N.Y, N.Y.), Harold and Pauline Price Foundation.

The BDES (Beaver Dam Eye Study) was supported by grant EY06594 (to RK and BEKK) from the National Institutes of Health, as well as Senior Scientific Investigator Awards (to RK and BEKK) and an unrestricted grant (to the University of Wisconsin Department of Ophthalmology and Visual Sciences) from Research to Prevent Blindness.

The Cambridge group was supported by the Medical Research Council, UK (grant G0000067 to JRWY, ATM), the Macular Disease Society (JRWY, ATM); the Guide Dogs for the Blind Association (ATM, JRWY) and the Department of Health's NIHR Biomedical Research Centre for Ophthalmology at Moorfields Eye Hospital and UCL Institute of Ophthalmology. We thank the clinicians who helped with recruitment, the Reading Centre at Moorfields Eye Hospital, London for grading fundus photographs and the subjects who participated in the research.

The EUGENDA-Cologne group was supported by a grant from the Retinovit foundation.

The Vanderbilt group was supported by the National Institutes of Health Grants AG019085 (JLH), EY023164 (JLH), EY022310 (JLH), EY012118 (JLH, AA, MAB), AG044089 (JDH), T32 EY007157 (JNCB) and a PhRMA Informatics fellowship (JNCB).

The MMAP-Penn group was supported by the Arnold and Mabel Beckman Initiative for Macular Research (CAC), Research to Prevent Blindness Inc (CAC), EyeSight Foundation of Alabama (CAC), and NIH EY023164 (DS).

The Oregon group was supported by National Eye Institute grants EY021532, and EY0105712, and an unrestricted departmental grant from Research to Prevent Blindness.

The Edinburgh group using the Scottish AMD study was funded by the Chief scientists Office (Scotland) CSO reference number: CZB/4/79 and would like to thank Alan Wright, Ana Armbrecht and Fraser Imrie for collecting the samples and all of the individuals who participated in this study.

The EU/JHU study acknowledges the support of the CEPH Biological Resource Centre by the French Ministère de l'Enseignement Supérieur et de la Recherche, Foundation Fighting Blindness Clinical Research Institute (FFB, CRI), an unrestricted grant to the Wilmer Eye Institute from Research to Prevent Blindness, and Baylor-Johns Hopkins Center for Mendelian Genetics (National Human Genome Research Institute, NHGRI/NIH; 1U54HG006542-01).

The Jerusalem study was supported by grants from the Israel Science fund (ISF) and the Israeli Ministry of Health.

The Southampton study acknowledges Southampton Wellcome Trust Clinical Research Facility for research nurse support in collecting DNA samples, Helen Griffiths (Clinical and Experimental Sciences, University of Southampton) for technical support in processing DNA and all the patients who contributed to this work. AJL supported at the University of Southampton by funding from The Wellcome Trust (076169/A), American Health Assistance Foundation (M2007110), Macula Vision Research Foundation, TFC Frost Charitable Trust, Brian Mercer Charitable Trust, Macular Society, Hobart Trust and the Gift of Sight appeal.

The Marshfield group was supported by grants NIH NCATS: UL1TR000427, NIH NHGRI: 1U01HG006389, and support from the Marshfield Clinic Research Foundation.

The Melbourne study was supported by the National Health and Medical Research Council Australia, project grant 1008979, Centre for Clinical Research Excellence #529923 - Translational Clinical Research in Major Eye Diseases. NHMRC Research Fellowship (PNB, #1028444). CERA receives Operational Infrastructure Support from the Victorian Government.

The Miami group was supported by National Institutes of Health Grants R01 EY012118 (MAP-V, WKS, JLK, SGS, MDC), EY023164 (MAP-V, WKS, JLH, MAP-V), EY022310 (MAP-V) and T32 EY023194 (RJS) and P30-EY005722. All Bascom Palmer Eye Institute authors are partially supported by NIH Center Core Grant P30EY014801 and an unrestricted grant from Research to Prevent Blindness, New York, NY, USA.

The MMAP-Michigan and AREDS groups were supported by Intramural Research Program of the National Eye Institute (ZO1 EY000475); the AREDS study was supported by the National Eye Institute/National Institutes of Health, (contract no.: HHS-NOI-EY-0-2127), Bethesda Maryland; the AREDS2 study was supported by the intramural program funds and contracts from the National Eye Institute/National Institutes of Health (NEI/NIH), Department of Health and Human Services, Bethesda, MD. Contract No. HHS-N-260-2005-00007-C. ADB Contract No. N01-EY-5-0007. The Michigan study was supported by the National Eye Institute (EY0022005) and the National Human Genome Research Institute (HG006513 HG007022), Foundation Fighting Blindness and National Institutes of Health/National Eye Institute Grant-EY016862.

The NHS/HPF studies were supported by EY021900, EY017362, EY13824, EY009611, CA87969, CA49449, and HL35464.

The Regensburg group was supported by BMBF-01ER1206 (to IMH), EFKS 2012_A147 (IMH), BMBF-01GP1308 (IMH), the Deutsche Forschungsgemeinschaft (grant WE 1259/19-1 and WE1259/19-2, BHFW), and the Alcon Research Institute (BHFW).

The Rotterdam-Clinic study was supported by ZoNMW project number: 170885606, MDfonds, Landelijke Stichting voor Blinden en Slechtzienden (LSBS).
The UCSD study was supported by grants from NIH (grants EY014428, EY018660, P30EY022589), 863 Program (2014AA021604), and Research to Prevent Blindness. ZS is supported by 863 Program (2014AA021604), ZY is supported by National Natural Science Foundation of China (81170883 and 81430008), KZ is supported by NIH grants (1R01EY018660-01A10) and VA Merit Award.
The EUGENDA-Neijmegen study was supported by MD Fonds, Gelderse Blindenstichting, Algemene Vereniging ter Voorkoming van Blindheid, Stichting Nederlands Oogheelkundig Onderzoek, Oogfonds.
The Utah study was supported by the ALSAM Foundation, an unrestricted grant from Research to Prevent Blindness to the Department of Ophthalmology and Visual Sciences, University of Utah, SOM, Moran Eye Center.
The Seoul National University Bundang group was supported by grants from the National Research Foundation of Korea, funded by the Ministry of Education, Science, and Technology (grant numbers; NRF-2009-0072603 and NRF-2012R1A1A2008943).
The Westmead/Sydney samples were collected in three studies that were supported by the National Health and Medical Research Council (NHMRC), Australia: Grant IDs 974159, 211069, 457349 and 512423 supported the Blue Mountains Eye Study that provided population-based controls; Grant ID 302010 supported the Cataract Surgery and Risk of Age-related Macular Degeneration study that provided clinic-based early and late AMD cases and controls; and Grant ID 571013 supported the Genes and Environment in late AMD study that provided clinic-based late AMD cases. NHMRC Senior Research Fellowship (JJW, 358702, 632909). NHMRC Senior Research Fellowship (JJW, 358702, 632909). The NHMRC had no role in the design or conduct of these studies.
The Columbia study was supported by the National Institutes of Health/NIH grants R01-EY013435 and P30-EY019007, and Research to Prevent Blindness (New York, NY).
The CWRU group was supported by VA Merit Review (NSP), Foundation Fighting Blindness (SAH), Research to Prevent Blindness (SAH); International Retinal Research Foundation (SKI).
CIDR Program contract number HHSN268201200008I.

We thank all participants of all the studies included for enabling this research by their participation to these studies. Computer resources for this project have been provided by the High Performance Computing Centers of the University of Michigan and the University of Regensburg. The Regensburg Team would like to thank Randy Rueckner for technical assistance.

Figure legends

Figure 1. **Genome-wide search reveals 34 loci and genes with rare variant burden for AMD.** (**a**) We conducted a genome-wide single variant association analysis for >12 million variants in 16,144 advanced AMD patients versus 17,832 controls. Shown is the Manhattan Plot exhibiting P-values for association highlighting novel ($P < 5x10^{-8}$ for the first time, green) and known (blue) AMD loci (see **Table 1**). (**b**) We computed independent effect size (log Odds Ratios) of each of the 52 identified variants (**Supplementary Table 4**). Shown are these effect sizes versus the frequency of the AMD risk increasing allele and a 80% power curve. (**c**) We conducted a genome-wide gene-based test for disease burden based on the protein-altering variants testing 17,044 RefSeq genes by the variable threshold test[52]. Shown is the Manhattan Plot with P-values, the red horizontal line indicating genome-wide significance ($P \leq 0.05/17,044 = 2.9x10^{-6}$) and the yellow line indicating AMD-locus-wide significance (given 703 genes in the 34 AMD loci, $P \leq 0.05/703 = 7.1x10^{-5}$). No gene outside the 34 loci is genome-wide significant; 14 genes are AMD-locus-wide significant (blue), four remain significant after locus-wide conditioning (bold letters, **Supplementary Table 9**).

Figure 2. Genes with top priority based on biological and statistical evidence combined. We queried 368 genes in the 34 narrow AMD regions (index and proxies, $r^2 \geq 0.5$, ±100kb) for biological (red; expression in retina/RPE/choroid, **Supplementary File 6**; ocular mouse phenotype, **Supplementary File 7**), statistical, (blue; ≥1 credible set variant in gene ±50 kb, **Supplementary File 3**; rare variant burden, **Table 2**), putative functional (green; ≥ 1 credible set variant in gene ±50 kb being protein-altering, 5'/3' UTR, other exonic, or putative promoter, **Supplementary File 3**), and molecular (magenta; enriched molecular pathway, drug target) evidence. We here focus on the gene(s) with the highest gene priority score (GPS) per locus (full list of genes in **Supplementary File 9**). Shown are **(a)** the 16 genes with highest GPS in the 15 novel AMD loci (one novel locus without any gene), and (**b)** the 25 genes with highest GPS in the 18 known AMD loci. Colored fields indicate yes and GPS counts number of colored fields per row.

Figure 3. Comparison of advanced AMD subtypes and intermediate versus advanced AMD. We compared associations of the 34 lead variants across different AMD phenotypes. Shown are effect sizes (log Odds Ratio) per minor allele in controls as well as 95% confidence intervals (widths and heights of diamonds). **(a**) Comparison of neovascular disease (10,749 CNV cases vs. 17,832 controls) and GA (3,235 GA cases vs. 17,832 controls) identified four variants (in loci *MMP9*, *ARMS2*/*HTRA1*, *CETP*, and *SYN3*/*TIMP3*) with significantly different association comparing CNV with GA ($P_{diff} < 0.05/34$, marked in red,

see also **Supplementary Table 12**). (**b)** Comparison of intermediate AMD (6,657 cases vs. 17,832 controls) with advanced AMD (16,144 cases vs. 17,832 controls) identifies 24 variants with nominally significant ($P < 0.05$, marked in red) association with intermediate AMD ($P_{binomial} = 4.8 \times 10^{-24}$), all of which have the same effect direction and less extreme effect sizes compared to advanced AMD (**Supplementary Table 13**).

Figure 4. Variance explained and absolute risk of disease based on the 52 identified variants. (**a**) Absolute disease risk (=proportion of affected) by genetic risk score intervals (deciles and top 10 percentiles in embedded bar plot) based on our cases-control-data weighted to model a general population with 5% disease prevalence (see also **Supplementary Table 16**). (**b**) Shown is disease liability explained by the 52 identified variants (bars) compared to the genomic heritability based on all genotyped variants (red lines) assuming disease prevalence of 1%, 5%, or 10%, respectively.

Table 1. Thirty-four loci for age-related macular degeneration. Our genome-wide single-variant association analysis identified 34 loci for advanced AMD with genome-wide significance ($P < 5 \times 10^{-8}$) based on logistic regression in 16,144 cases and 17,832 controls of European ancestry. Shown are P-values and effect sizes (Odds Ratios, OR) for the variant with the smallest P-value per locus (lead variant) and the number of independent signals per locus (see **Supplementary Table 4**)

Lead Variant	Chr	Position[a]	Major/ minor allele	Locus name[b]	# Sig-nals[c]	MAF		Association	
						Cases	Controls	OR	*P*
KNOWN (previously reported with genome-wide significance, $P < 5 \times 10^{-8}$)									
rs10922109	1	196,704,632	C/A	*CFH*	8	0.223	0.426	0.38	9.6×10^{-618}
rs62247658	3	64,715,155	T/C	*ADAMTS9-AS2*	1	0.466	0.433	1.14	1.8×10^{-14}
rs140647181	3	99,180,668	T/C	*COL8A1*	2	0.023	0.016	1.59	1.4×10^{-11}
rs10033900	4	110,659,067	C/T	*CFI*	2	0.511	0.477	1.15	5.4×10^{-17}
rs62358361	5	39,327,888	G/T	*C9*	1	0.016	0.009	1.80	1.3×10^{-14}
rs116503776	6	31,930,462	G/A	*C2/CFB/SKIV2L*	4	0.090	0.148	0.57	1.2×10^{-103}
rs943080	6	43,826,627	T/C	*VEGFA*	1	0.465	0.497	0.88	1.1×10^{-14}
rs79037040	8	23,082,971	T/G	*TNFRSF10A*	1	0.451	0.479	0.90	4.5×10^{-11}
rs1626340	9	101,923,372	G/A	*TGFBR1*	1	0.189	0.209	0.88	3.8×10^{-10}
rs3750846	10	124,215,565	T/C	*ARMS2/HTRA1*	1	0.436	0.208	2.81	6.5×10^{-735}
rs9564692	13	31,821,240	C/T	*B3GALTL*	1	0.277	0.299	0.89	3.3×10^{-10}
rs61985136	14	68,769,199	T/C	*RAD51B*	2	0.360	0.384	0.90	1.6×10^{-10}
rs2043085	15	58,680,954	T/C	*LIPC*	2	0.350	0.381	0.87	4.3×10^{-15}
rs5817082	16	56,997,349	C/CA	*CETP*	2	0.232	0.264	0.84	3.6×10^{-19}
rs2230199	19	6,718,387	C/G	*C3*	3	0.266	0.208	1.43	3.8×10^{-69}
rs429358	19	45,411,941	T/C	*APOE*	2	0.099	0.135	0.70	2.4×10^{-42}
rs5754227	22	33,105,817	T/C	*SYN3/TIMP3*	1	0.109	0.137	0.77	1.1×10^{-24}
rs8135665	22	38,476,276	C/T	*SLC16A8*	1	0.217	0.195	1.14	5.5×10^{-11}
NOVEL (reported with genome-wide significance, $P < 5 \times 10^{-8}$, for the first time)									
rs11884770	2	228,086,920	C/T	*COL4A3*	1	0.258	0.278	0.90	2.9×10^{-8}
rs114092250	5	35,494,448	G/A	*PRLR/SPEF2*	1	0.016	0.022	0.70	2.1×10^{-8}
rs7803454	7	99,991,548	C/T	*PILRB/PILRA*	1	0.209	0.190	1.13	4.8×10^{-9}
rs1142	7	104,756,326	C/T	*KMT2E/SRPK2*	1	0.370	0.346	1.11	1.4×10^{-9}
rs71507014	9	73,438,605	GC/G	*TRPM3*	1	0.427	0.405	1.10	3.0×10^{-8}
rs10781182	9	76,617,720	G/T	*MIR6130/RORB*	1	0.328	0.306	1.11	2.6×10^{-9}
rs2740488	9	107,661,742	A/C	*ABCA1*	1	0.255	0.275	0.90	1.2×10^{-8}
rs12357257	10	24,999,593	G/A	*ARHGAP21*	1	0.243	0.223	1.11	4.4×10^{-8}
rs3138141	12	56,115,778	C/A	*RDH5/CD63*	1	0.222	0.207	1.16	4.3×10^{-9}
rs61941274	12	112,132,610	G/A	*ACAD10*	1	0.024	0.018	1.51	1.1×10^{-9}
rs72802342	16	75,234,872	C/A	*CTRB2/CTRB1*	1	0.067	0.080	0.79	5.0×10^{-12}
rs11080055	17	26,649,724	C/A	*TMEM97/VTN*	1	0.463	0.486	0.91	1.0×10^{-8}
rs6565597	17	79,526,821	C/T	*NPLOC4/TSPAN10*	1	0.400	0.381	1.13	1.5×10^{-11}
rs67538026	19	1,031,438	C/T	*CNN2*	1	0.460	0.498	0.90	2.6×10^{-8}
rs142450006	20	44,614,991	TTTTC/T	*MMP9*	1	0.124	0.141	0.85	2.4×10^{-10}
rs201459901	20	56,653,724	T/TA	*C20orf85*	1	0.054	0.070	0.76	3.1×10^{-16}

Chr = Chromosome; MAF = minor allele frequency; OR = Odds Ratio [a] Chromosomal position is given based on NCBI RefSeq hg19; [b] The locus name is a label of the region using the nearest gene(s), but does not necessarily state the responsible gene; [c] number of independent variants in this locus; hg19 = human genome reference assembly (version 19)

Table 2. Four genes with a significant rare variant burden within the 34 AMD loci independent from other identified variants. We computed a gene-based burden test of rare protein-altering variants comparing 16,144 advanced AMD cases and 17,832 controls. Shown are P-values from the variable threshold test (up to 100 million permutations) and Odds Ratios from the collapsed burden test, both adjusted for the other identified variants in the respective locus (locus-wide conditioning). Four genes (among the 703 genes in the 34 AMD locus regions) showed a significant ($P < 0.05/703 = 7.1 \times 10^{-5}$) burden. Details about the corresponding rare variants underlying the observed burden can be found in **Supplementary File 4.** Results for the 14 genes that show significant burden within the 34 AMD loci without locus-wide conditioning are shown in **Supplementary Table 9**. Rare variants were defined here as variants with minor allele frequency in cases and controls < 1% in each of the ancestries, European, Asian, and African.

Gene	Optimal Threshold for Rare Variants	Number of Variants below Optimal RAC	Summed Rare Allele Count (Frequency [%])		*P*[a]	Odds Ratio
	Count (%)	Total (Exome Chip Base + Custom)	Cases N = 16,144	Controls N = 17,832		
CFH	10 (0.015%)	37 (9+28)	88 (0.273%)	38 (0.107%)	1.2×10^{-6}	2.94
CFI	46 (0.068%)	43 (17+26)	213 (0.660%)	82 (0.230%)	1.0×10^{-8}	2.95
TIMP3	14 (0.021%)	9 (1+8)	29 (0.0898%)	1 (0.00280%)	9.0×10^{-8}	31.21
SLC16A8	648 (0.954%)	9 (7+2)	487 (1.51%)	392 (1.10%)	3.1×10^{-6}	1.40

RAC = rare allele count; [a] P-values are from the variable threshold test conditioned on other identified variants in the locus (locus-wide conditioned).

ONLINE METHODS

Study data and phenotype. In the International AMD Genomics Consortium (IAMDGC), we gathered 26 studies with each including (i) advanced AMD cases with GA and/or CNV in at least one eye and age at first diagnosis ≥ 50 years, (ii) intermediate AMD cases with pigmentary changes in the RPE or more than five macular drusen greater than 63μm and age at first diagnosis ≥ 50 years, or (iii) controls without known advanced or intermediate AMD. Recruitment and ascertainment strategies varied by study (**Supplementary Tables 1 and 2**): Advanced and intermediate AMD cases were mostly recruited from ophthalmology clinics (61.6% of cases), but also in spouses and friends of cases (2.1%), from general population (18.5%), or via mixed approaches (17.8%); controls were recruited among elderly individuals at ophthalmology clinics (53.0% of controls), among spouses and friends of cases (2.6%), from general population (26.4%), or via mixed approaches (18.0%). Of all subjects, 94.5% ascertained disease status via fundus photography or fundus exam; one study (5.5% of subjects) validated interview information through the patients' ophthalmologist. Of the 26 studies, 18 studies used Fluorescein Angiography or Optic Coherence Tomography for differentiating GA from neovascular disease. Grading scales, used to ascertain intermediate AMD, differed – as usual – across studies. All groups collected data according to the Declaration of Helsinki principles. Study participants provided informed consent and protocols were reviewed and approved by local ethics committees.

DNA and chip design. We gathered DNA samples of more than 50,000 individuals. Groups with very limited amounts of available DNA contributed aliquots after whole-genome amplification (8% of subjects).

We utilized a custom-modified HumanCoreExome array by Illumina, Inc., which includes (i) tagging variants across the genome (genome chip content) and (ii) a catalogue of protein-altering variants (exome chip content). Our customization of the array included three additional tiers to enrich for variants from 22 AMD loci implicated by our previous genome-wide association analysis[6] based on 19 index variants with genome-wide significance, 3 with consistent effect direction in the replication stage and $4x10^{-7} \leq P \leq 2x10^{-6}$) by selecting (iii) tagging variants (pair-wise tagging $r^2 < 0.8$) from Phase I 1000G/HapMap[53,54] common variants (minor allele frequency, MAF, ≥ 1 % in European or East Asian individuals) using Tagger implemented in Haploview[55] within ±100kb of the 22 index variants expanded to cover all correlated variants (r^2 [EUR] > 0.5) and the complete gene (transcript ±1 kb), (iv) protein-altering variants within 500 kb of the 22 index variants as identified from public general population data bases (dbSNP[56], the NHLBI Exome Sequencing Project[57], the Phase I 1000 Genomes Project, see **Web Resources**), and (v) protein-altering variants within the 500 kb of the 22 index variants identified by re-sequencing AMD case-control study data

(targeted re-sequencing of 2,335 AMD cases and 789 controls[10,58] and whole-genome sequencing 60 AMD cases and 60 controls; G. Abecasis and A. Swaroop). The customization further included (vi) the 1,000 top independent (> 2 Mb distant) variants from the previous analysis and additional 100 top variants from each the previous CNV only and the previous GA only analysis, (vii) and 375 variants in *ABCA4*, including known variants causing Stargardt disease[59], benign variants, and those of unknown significance, as well as 10 known and 44 predicted cysteine mutations in *TIMP3*, motivated by the known variants causing Sorsby's fundus dystrophy[33,34] (also B. Weber, personal communication).

Annotation. Variant identifiers were based on NCBI dbSNP v137. Chromosomal position and functional annotation of the variant was based on the NCBI Reference Sequence Human Genome Build 19 (RefSeq hg19)[60] and SeattleSeq Annotation 138[61] (see **Web Resources**). We particularly focus on protein-altering variants including non-synonymous coding variants (missense, stop loss, in-frame insertion/deletion, frameshift, premature stop codon) and splice sites. We converted the description of splice site variants to HGVS nomenclature using Mutalyzer version 2.0.beta-33[62] (see **Web Resources**).

Genotypes. We genotyped all subjects centrally at the Center for Inherited Diseases Research (CIDR), Johns Hopkins University School of Medicine, Baltimore, MD, USA. From the 569,645 genotyped variants, our stringent quality control procedure excluded poorly genotyped variants as evidenced by genotype call rates < 98.5% (5.8%), deviations from Hardy-Weinberg equilibrium with $P < 10^{-6}$ (0.34%), variants that mapped at multiple genome locations (0.25%) or variants failing other criteria, resulting in 521,950 (91.6%) variants passing all quality criteria. After excluding monomorphic variants (15.8%), we yielded 264,655 common variants distributed across autosomes, sex chromosomes, and the mitochondria, as well as 163,714 directly genotyped protein-altering variants including 8,290 from previously implicated AMD loci (**Supplementary Table 3A**). For these variants, genotype call rates averaged 99.9% (99.1% for subjects with amplified DNA).

We phased the autosomal and X-chromosomal genotype data using SHAPEIT (200 states, 2.5 Mb windows)[63], then imputed genotypes based on the 1000 Genomes Project[64] reference panel (1000G Phase I, version 3, SHAPEIT2 Reference) using MINIMAC[65] (reference-based 2.5 Mb chunks, 500 kb buffer regions). We then merged study variants that were excluded during imputation (not found in the reference panel) back into the final data set. We excluded common variants (CAF ≥ 1%) with bad imputation quality, $R^2 < 0.3$, and adopted a more stringent exclusion criterion for rare variants (CAF < 1%), $R^2 < 0.8$, for the initial identification of lead variants. This yielded a total of 12,023,830 genotyped (439,350) or imputed (11,584,480) quality-controlled variants (**Supplementary Table 3A**).

Analyzed subjects. Using the genomic information for subject-level quality control, we excluded duplicated and related individuals (kinship coefficient $\phi \geq 0.0884$, i.e. 3rd degree relatives or closer)[66], subjects with discrepancies between reported gender and sex chromosomal information or with atypical sex chromosome configurations[67], or subjects with genotyping call rates < 98.5%; we derived ancestry based on the first two principal components using autosomal genotyped variants together with genotype information of the samples from the Human Genome Diversity Project (HGDP)[68]. Our final data set contained 43,566 successfully genotyped unrelated subjects including 16,144 advanced AMD cases and 17,832 controls of European ancestry, 6,657 intermediate AMD cases of European ancestry, and 2,933 subjects (advanced AMD or controls) of Asian or African ancestries (**Supplementary Table 3B**).

Genomic heritability and genomic correlation. Combined contribution of genotyped variants to disease was evaluated using a variance-component based heritability analysis[69]. This analysis used genotypes to build a similarity matrix, summarizing the overall genetic kinship between each pair of individuals, and then examined the correspondence between genetic and phenotypic similarity. We estimated the explained variance on all genotyped, autosomal variants using restricted maximum likelihood (REML) analysis implemented in GCTA[28] (see **Web Resources**). We jointly estimated the contributions of rare (MAF in controls < 1 %) and common (MAF in controls ≥ 1%) genotyped variants by first separately calculating their genetic relationship matrices before adding both to the model. Obtained estimates of variance explained were transformed from the observed scale to the liability scale assuming various levels of disease prevalence[69].

We estimated the genomic correlation between different disease sub-phenotypes using bivariate REML analyses implemented in GCTA and only included common (MAF in controls ≥ 1%) genotyped variants [29]. We compared 10,749 cases with CNV versus 3,325 cases with GA (excluding the 2,070 cases with mixed CNV and GA) and we compared 6,657 intermediate AMD cases with 16,144 advanced AMD cases. For both analyses, we used the control subjects as reference and avoided shared controls between traits by randomly splitting the 17,832 unrelated European control individuals into two sub-samples of 8,916 individuals.

Genome-wide single variant association analysis. Single-variant association tests analyzing the 16,144 advanced AMD cases and 17,832 controls of European ancestry were based on the Firth bias-corrected likelihood ratio test[70], which is recommended for genetic association studies that include rare variants[71], as implemented in EPACTS (see **Web**

Resources). Analyses were adjusted for two principal components and source of DNA (whole-blood or whole-genome amplified DNA). Allele dosages of the imputed data were utilized, Sensitivity analyses were conducted to evaluate the influence of alternative association tests, alternative covariate adjustment including age or sex, or up to 10 principal components instead of two, as well as the influence of restricting to population-based controls, or to controls aged 50 years or older. Genomic control correction[72] was used to account for potential population stratification using all genotyped variants with minor allele count ≥ 20 outside of 20 previously described AMD loci[6,9]. As usual for genome-wide association studies, we considered P-values ≤ 5×10^{-8} as genome-wide significant.

To identify independently associated variants, we adopted a sequential forward selection approach: We first computed single variant association for each of the > 12 million variants. Then we selected the variant with the smallest P-value and its flanking ±5 Mb region, repeating the process until no genome-wide significant variant ($P \leq 5 \times 10^{-8}$) was left yielding a number of 10 Mb regions. Within each of these large regions, we re-analyzed each variant conditioning on the top variant, and repeated this process by adding the previously identified genome-wide significant variant(s) within the respective 10 Mb region. This yielded one or more independently associated genome-wide significant variant(s) per 10 Mb region.

A locus region was defined by a genome-wide significant variant and its correlated variants ($r^2 \geq 0.5$) ± 500kb; overlapping locus regions were merged to one locus, so some loci contained more than one index variant (details in **Supplementary Figure 3**).

In order to derive independent effect sizes (log odds ratios) for all identified variants, we computed a fully conditioned logistic regression model including all identified variants.

Bayesian approach to prioritize variants. In order to summarize the statistical evidence of a variant for its association strength, we computed the Bayes factor for each variant, which is a measure of the strength of the association that is comparable irrespective of variant frequency or study sample size. It provides the probability of the genotype configuration at a variant (in cases and controls) under the alternative hypothesis (association) divided by the probability of the genotype configuration under the null hypothesis (no association). It is computed using the association results per variant [73]. The posterior probability of each variant is then computed as the Bayes factor relative to the sum of all variants' Bayes factors across one locus region and can be thought of as the relative strength of evidence in favor of each SNP studied in the respective region. This assumes that there is one causal variant per region and that the causal variant is in the analyzed data set.

Expanding to loci with multiple association signals and thus a single alleged causal variant per signal, we used the association results per SNP obtained by conditioning on the other independent variants at that locus for computing the Bayes factor.

We derived 95% credible sets of variants per signal, which is the minimal set of variants, for which the sum of the posterior probabilities accumulates beyond 95%. This approach was recommended for fine-mapping of association signals and for prioritizing variants[74]. Assuming that there is only one causal variant in an association signal and that the causal variant is contained among the analyzed variants, such a credible set of variants contains the causal variant with 95% probability.

We annotated functionality of the variants in each of the 95% credible sets (see above).

Gene-based burden analysis. Single variant analyses have limited power to depict rare variants with association. Gene-based burden tests evaluating accumulated association from multiple rare variants per gene have been shown to complement such analyses and improve power to detect a burden of disease. We computed the burden of disease using the variable threshold test[52] as implemented in EPACTS. These analysis assume that all variants in a gene either increase or decrease disease risk. When variants with opposite directions of effect reside in the same gene, power will be reduced. An analysis with SKAT and SKAT-O, which both allow for variants with opposite directions of effect to reside in the same gene, did not identify additional signals (data not shown).

We focused this analysis on protein-altering variants, since we assumed that the other (not protein-altering) variants would outnumber these predicted deleterious variants by far and would thus dilute a disease burden from the deleterious variants. Assuming a negative selection against such deleterious variants that cause their frequency to be low across ancestries, we restricted our rare variant definition to variants with MAF < 1% (cases and controls combined) in each of our ancestry groups (African, Asian, and European). We utilized the genotypes of these rare protein-altering variants if genotyped directly, or rounded imputed allele dosages to the next best genotype if imputed; imputed variants were restricted to those of highest imputation quality (RSQ >= 0.8).

We assessed statistical significance by adaptive permutation testing with variable thresholds (up to 100 million permutations; minimal P-value = 1×10^{-8})[52]. When rare variants appear on a haplotype associated with disease through a common variant allele already identified for AMD, the rare variant burden would depict a mere shadow of the already identified variant. Therefore, we repeated the variable threshold test conditioned on the variant(s) identified in the respective locus by single variant analysis (locus-wide conditioning), to unravel a gene-based burden of rare variants independent of risk variants identified in single variants tests.

First, we searched for rare variant disease burden genome-wide applying a genome-wide Bonferroni-corrected significance threshold of $0.05 / 17{,}044 = 2.9 \times 10^{-6}$ (17,044 genes

genome-wide with at least 1 variant included in the analysis, i.e. with ≥ 1 rare protein-altering variant). In a second view on this, we focused on our 34 identified AMD loci and here applied a significance threshold based on the 703 genes overlapping with the locus regions ($P < 0.05 / 703 = 7.1 \times 10^{-5}$). Odds ratio estimates of the burden were derived by logistic regression using the Wald test on the collapsed burden.

There was an overlap of the sequenced subjects with the chip data subjects: of the 3264 subjects in the overlap, 3084 had passed our quality control including 2959 unrelated subjects of European ancestry with either late AMD (858), early AMD (1451), or no AMD (650). We conducted a sensitivity analysis for the burden test excluding the 858 advanced AMD subjects and the 650 control subjects (thus comparing15,286 advanced AMD subjects to 17,182 control subjects).

Follow-up queries for genes underneath the association signals. In order to derive information for all genes underneath our 52 identified association signals (spread across the 34 AMD loci), we built a gene list containing all genes that overlapped with a more narrow definition of locus regions: We have been using a particularly comprehensive definition of the locus region during the signal identification step (index variants and proxies, $r^2 \geq 0.5$, ±500kb), to avoid far-reaching linkage disequilibrium that may generate shadow signals (particularly in the light of strong associations in the *CFH*, *C3*, *C2*/*CFI*, and *ARMS2/HTRA1* loci) and to optimally differentiate independent signals within a locus. We have also used this wide locus region definition for the rare variant burden test again to fully correct for independent signals in the respective wider locus regions and to be conservative in the multiple testing corrections for the AMD-locus-wide burden test search. However, this wide definition is less adequate when prioritizing genes around the identified signals under the assumption that most protein-altering or regulating variants exert their effects in cis[42]. We thus focused the gene list for further queries to a more narrow locus region definition (index variants and proxies, $r^2 \geq 0.5$, ± 100kb) and yield 368 overlapping RefSeq genes (**Supplementary File 5**).

Gene expression. For the 368 genes in our gene list (see above), we sought to obtain gene expression in relevant tissues, retina, RPE, and choroid, in two independent data sets.

In the first laboratory (Dwight Stambolian Lab; University of Pennsylvania), we used RNA-Seq to characterize the chorioretinal transcriptomes in a discovery set of eight normal human eyes (two eyes from each of four persons)[75]. For each eye, we sequenced four RNA-Seq samples and generated close to 100 million 101-bp paired-end reads per sample. We mapped the sequence reads to the reference human genome (hg19) using GSNAP[76]. Our data are of high quality with 76–94% of the reads mapped to the human genome and 60–

81% retained after stringent quality control filtering, among which 86–93% mapped to genes defined by RefSeq. We considered the overall gene expression in each sample. Using filtered mapped reads, we estimated the expression levels of 23,569 RefSeq protein-coding genes using the fragments per kilobase of gene per million mapped fragment (FPKM) metric[77]. With coverage depth ranging from 66 to 133 million paired-end reads per sample, we detected expression of the majority of known protein-coding genes[75]. We considered genes and transcripts to be expressed, if FPKM > 0. Among the 23,569 genes with available expression for retina tissue or RPE/choroid/sclera, respectively, 290 genes for retina and 300 for RPE/choroid/sclera tissue overlapped with the 368 genes in the gene list.

In the second independent laboratory (Weber lab; University of Regensburg), we used RNA-Seq to estimate the relative abundance of known and novel transcripts in human retina, RPE and RPE-related cell types. Each tissue/cell line was sequenced as biological replicate, e.g. RNA was retrieved from cells of two individuals. The NextFlex Directional RNASeq library preparation kit (UDP based) from Bioo was used and between 30 and 60 million 75bp paired-end reads for each library were generated. The Tuxedo Tools pipeline (BowTie, TopHat, and Cufflinks) was used to map the reads to the genome and transcriptome and to quantify the abundance of transcripts measured as fragments per kilobase of gene per million mapped fragments (FPKM)[77]. We considered genes and transcripts to be expressed, if the respective FPKM value of the gene/transcript was greater than the first quartile of all FPKM values obtained from the tissues. Among the 20,590 with available expression available in retina tissue or RPE/choroid, respectively, 316 genes for retina and for REP/choroid overlapped with the 368 genes in gene list.

A consensus rating of gene expression observed in the two labs was derived as follows: Expression of a gene in one set of tissues (retina or RPE/choroid) was inferred, if both labs detected expression in the respective set of tissues; if at least one of the labs did not observe expression, the gene was considered as not expressed; gene expression of all other genes (one lab observing expression and the other with missing, or both labs with missing data) was regarded as missing.

Mouse model phenotypes. For the 368 genes in our gene list, we queried the Mouse Genome Informatics (MGI)[78] and the International Mouse Phenotyping Consortium (IPMC)[79] data bases (see **Web Resources**), and manually curated results by information from published literature. We determined whether a gene exhibited a relevant eye-phenotype (i.e. retina, RPE, or choroid phenotypes) in established genetic mouse models (knock-out, knock-in, or trans-genic mice).

Enrichment for molecular pathways. For the 368 overlapping genes, we performed functional enrichment analysis using INRICH[80] with default settings unless stated otherwise. Target intervals of this analysis were the narrow AMD locus regions (index variants and proxies, $r^2 \geq 0.5$, ± 100kb, **Supplementary Table 5**). Since there is no consensus approach to pathway analysis, we queried multiple data bases: (i) Kyoto Encyclopedia of Genes and Genomes (KEGG)[81], (ii) Reactome[82], and (iii) Gene Ontology (GO) Consortium[83] (see **Web Resources**). For example, while KEGG is a manually curated database on metabolic pathways, GO also includes automatic annotations and more comprehensive set of cellular processes and molecular functions. To reduce the multiple testing burden, we used gene sets with 5 to 200 genes that overlapped at least three overlapping target intervals. All our imputed/genotyped common (MAF in cases and controls combined ≥ 1%) variants genome-wide in these target regions were used to inform this analysis regarding variant density; no P-value threshold was used. We carried out the analysis with 1,000,000 replicates and 50,000 bootstrap rounds to yield corrected P-values, matching selected target regions in terms of gene count, variant density (80-120%) and total number of variants.

Drug pathways and targets. In order to derive information on whether the product of a gene among the 368 genes in our gene list was a direct drug target, we searched the DrugBank database (Version 4.1) which contains 4,207 drug targets (= genes) and 7,740 drugs [84](see **Web Resources**).

Explained variability in disease liability. Based on the 52 identified AMD variants, we estimated the explained proportion of disease liability explained by these variants (see **Web Resources**)[85] using the log Odds Ratio estimates from the model including all 52 identified variants (fully conditioned) to derive independent effect sizes. We compared this proportion explained by the 52 variants with the earlier derived genomic heritability based on all genotyped variants (see above).

Genetic risk score and relative and absolute genetic risk of AMD. For each individual, we computed a genetic risk score (GRS) as the effect size weighted sum of the AMD risk increasing alleles for all 52 independent variants divided by the sum of all effect sizes. For the weighting, the log Odds Ratios for each of the 52 variants were derived from the fully adjusted model (including all 52 variants), to assure independence of effect sizes.

In order to also derive a realistic genetic risk score distribution, we modeled a general population based on our case-control data by weighing each case individual using

$$w_{case} = Prevalence \,/\, (N_{cases} \,/\, (N_{cases} + N_{controls}))$$

and each control individual using

$$w_{control} = (1 - \textit{Prevalence}) / (N_{controls} / (N_{cases} + N_{controls})),$$

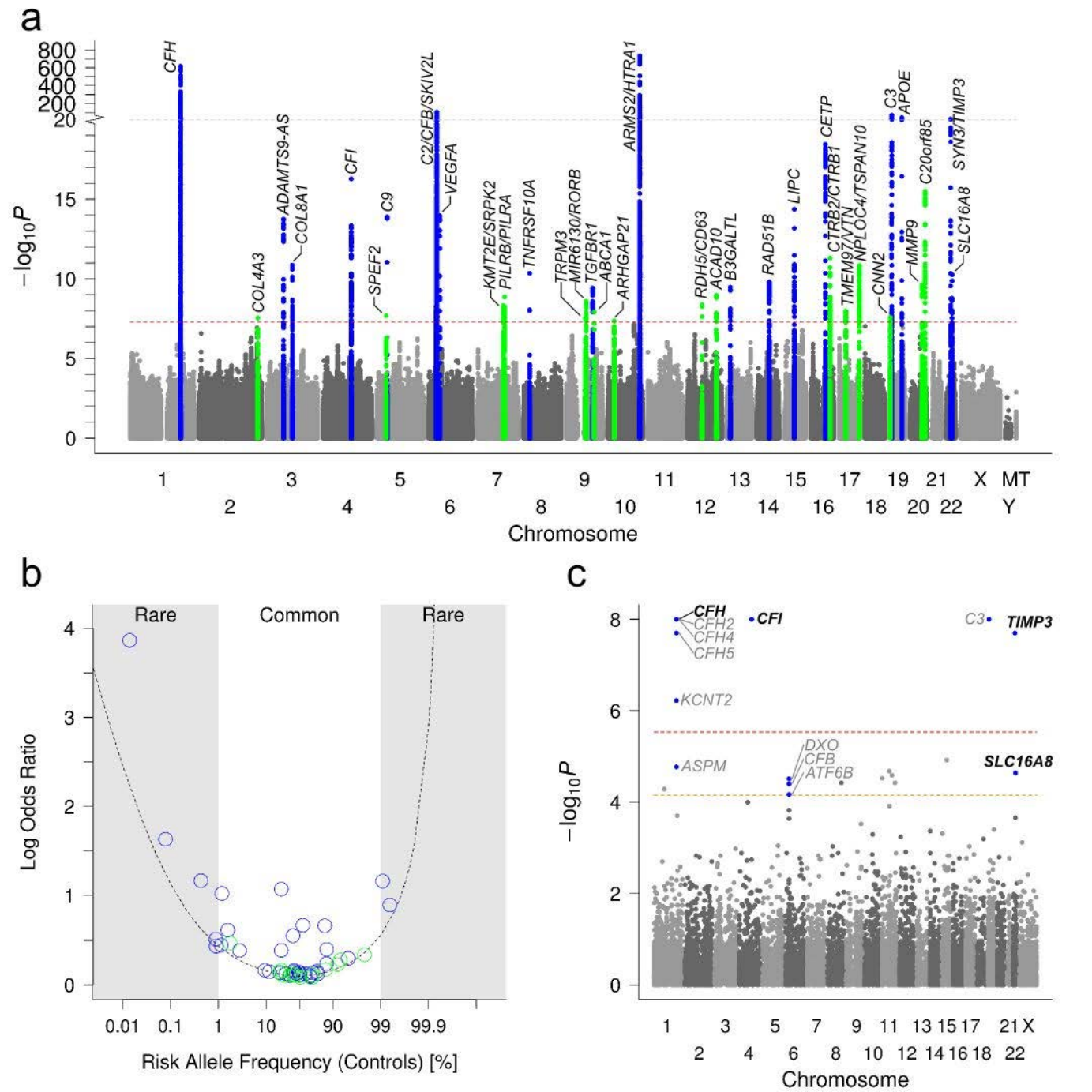
with *Prevalence* being an assumed prevalence of advanced AMD in the general population. We computed several scenarios using prevalence estimates of 1%, 5%, or 10% reflecting approximate prevalence of advanced AMD in the general population above the age of 50, 75, or 85 years of age, respectively. For this modeled general population, we derived the GRS distribution and its deciles.

We derived relative risk estimates (as Odds Ratios) for each GRS decile with the first decile as reference. This relative risk estimate per se is independent of the prevalence except that the decile to form the genetic risk groups used the GRS distribution as expected in a general population (which requires a prevalence assumption). We also computed absolute risk estimates per GRS decile, which is given by the proportion of advanced AMD cases applying the weights, again, as described above. This estimate depends on the prevalence.

References - Online Methods:

53. The International HapMap Consortium *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-61 (2007).
54. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
55. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-5 (2005).
56. Sherry, S.T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308-11 (2001).
57. Tennessen, J.A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64-9 (2012).
58. Age-Related Eye Disease Study Research Group. A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and E and beta carotene for age-related cataract and vision loss: AREDS report no. 9. *Arch Ophthalmol* **119**, 1439-52 (2001).
59. Fritsche, L.G. *et al.* A subgroup of age-related macular degeneration is associated with mono-allelic sequence variants in the ABCA4 gene. *Invest Ophthalmol Vis Sci* **53**, 2112-8 (2012).
60. Pruitt, K.D. *et al.* RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res* **42**, D756-63 (2014).
61. Ng, S.B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272-6 (2009).
62. Wildeman, M., van Ophuizen, E., den Dunnen, J.T. & Taschner, P.E. Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. *Hum Mutat* **29**, 6-13 (2008).
63. Delaneau, O., Marchini, J. & Zagury, J.F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-81 (2012).
64. Ristau, T. *et al.* Allergy is a protective factor against age-related macular degeneration. *Invest Ophthalmol Vis Sci* **55**, 210-4 (2014).
65. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* **44**, 955-9 (2012).
66. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-73 (2010).
67. Turner, S. *et al.* Quality control procedures for genome-wide association studies. *Curr Protoc Hum Genet* **Chapter 1**, Unit1 19 (2011).

68. Cavalli-Sforza, L.L. The Human Genome Diversity Project: past, present and future. *Nat Rev Genet* **6**, 333-40 (2005).
69. Lee, S.H., Wray, N.R., Goddard, M.E. & Visscher, P.M. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* **88**, 294-305 (2011).
70. Firth, D. Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27-38 (1993).
71. Ma, C., Blackwell, T., Boehnke, M., Scott, L.J. & Go, T.D.i. Recommended joint and meta-analysis strategies for case-control association testing of single low-count variants. *Genet Epidemiol* **37**, 539-50 (2013).
72. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).
73. Stephens, M. & Balding, D.J. Bayesian statistical methods for genetic association studies. *Nat Rev Genet* **10**, 681-90 (2009).
74. Wellcome Trust Case Control Consortium *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat Genet* **44**, 1294-301 (2012).
75. Li, M. *et al.* Comprehensive analysis of gene expression in human retina and supporting tissues. *Hum Mol Genet* **23**, 4001-14 (2014).
76. Wu, T.D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873-81 (2010).
77. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-5 (2010).
78. Blake, J.A., Bult, C.J., Eppig, J.T., Kadin, J.A. & Richardson, J.E. The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. *Nucleic Acids Res* **42**, D810-7 (2014).
79. Brown, S.D. & Moore, M.W. Towards an encyclopaedia of mammalian gene function: the International Mouse Phenotyping Consortium. *Dis Model Mech* **5**, 289-92 (2012).
80. Lee, P.H., O'Dushlaine, C., Thomas, B. & Purcell, S.M. INRICH: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics* **28**, 1797-9 (2012).
81. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).
82. Croft, D. *et al.* The Reactome pathway knowledgebase. *Nucleic Acids Res* **42**, D472-7 (2014).
83. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-9 (2000).
84. Law, V. *et al.* DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res* **42**, D1091-7 (2014).
85. So, H.C., Gui, A.H., Cherny, S.S. & Sham, P.C. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet Epidemiol* **35**, 310-7 (2011).

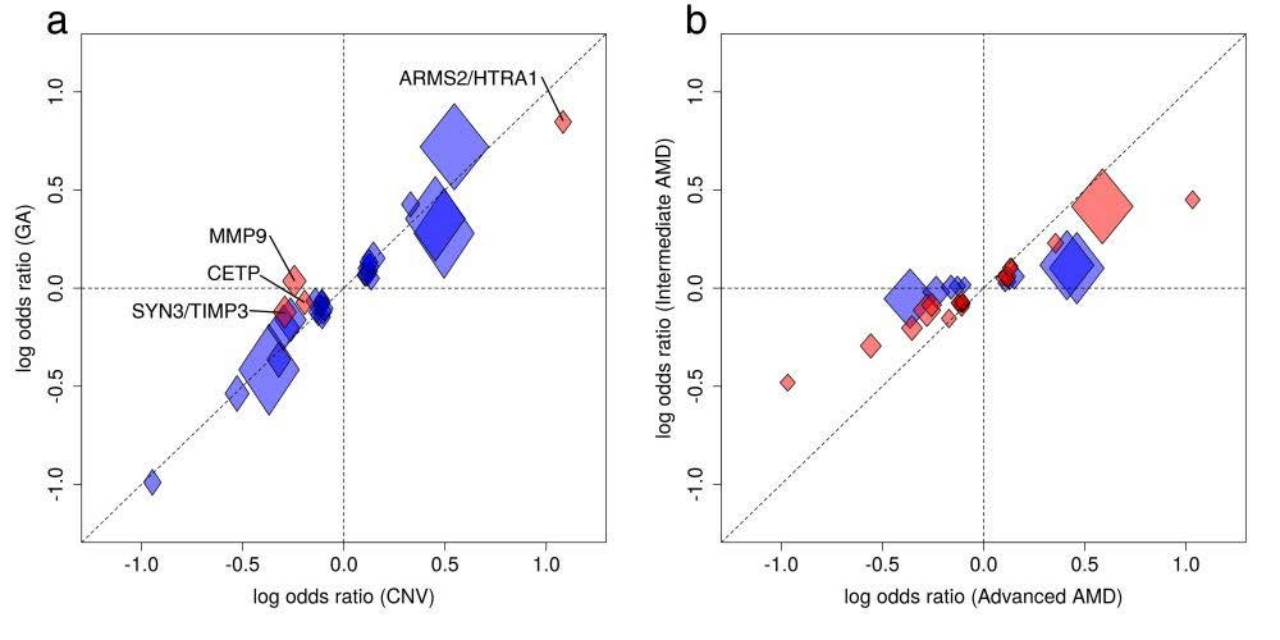

a
−log10P
CFH
COL4A3
ADAMTS9-AS
COL8A1
CFI
SPEF2
C9
C2/CFB/SKIV2L
VEGFA
KMT2E/SRPK2
PILRB/PILRA
TNFRSF10A
TRPM3
MIR6130/RORB
TGFBR1
ABCA1
ARHGAP21
ARMS2/HTRA1
RDH5/CD63
ACAD10
B3GALTL
RAD51B
LIPC
CETP
CTRB2/CTRB1
TMEM97/VTN
NPLOC4/TSPAN10
CNN2
C3
APOE
MMP9
C20orf85
SYN3/TIMP3
SLC16A8
Chromosome
b
Rare
Common
Rare
Log Odds Ratio
Risk Allele Frequency (Controls) [%]
c
CFH
CFH2
CFH4
CFH5
CFI
C3
TIMP3
KCNT2
ASPM
DXO
CFB
ATF6B
SLC16A8
−log10P
Chromosome

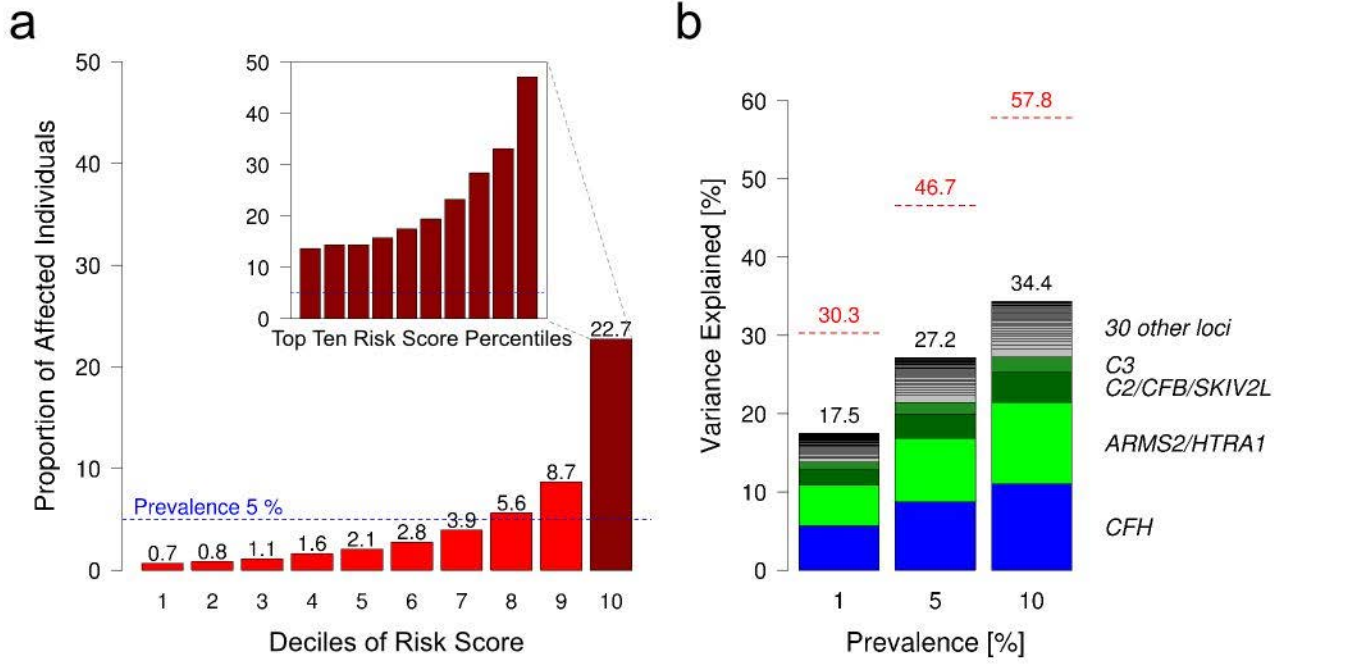
a

				Level of Evidence											
				Biology		Statistics		Annotation				Pathways	
				1	1	1	1	1	1	1	1	1	1
Gene	Locus #	Locus name	Gene Priority Score	Expression in retina/RPE/choroid	Eye phenotype in knock-out mouse	≥1 variant in 95% credible sets	Rare variant burden	Protein altering	5' UTR/3' UTR	Other exonic	Promoter region	Enriched molecular pathway	Drug target
COL4A3	2	|COL4A3|	4										
SPEF2	7	|PRLR/SPEF2|	3										
SRPK2	10	|KMT2E/SRPK2|	3										
PILRB	11	|PILRB/PILRA|	5										
TRPM3	14	|TRPM3|	4										
ABCA1	16	|ABCA1|	4										
ARHGAP21	17	|ARHGAP21|	3										
MMP19	19	|RDH5/CD63|	4										
RDH5	19	|RDH5/CD63|	4										
PTPN11	20	|ACAD10|	4										
BCAR1	25	|CTRB2/CTRB1|	4										
VTN	26	|TMEM97/VTN|	5										
TSPAN10	27	|NPLOC4/TSPAN10|	5										
GPX4	29	|CNN2|	4										
MMP9	31	|MMP9|	5										
C20orf85	32	|C20orf85|	0										

b

				Level of Evidence											
				Biology		Statistics		Annotation				Pathways	
				1	1	1	1	1	1	1	1	1	1
Gene	Locus #	Locus name	Gene Priority Score	Expression in retina/RPE/choroid	Eye phenotype in knock-out mouse	≥1 variant in 95% credible sets	Rare variant burden	Protein altering	5' UTR/3' UTR	Other exonic	Promoter region	Enriched molecular pathway	Drug target
CFH	1	|CFH|	7										
ADAMTS9	3	|ADAMTS9-AS2|	2										
ADAMTS9-AS2	3	|ADAMTS9-AS2|	2										
MIR548A2	3	|ADAMTS9-AS2|	2										
COL8A1	4	|COL8A1|	3										
CFI	5	|CFI|	5										
C9	6	|C9|	3										
CFB	8	|C2/CFB/SKIV2L|	7										
VEGFA	9	|VEGFA|	4										
TNFRSF10A	12	|TNFRSF10A|	3										
COL15A1	15	|TGFBR1|	3										
TGFBR1	15	|TGFBR1|	3										
HTRA1	18	|ARMS2/HTRA1|	3										
ARMS2	18	|ARMS2/HTRA1|	2										
B3GALTL	21	|B3GALTL|	2										
RAD51B	22	|RAD51B|	2										
LIPC	23	|LIPC|	3										
CETP	24	|CETP|	2										
HERPUD1	24	|CETP|	2										
NLRC5	24	|CETP|	2										
SLC12A3	24	|CETP|	2										
C3	28	|C3|	6										
APOE	30	|APOE|	5										
TIMP3	33	|SYN3/TIMP3|	5										
SLC16A8	34	|SLC16A8|	5										


a
b
ARMS2/HTRA1
MMP9
CETP
SYN3/TIMP3
log odds ratio (GA)
log odds ratio (CNV)
log odds ratio (Intermediate AMD)
log odds ratio (Advanced AMD)
1.0
0.5
0.0
-0.5
-1.0


a
Proportion of Affected Individuals
Deciles of Risk Score
Top Ten Risk Score Percentiles
Prevalence 5 %
0.7
0.8
1.1
1.6
2.1
2.8
3.9
5.6
8.7
22.7
b
Variance Explained [%]
Prevalence [%]
30.3
46.7
57.8
17.5
27.2
34.4
30 other loci
C3
C2/CFB/SKIV2L
ARMS2/HTRA1
CFH

Web Resources:

Full GWAS results: http://csg.sph.umich.edu/abecasis/public/amd2015/

The following Web Resources have been utilized:

GWAS catalog http://www.ebi.ac.uk/gwas/home),

Exome Variant Server, NHLBI GO Exome Sequencing Project:

http://evs.gs.washington.edu/EVS/

EPACTS: http://www.sph.umich.edu/csg/kang/epacts/index.html

SHAPEIT: https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html

MINIMAC: http://genome.sph.umich.edu/wiki/Minimac

1000 Genomes Reference Panel:

http://www.sph.umich.edu/csg/abecasis/MACH/download/1000G.2013-09.html

The Human Genome Diversity Project data:

http://genome.sph.umich.edu/wiki/LASER and http://www.hagsc.org/hgdp

SeattleSeq: http://snp.gs.washington.edu/SeattleSeqAnnotation138/index.jsp

Mutalyzer: https://mutalyzer.nl

NCBI Reference Sequence (RefSeq, downloaded December, 2012):

http://www.ncbi.nlm.nih.gov/refseq/

Human Splicing Finder 3.0: http://www.umd.be/HSF3/index.html

PubMed (retrieved November 11, 2014): http://www.pubmed.org

Mouse Genome Informatics (MGI) databases: http://www.informatics.jax.org

International Mouse Phenotyping Consortium Database: https://www.mousephenotype.org

INRICH: http://atgu.mgh.harvard.edu/inrich/

KEGG: Kyoto Encyclopedia of Genes and Genomes (KEGG): http://www.genome.jp/kegg/

MSigDB database v4.0: http://www.broadinstitute.org/gsea/index.jsp

Reactome (downloaded January 12th, 2015): http://www.reactome.org

Gene Ontology (GO) Consortium (downloaded January 12th, 2015): http://geneontology.org

DrugBank (downloaded June 4, 2014): http://www.drugbank.ca

GCTA: http://www.complextraitgenomics.com/software/gcta/

Variance explained by genetic variants:

https://sites.google.com/site/honcheongso/software/varexp

SUPPLEMENTARY INFORMATION

Content

Author Contributions

Clinical Ascertainment, Contribution of Samples, Study Coordination, and Data Analysis: Gonçalo R. Abecasis, Anita Agarwal, Jeeyun Ahn, Rando Allikmets, Isabelle Audo, Paul N. Baird, Elisa Bala, Mustapha Benchaboune, Hélène Blanché, John Blangero, Frédéric Blon, Alexis Boleda, Caroline Brandl, Kari E. Branham, Murray H. Brilliant, Kathryn P. Burdon, Melinda S. Cain, Peter Campochiaro, Albert Caramoy, Daniel Chen, David Cho, Itay Chowers, Ian J. Constable, Jamie E. Craig, Angela Cree, Christine Curcio, Margaret DeAngelis, Jean-François Deleuze, Anneke I. den Hollander, Bal Dhillon, Lebriz Ersoy, Lindsay A. Farrer, Sascha Fauser, Henry Ferreyra, Ken Flagg, Johanna R. Foerster, Lars G. Fritsche, Linn Gieser, Bamini Gopinath, Michael B. Gorin, Srinivas Goverdhan, Robyn H. Guymer, Shira Hagbi-Levi, Stephanie A. Hagstrom, Jonathan L. Haines, Janette Hall, Michael A. Hauser, Caroline Hayward, Scott J. Hebbring, John R. Heckenlively, Iris M. Heid, Alex W. Hewitt, Joshua D. Hoffman, Frank G. Holz, Carel B. Hoyng, David J. Hunter, Timothy Isaacs, Sudha K. Iyengar, Matthew P. Johnson, Nicholas Katsanis, Jane Khan, Ivana K. Kim, Terrie E. Kitchner, Caroline C. W. Klaver, Barbara E. K. Klein, Michael L. Klein, Ronald Klein, Jaclyn L. Kovach, Alan M. Kwong, Stewart Lake, Thomas Langmann, Reneé Laux, Yara T. E. Lechanteur, Kristine E. Lee, Thierry Léveillard, Mingyao Li, Helena Hai Liang, Gerald Liew, Danni Lin, Andrew Lotery, Hongrong Luo, David A. Mackey, Guanping Mao, Tammy M. Martin, Ian L. McAllister, J. Allie McGrath, Joanna E. Merriam, John C. Merriam, Stacy M. Meuer, Paul Mitchell, Saddek Mohand-Saïd, Anthony T. Moore, Emily L. Moore, Chelsea E. Myers, Anton Orlin, Mohammad I. Othman, Hong Ouyang, Kyu Hyung Park, Neal S. Peachey, Margaret A. Pericak-Vance, Eric A. Postel, Christina Rennie, Andrea J. Richardson, Guenther Rudolph, José-Alain Sahel, Nicole T. M. Saksens, Debra A. Schaumberg, Tina Schick, Hendrik P. N. Scholl, Stephen G. Schwartz, William K. Scott, Sebanti Sengupta, Humma Shahid, Giuliana Silvestri, R. Theodore Smith, Eric Souied, Emmanuelle Souzeau, Dwight Stambolian, Zhiguang Su, Anand Swaroop, Ava G. Tan, Barbara Truitt, Evangelia E. Tsironi, Cornelia M. van Duijn, Claudia N. von Strachwitz, Brendan J. Vote, Jie Jin Wang, Bernhard H. F. Weber, Daniel E. Weeks, Cindy Wen, Armin Wolf, Zhenglin Yang, John R. W. Yates, Donald Zack, Kang Zhang

Phenotype Committee: Ivana K. Kim (lead), Sudha K. Iyengar (lead), Margaret DeAngelis (lead), Gabriëlle H. S. Buitendijk, Emily Y. Chew, Itay Chowers, Anneke I. den Hollander, Sascha Fauser, Michael B. Gorin, Jonathan L. Haines, Iris M. Heid, Alex W. Hewitt, Caroline C. W. Klaver, Barbara E. K. Klein, Michael L. Klein, Ronald Klein, Thierry Léveillard, Andrew Lotery, Kyu Hyung Park, Jie Jin Wang, Kang Zhang

Data Curation: Jennifer L. Bragg-Gresham, Margaret DeAngelis, Lars G. Fritsche, Mathias Gorski, Wilmar Igl, Ivana K. Kim

Data Analysis

Team 2: single variant analysis: Lars G. Fritsche (lead), Iris M. Heid (lead), Gonçalo R. Abecasis (lead), Wilmar Igl (lead), Jennifer L. Bragg-Gresham, Gabriëlle H. S. Buitendijk, Valentina Cipriani, Margaret DeAngelis, Mathias Gorski, Felix Grassmann, Michelle Grunin, Jonathan L. Haines, Robert P. Igo Jr., Sudha K. Iyengar, Caroline C. W. Klaver, Matthias Olden, Klaus Stark, Xiaowei Zhan

Team 3: pathway and rare variant burden analysis: Lars G. Fritsche (lead), Jessica N. Cooke Bailey (lead), Matthew Schu (lead), Gonçalo R. Abecasis, Milam A. Brantley Jr., Matthew Brooks, Gabriëlle H. S. Buitendijk, Monique D. Courtenay, Margaret DeAngelis, Eiko K. de Jong, Anneke I. den Hollander, Lindsay A. Farrer, Felix Grassmann, Jonathan L. Haines, Iris M. Heid, Joshua D. Hoffman, Wilmar Igl, Robert P. Igo Jr., Sudha K. Iyengar, Yingda Jiang, Margaux A. Morrison, Matthias Olden, Margaret A. Pericak-Vance, Rebecca J. Sardell, William K. Scott, Klaus Stark, Anand Swaroop, Bernhard H. F. Weber, Daniel E. Weeks, Xiaowei Zhan

Team 4: analysis of non-SNP variation: Robert P. Igo Jr. (lead), Sudha K. Iyengar (lead), Paul N. Baird (lead), Gonçalo R. Abecasis, Monique D. Courtenay, Lars G. Fritsche, Jonathan L. Haines

Team 5: functional data analysis: Dwight Stambolian (lead), Bernhard H. F. Weber (lead), Margaret DeAngelis (lead), Sudha K. Iyengar (lead), Valentina Cipriani, Jessica N. Cooke Bailey, Monique D. Courtenay, Eiko K. de Jong, Anneke I. den Hollander, Sascha Fauser, Lars G. Fritsche, Felix Grassmann, Jonathan L. Haines, Caroline Hayward, Iris M. Heid, Wilmar Igl, Denise J. Morgan, Margaux A. Morrison, Rinki Ratnapriya, Chloe M. Stanton, Anand Swaroop, Xiaowei Zhan

Design of Overall Experiment: Gonçalo R. Abecasis, Margaret DeAngelis, Lars G. Fritsche, Jonathan L. Haines, Iris M. Heid, Sudha K. Iyengar, Margaret A. Pericak-Vance, Bernhard H. F. Weber

Genotyping and QC: Kimberly F. Doheny (lead), Jane Romm (lead), Lars G. Fritsche (lead), Mathias Gorski (lead), Gonçalo R. Abecasis, Jennifer L. Bragg-Gresham, Monique D.

Courtenay, Felix Grassmann, Jonathan L. Haines, Iris M. Heid, Joshua D. Hoffman, Wilmar Igl, Matthias Olden, Xiaowei Zhan

Writing Team: Lars G. Fritsche (lead), Iris M. Heid (lead), Gonçalo R. Abecasis, Jessica N. Cooke Bailey, Margaret DeAngelis, Jonathan L. Haines, Wilmar Igl, Sudha K. Iyengar, Ivana K. Kim, Dwight Stambolian, Bernhard H. F. Weber

Critical review of manuscript: Gonçalo R. Abecasis, Rando Allikmets, Paul N. Baird, Murray H. Brilliant, Itay Chowers, Jessica N. Cooke Bailey, Margaret DeAngelis, Sascha Fauser, Anneke I. den Hollander, Lindsay A. Farrer, Lars G. Fritsche, Michael B. Gorin, Stephanie A. Hagstrom, Jonathan L. Haines, Caroline Hayward, Iris M. Heid, Alex W. Hewitt, Wilmar Igl, Sudha K. Iyengar, Ivana K. Kim, Caroline C. W. Klaver, Barbara E. K. Klein, Michael L. Klein, Ronald Klein, Thierry Léveillard, Andrew Lotery, Paul Mitchell, Anthony T. Moore, Kyu Hyung Park, Neal S. Peachey, Margaret A. Pericak-Vance, Debra A. Schaumberg, Dwight Stambolian, Anand Swaroop, Jie Jin Wang, Bernhard H. F. Weber, Daniel E. Weeks, John R. W. Yates, Kang Zhang

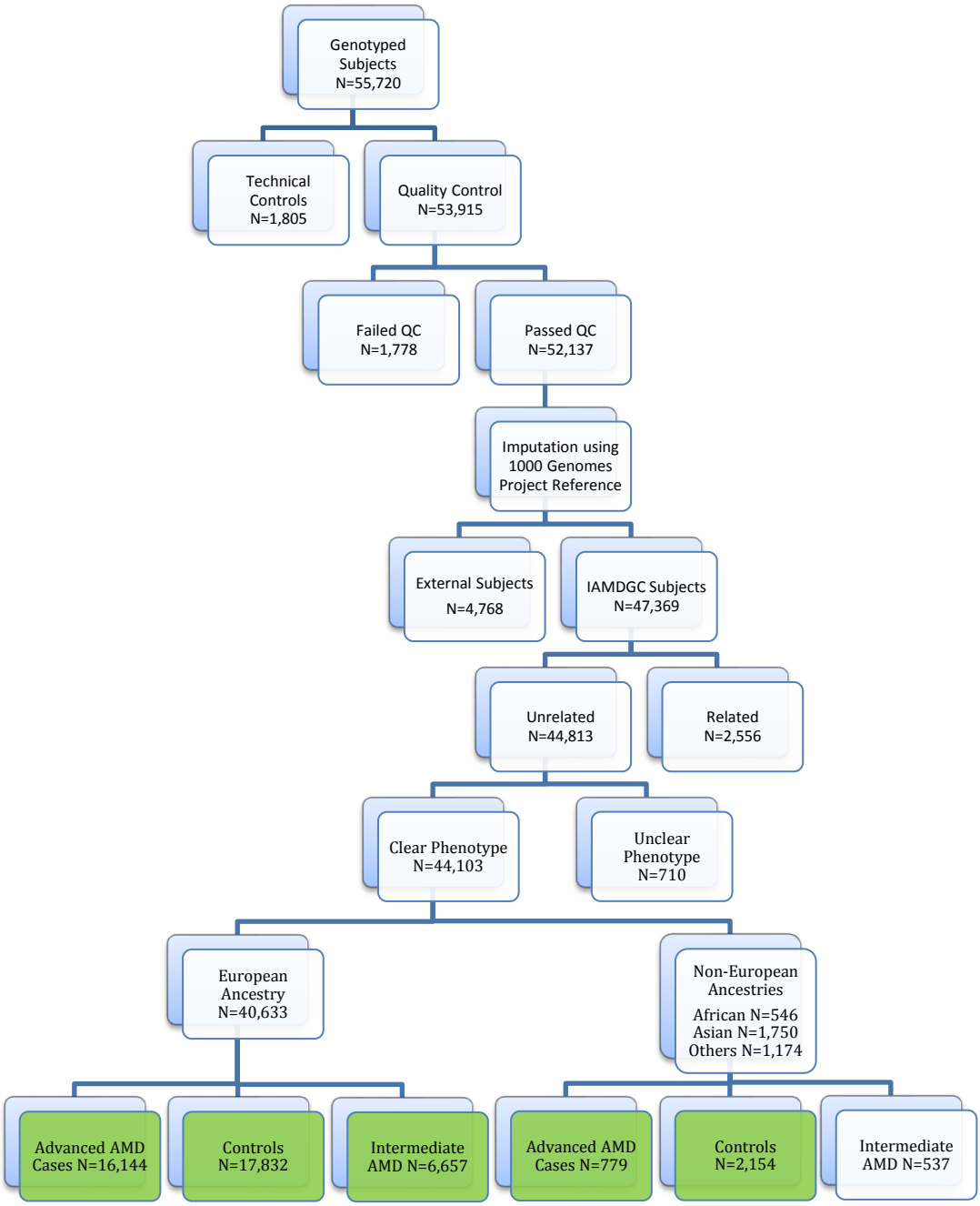
Steering Committee of IAMDGC consortium: Anand Swaroop, Gonçalo R. Abecasis, Alex W. Hewitt, Murray H. Brilliant, Kang Zhang, Bernhard H. F. Weber, Iris M. Heid, Margaret DeAngelis, Lindsay A. Farrer, Kyu Hyung Park, Ivana K. Kim, Dwight Stambolian, Thierry Léveillard, Andrew Lotery, Itay Chowers, Sudha K. Iyengar, Stephanie A. Hagstrom, Neal S. Peachey, Barbara E. K. Klein, Ronald Klein, Debra A. Schaumberg, Margaret A. Pericak-Vance, Paul Mitchell, Jie Jin Wang, Rando Allikmets, Anthony T. Moore, John R. W. Yates, Jonathan L. Haines, Sascha Fauser, Anneke I. den Hollander, Paul N. Baird, Michael L. Klein, Michael B. Gorin, Daniel E. Weeks, Caroline Hayward, Caroline C. W. Klaver

Senior Executive Committee of IAMDGC consortium: Gonçalo R. Abecasis, Margaret DeAngelis, Jonathan L. Haines, Sudha K. Iyengar, Margaret A. Pericak-Vance, Bernhard H. F. Weber

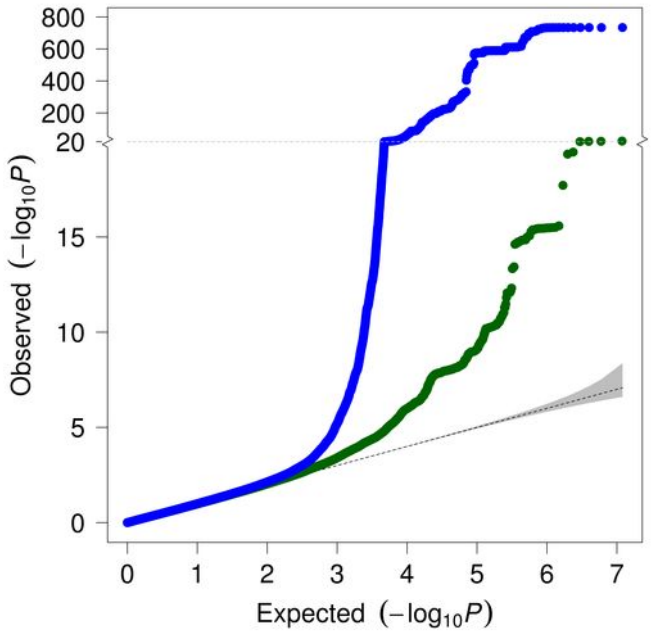
Conflicts of Interest

Inventor status for patents held by University of Pittsburgh regarding the 10q26 AMD susceptibility locus (DEW, MBG). VC, ATM and JRW are co-inventors or beneficiaries of patents related to genetic discoveries in AMD. IC serves as a consultant for Novartis, Bayer, Allergan, and Lycored. Royalties for AMD patents held by the University of Regensburg (LGF, BHFW), Royalties for AMD patents held by the University of Michigan (GRA, AS, MIO, KEB), Scientific Advisory Board for Regeneron Genetics Center (GRA). PM holds a consultant position for Bayer Inc. and Novartis Inc. AL has acted as a consultant to Bayer, Allergan, Roche and Novartis Pharmaceuticals. SGS has acted as a consultant to Alimera and Bausch + Lomb and has received writing fees from Vindico.

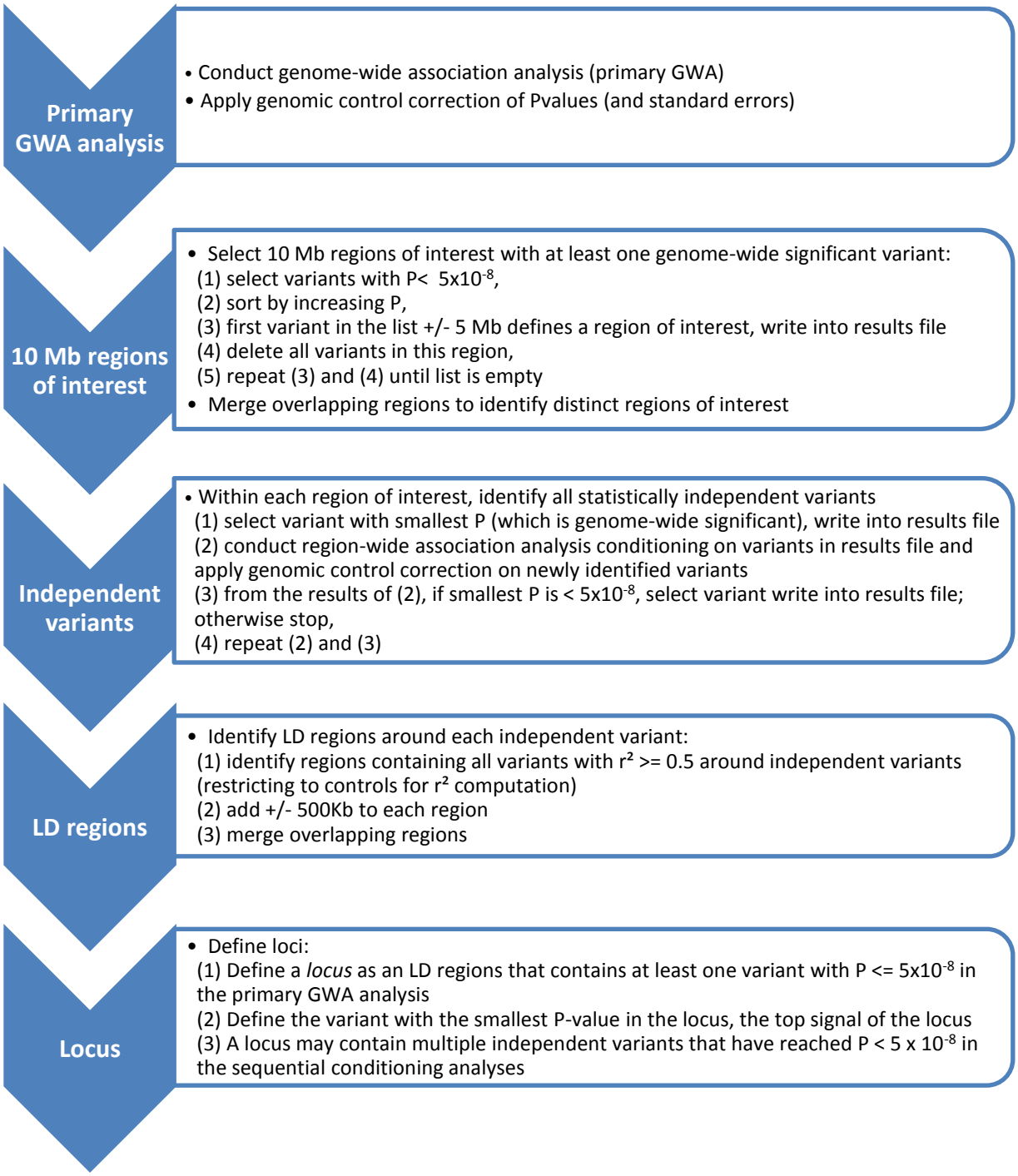
Supplementary Figure 1: Flow chart of subject processing and quality control. We removed technical controls and performed quality control (QC, e.g. exclusion of subjects with low call rate, violation of Hardy-Weinberg-equilibrium, or unexpected duplicates). We then imputed all remaining subjects together with the 1000 Genomes Project reference panel (Phase I). We excluded external subjects (sample collection unconnected to this project) and classified subjects by genetically inferred relatedness and ancestry. Advanced AMD cases with age below 50 years (N=211) or subjects with missing phenotypes (N=475) were classified as "unclear phenotype". Subjects with advanced AMD and controls of any ancestry as well as European subjects with intermediate AMD were analyzed (green boxes).



Supplementary Figure 2. Quantile-quantile plot for genome-wide single-variant association analysis. Shown are the observed P-values (-log10 *P*) from the single-variant association analysis (16,144 advanced AMD cases versus 17,832 controls) for all variants (blue) and without the variants in the known AMD loci (green) compared to those expected under the null hypothesis (no association). The black dotted line indicates the identity (no association) and the 95% confidence interval. The observed P-values are genomic control corrected by the lambda factor of 1.130.

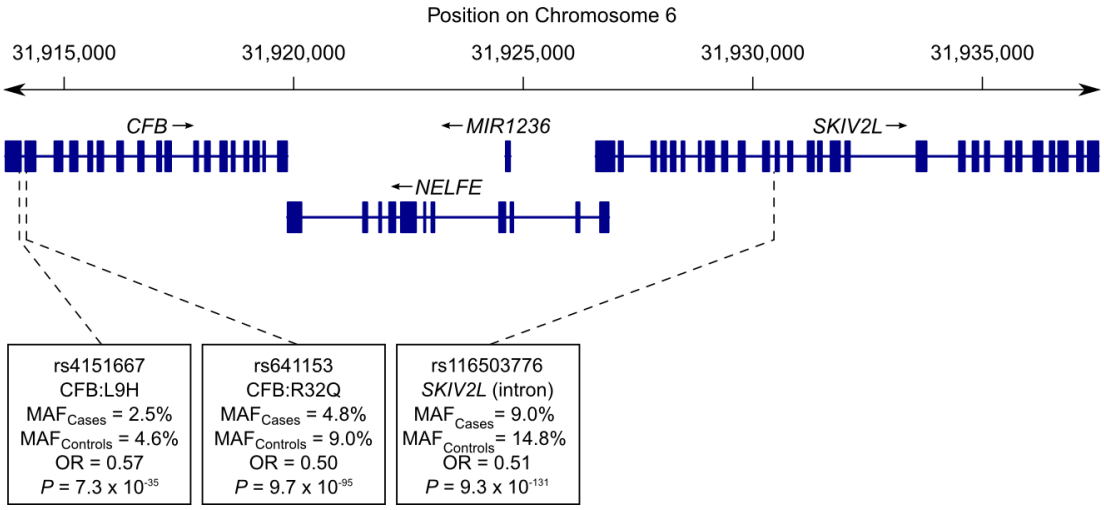


Supplementary Figure 3: **Locus identification procedure.**



Supplementary Figure 4: A counter example for credible set variants being able to depict the most likely causal variant(s) in the case of haplotype effects. Haplotype analysis elucidated that rs116503776 (*C2/CFB/SKIV2L)*, which is the sole 95% credible set variant in this signal, tags two previously described *CFB* missense variants, rs4151667 (*CFB*:L9H) and rs641153 (*CFB*:R32Q). Using the 16,144 patients and 17,832 controls, we derived the SNP and haplotype associations for the three variants, with haplotypes estimated during imputation. Shown are Odd Ratios (OR) and P-values from Firth's bias-corrected logistic regression, adjusting for principal components, DNA source, and for the other index variants in the locus (rs144629244, rs114254831, rs181705462, locus-wide conditioning). **(A)** The rs116503776 (*SKIV2L* intron) showed a stronger association than *CFB*:L9H or *CFB*:R32Q and is thus included into the 95% credible set as the statistically most likely causal variant (in fact as the sole credible set variant, posterior probability = 1.00; **Supplementary Table 7**). **(B)** Haplotype analysis revealed that it is not the A-allele of rs116503776 that carries the risk (H4), but rather its coinciding with the A-allele of L9H or R32Q (H2 and H3, respectively), which is supported by the rare haplotype with the R32Q A-allele without the rs116503776 A-allele (H5) carrying risk.

A



B

Haplotype No.*	Haplotype			Haplotype Frequency [%]		OR	P
	rs4151667 L9H	rs641153 R32Q	rs116503776	Cases	Controls		
H1	T	G	G	90.9	85.1	Reference	
H2	T	A	A	4.8	8.9	0.49	4.8×10^{-94}
H3	A	G	A	2.5	4.6	0.54	8.9×10^{-39}
H4	T	G	A	1.7	1.3	0.89	0.22
H5	T	A	G	0.1	0.2	0.53	0.0066

*Haplotype H6 “A-G-G” with frequency < 0.01 % was excluded.

Supplementary Table 1: Number of subjects per study. Stated are the numbers of unrelated individuals of European or non-European ancestry contained in each study after quality control. Quality control ensured that subjects were unique to one study.

Study Name	Advanced AMD			Intermediate AMD	Controls	Total
	Mixed [(a)]	CNV[(b)]	GA[(c)]			
AREDS	226	1,197	448	2,325	329	4,525
BDES	41	42	39	538	798	1,458
CWRU	322	294	237	171	646	1,670
Cambridge	127	579	139	0	421	1,266
EUGENDA - Cologne	12	280	9	231	577	1,109
Columbia	81	321	89	239	546	1,276
EU / JHU	36	518	254	260	668	1,736
Edinburgh	0	180	44	145	194	563
Jerusalem	0	305	1	43	207	556
Marshfield	149	10	26	563	2,738	3,486
Melbourne	51	457	67	0	404	979
Miami	69	525	114	349	364	1,421
MMAP - Michigan	192	351	174	98	633	1,448
NHS / HPF	0	154	0	258	1,029	1,441
Oregon	69	429	164	0	276	938
MMAP - Penn	96	459	146	212	853	1,766
Pittsburgh	210	254	115	130	130	839
Regensburg	365	935	367	159	1,149	2,975
Rotterdam	10	102	13	0	0	125
Southampton	17	254	63	167	588	1,089
UCSD	72	1,040	172	149	2,049	3,482
EUGENDA - Neijmegen	12	314	35	220	445	1,026
Utah	6	676	59	77	1,107	1,925
UWA, LEI & Flinders	2	1,040	432	0	2,493	3,967
Vanderbilt	0	391	78	181	552	1,202
Westmead / Sydney	24	299	43	142	790	1,298
Total	**2,189**	**11,406**	**3,328**	**6,657**	**19,986**	**43,566**

(a) subjects with CNV in one eye and GA in the other eye, (b) subjects with CNV in at least one eye, (c) subjects with GA in at least one eye and no evidence of CNV in either eye

Supplementary Table 2. Recruitment and ascertainment per study. Detailed information was gathered regarding recruitment and ascertainment for each of the 26 studies from the study ophthalmologists and principal investigators. Grading of intermediate AMD include AREDS[1], a modified AREDS (Clinical-Age-Related Maculopathy Staging, CARMS)[2], the Wisconsin[3], the International[4], and the EUGENDA[5] grading systems.

Study Name	Recruitment			Ascertainment via fundus photo/exam			Ascertain-ment via FA / OCT	Grading system	Ref.
	Adv. AMD	Int. AMD	Controls	Adv. AMD	Int. AMD	Controls	CNV	Int. AMD	
AREDS	Clinic	Clinic	Clinic	Yes	Yes	Yes	OCT	AREDS	1,6
BDES	Population	Population	Population	Yes	Yes	Yes	OCT	Modified Wisconsin	7
Cambridge	Clinic	NA	Spouses/friends	Yes	NA	Yes	OCT	International	8
EUGENDA-Cologne	Clinic/volunt.	Clinic/volunt.	Population/volunt.	Yes	Yes	Yes	FA/OCT	EUGENDA	5
Columbia	Clinic	Clinic	Clinic	Yes	Yes	Yes	OCT	Modified International	9
CWRU	Clinic	Clinic	Clinic	Yes	Yes	Yes	No	AREDS	10
Edinburgh	Clinic	Clinic	Clinic	Yes	Yes	Yes	No	International	8
EU / JHU [(a)]	Mixed	Mixed	Mixed	Yes	Yes	Yes/No [(a)]	FA/OCT[(a)]	International/AREDS	11
Jerusalem	Clinic	Clinic	Clinic	Yes	Yes	Yes	FA/OCT	AREDS	12,13
Marshfield	Clinic	Clinic	Population	Yes	Yes	Yes	FA/OCT	customized	14
Melbourne	Clinic	Clinic	Population	Yes	Yes	Yes	FA/OCT	International	11
Miami	Clinic	Clinic	Spouses/friends	Yes	Yes	Yes	FA/OCT	Modified AREDS	15
MMAP-Michigan	Clinic	Clinic	Clinic	Yes	Yes	Yes	No	International	16
NHS / HPFS	Population	Population	Population	No [(b)]	No [(b)]	No [(b)]	No	None	17
Oregon	Clinic	Clinic	Clinic	Yes	Yes	Yes	FA	AREDS	11
MMAP-Penn	Clinic	Clinic	Clinic	Yes	Yes	Yes	No	AREDS	16
Pittsburgh	Clinic	Clinic	Clinic	Yes	Yes	Yes	No	customized	18
Regensburg	Clinic	Clinic	Spouses/friends	Yes	Yes	Yes	No	International	19
Rotterdam	Clinic	NA	NA	Yes	NA	NA	OCT	Wisconsin/modified Int.	20
Southampton	Clinic	Clinic	Clinic	Yes	Yes	Yes	FA/OCT	AREDS	21,22
UCSD	Clinic	Clinic	Clinic	Yes	Yes	Yes	OCT	AREDS	5
EUGENDA-Neijmegen	Clinic	Popul./clinic	Population/clinic	Yes	Yes	Yes	FA/OCT	EUGENDA	23
Utah [(c)]	Mixed	Mixed	mixed	Yes [(c)]	Yes [(c)]	Yes [(c)]	FA/OCT	AREDS	24-26
UWA/LEI & Flinders[(d)]	Mixed	Mixed	Mixed	Yes	Yes	Yes	No	International	15
Vanderbilt	Clinic	Clinic	Spouses/friends	Yes	Yes	Yes	FA/OCT	Modified AREDS	27-29
Westmead/Sydney[(e)]	Clinic	Clinic	Mixed	Yes	Yes	Yes	FA/OCT	Modified Wisconsin	

Clinic = recruitment via ophthalmology clinic; Population = recruitment via general population study; Volunt. = recruitment of volunteers; Int. AMD = Intermediate AMD, Adv. AMD = Advanced AMD; FA = Fluorescin Angiography; OCT = Optical Coherence Tomography; CNV=choroidal neovascularization; Ref = Reference; EUG = EUGENDA (a) EU/JHU includes several studies: Jerusalem, Paris, Southampton, Baltimore, Bonn, and Créteil: all with FA/OCT (except Paris: OCT, Southampton: FA); (b) NHS/HPF with cases via practitioner-confirmed self-report, controls self-reported unaffected. (c) Utah includes four collections: SNUBH, Greece (cases/controls from clinics), NESC (cases from clinic, controls from population), Timor (cases/controls from population); all cases ascertained via fundus photography/exam. (d) UWA recruited from population, LEI & Flinders from clinic; (e) Sydney-Westmead includes: Blue Mountains Eye Study (controls from population), unpublished study by Paul Mitchell (cases from clinic), Cataract Surgery and AMD study (cases/controls from clinic).

Supplementary Table 3: Study characteristics. Shown are: (**A)** the number of genotyped and imputed variants by predicted function, and **(B)** the number of subjects, in the analysed data set after quality control.

A

	# Variants total	# rare variants MAF[a] < 1%	# moderately common variants MAF[a] [1 – 10%)	# very common variants MAF[a] ≥ 10%
Genotyped variants[b]				
Non-synonymous excluding frameshift and stop codon	152,046	133,064	10,319	8,663
Premature stop codon	6,820	6,660	107	53
Frame-shift	2,328	2,280	24	24
Splice Site	2,520	2,428	51	41
Other	275,636	30,263	31,003	214,370
Total genotyped	439,350	174,695	41,504	223,151
Imputed variants [b]				
Non-synonymous [c]	12,180	3,600	4,212	4,368
Premature Stop Codons	233	59	73	101
Frame-shift	349	75	138	136
Splice Site	171	53	62	56
Other	11,571,547	2,871,531	3,788,768	4,911,248
Total imputed	11,584,480	2,875,318	3,793,253	4,915,909
TOTAL	12,023,830	3,050,013	3,834,757	5,139,060

[a] MAF = minor allele frequencies in the 17,832 European controls. [b] Annotated based on RefSeq hg19 using SeattleSeq Annotation 138.

B

	Sample Size	Female	Age (yrs)[d]
	N	**%**	**Mean (SD)**
European Ancestry			
Advanced AMD	16,144	60%	76.79 (8.26)
GA[a]	3,235	59%	76.81 (8.62)
CNV[b]	10,749	60%	76.68 (8.14)
Mixed[c]	2,160	60%	77.28 (8.25)
Intermediate AMD	6,657	59%	74.30 (8.76)
Controls	17,832	56%	70.71 (9.70)
Total European Ancestry	40,633	58%	73.70 (9.42)
Non-European Ancestry[e]			
Advanced AMD	779	50%	72.75 (9.15)
Controls	2,154	55%	67.00 (11.56)
Total Non-European Ancestry	2,933	55%	69.08 (10.95)
Total	43,566	58%	73.33 (9.63)

GA = geographic atrophy, CNV = choroidal neovascularisation, SD = standard deviation.
[a] Patients with GA in at least one eye and no evidence of CNV in any eye; [b] Patients with CNV in at least one eye; [c] Patients with CNV in at least one eye and GA in at least one eye and both eyes affected; [d] For cases, age at onset if available, otherwise age at exam; for controls, age at exam; [e] Non-European Ancestries included 1,572 Asians (473 advanced AMD cases, 1,099 controls), 413 Africans (52 advanced AMD cases, 361 controls), and 948 subjects with other ancestries (254 advanced AMD cases, 694 controls).

Supplementary Table 4: Identification of 52 independent AMD risk variants in 34 loci. The single variant genome-wide association analysis in 16,144 advanced AMD cases and 17,832 controls of European ancestry identified 52 independent variants through a sequential forward selection regression approach (see **Supplementary Figure 3**) within 34 loci (locus region defined by the independent variants, their correlated variants, $r^2 \geq 0.5$, ±500kb, merging overlapping regions). Horizontal lines separate loci that were analysed jointly during sequential forward selection due to their physical proximity (within 10 Mb). Shown are effect sizes (Odds Ratios, OR) and P-values of the primary analysis (unconditioned single variant association analysis), the sequential forward selection as the variant identifying analysis, and the fully conditioned analysis (all 52 index variants in the model) to derive independent ORs.

Signal number[a]	Locus name	Index variant[e]	RSQ[f]	Chr:Position	Major / minor allele	Minor allele frequency		Primary analysis		Sequential forward selection [c]		Fully conditioned analysis [d]	
						Cases	Controls	OR	*P*	OR	P	OR	P
1.1	*CFH*	rs10922109	1.00	1:196,704,632	C/A	.223	.426	0.38	9.6×10^{-618}	...	...	0.51	1.0×10^{-131}
1.2	*CFH*	rs570618	1.00	1:196,657,064	G/T	.580	.364	2.38	2.0×10^{-590}	1.65	6.0×10^{-112}	1.74	9.2×10^{-76}
1.3	*CFH*	rs121913059	-	1:196,716,375	C/T	.003	.00014	20.28	8.9×10^{-24}	33.2	1.8×10^{-32}	47.63	2.2×10^{-35}
1.4	*CFH*	rs148553336	-	1:196,613,173	T/C	.003	.009	0.29	8.6×10^{-26}	0.31	8.9×10^{-21}	0.31	8.8×10^{-17}
1.5	*CFH*	rs187328863	0.83	1:196,380,158	C/T	.054	.028	2.27	1.1×10^{-68}	1.43	9.2×10^{-13}	1.47	2.8×10^{-12}
1.6	*CFH (CFHR3/CFHR1)*[b]	rs61818925	0.87	1:196,815,450	G/T	.284	.385	0.60	6.0×10^{-165}	1.18	1.3×10^{-10}	1.18	6.3×10^{-9}
1.7	*CFH*	rs35292876	-	1:196,706,642	C/T	.021	.009	2.42	8.2×10^{-37}	1.55	2.9×10^{-9}	1.54	9.5×10^{-8}
1.8	*CFH*	rs191281603	0.42	1:196,958,651	C/G	.007	.006	1.07	0.68	0.39	1.4×10^{-8}	0.41	7.7×10^{-7}
2	*COL4A3*	rs11884770	0.98	2:228,086,920	C/T	.258	.278	0.90	2.9×10^{-8}	...	...	0.92	2.6×10^{-4}
3	*ADAMTS9-AS2*	rs62247658	1.00	3:64,715,155	T/C	.466	.433	1.14	1.8×10^{-14}	...	...	1.14	7.8×10^{-11}
4.1	*COL8A1*	rs140647181	0.77	3:99,180,668	T/C	.023	.016	1.59	1.4×10^{-11}	...	...	1.85	1.6×10^{-14}
4.2	*COL8A1*	rs55975637	0.99	3:99,419,853	G/A	.132	.117	1.15	1.3×10^{-8}	1.16	2.3×10^{-9}	1.16	3.8×10^{-7}
5.1	*CFI*	rs10033900	-	4:110,659,067	C/T	.511	.477	1.15	5.4×10^{-17}	...	...	1.15	1.2×10^{-13}
5.2	*CFI*	rs141853578	-	4:110,685,820	C/T	.003	.0008	3.64	6.3×10^{-10}	3.87	8.6×10^{-11}	5.12	7.4×10^{-12}
6	*C9*	rs62358361	0.98	5:39,327,888	G/T	.016	.009	1.80	1.3×10^{-14}	...	...	1.67	7.2×10^{-9}
7	*PRLR/SPEF2*	rs114092250	0.88	5:35,494,448	G/A	.016	.021	0.70	2.1×10^{-8}	0.70	3.6×10^{-8}	0.71	9.5×10^{-6}
8.1	*C2/CFB/SKIV2L*	rs116503776	-	6:31,930,462	G/A	.090	.148	0.57	$1.2x\ 10^{-103}$	...	...	0.51	5.0×10^{-96}
8.2	*C2/CFB/SKIV2L*	rs144629244	-	6:31,946,792	G/A	.016	.012	1.39	2.6×10^{-6}	2.50	5.9×10^{-35}	2.79	1.0×10^{-32}
8.3	*C2/CFB/SKIV2L (PBX2)*[b]	rs114254831	-	6:32,155,581	A/G	.284	.260	1.13	9.4×10^{-12}	1.15	1.8×10^{-14}	1.13	8.8×10^{-9}
8.4	*C2/CFB/SKIV2L*	rs181705462	1.00	6:31,947,027	G/T	.018	.012	1.55	3.1×10^{-10}	1.49	1.0×10^{-8}	1.56	2.8×10^{-8}
9	*VEGFA*	rs943080	-	6:43,826,627	T/C	.465	.497	0.88	1.1×10^{-14}	...	...	0.87	5.8×10^{-13}
10	*KMT2E/SRPK2*	rs1142	0.99	7:104,756,326	C/T	.370	.346	1.11	1.4×10^{-9}	...	...	1.14	1.3×10^{-10}
11	*PILRB/PILRA*	rs7803454	1.00	7:99,991,548	C/T	.209	.190	1.13	4.8×10^{-9}	1.13	3.7×10^{-9}	1.15	2.8×10^{-9}

12	*TNFRSF10A*	rs79037040	-	8:23,082,971	T/G	.451	.479	0.90	4.5×10^{-11}	...	...	0.89	5.1×10^{-9}
13	*MIR6130/RORB*	rs10781182	1.00	9:76,617,720	G/T	.328	.306	1.11	2.6×10^{-9}	1.10	...	1.12	1.5×10^{-6}
14	*TRPM3*	rs71507014	0.99	9:73,438,605	GC/G	.427	.405	1.10	3.0×10^{-8}	...	3.6×10^{-8}	1.11	2.3×10^{-8}
15	*TGFBR1*	rs1626340	1.00	9:101,923,372	G/A	.189	.209	0.88	3.8×10^{-10}	...	...	0.88	4.0×10^{-7}
16	*ABCA1*	rs2740488	0.96	9:107,661,742	A/C	.255	.275	0.90	1.2×10^{-8}	0.90	1.0×10^{-8}	0.89	6.0×10^{-7}
17	*ARHGAP21*	rs12357257	0.98	10:24,999,593	G/A	.243	.223	1.11	4.4×10^{-8}	...	...	1.12	1.8×10^{-6}
18	*ARMS2/HTRA1*	rs3750846	1.00	10:124,215,565	T/C	.436	.208	2.81	6.5×10^{-735}	...	...	2.93	6.0×10^{-645}
19	*RDH5/CD63*	rs3138141	0.59	12:56,115,778	C/A	.222	.207	1.16	4.3×10^{-9}	...	...	1.18	4.7×10^{-8}
20	*ACAD10*	rs61941274	0.73	12:112,132,610	G/A	.024	.018	1.51	1.1×10^{-9}	...	...	1.60	3.2×10^{-9}
21	*B3GALTL*	rs9564692	1.00	13:31,821,240	C/T	.277	.299	0.89	3.3×10^{-10}	...	...	0.90	1.0×10^{-6}
22.1	*RAD51B*	rs61985136	0.97	14:68,769,199	T/C	.360	.384	0.90	1.6×10^{-10}	...	...	0.88	8.2×10^{-10}
22.2	*RAD51B*	rs2842339	-	14:68,986,999	A/G	.107	.094	1.14	1.4×10^{-6}	1.17	1.6×10^{-8}	1.18	3.3×10^{-7}
23.1	*LIPC*	rs2043085	-	15:58,680,954	T/C	.350	.381	0.87	4.3×10^{-15}	...	...	1.15	7.7×10^{-13}
23.2	*LIPC*	rs2070895	-	15:58,723,939	G/A	.195	.217	0.87	2.4×10^{-11}	0.87	1.8×10^{-11}	0.86	1.8×10^{-10}
24.1	*CETP*	rs5817082	0.99	16:56,997,349	C/CA	.232	.264	0.84	3.6×10^{-19}	...	...	0.87	2.7×10^{-8}
24.2	*CETP*	rs17231506	1.00	16:56,994,528	C/T	.348	.315	1.16	2.2×10^{-18}	1.11	1.7×10^{-8}	1.11	1.2×10^{-6}
25	*CTRB2/CTRB1*	rs72802342	0.84	16:75,234,872	C/A	.067	.080	0.79	5.0×10^{-12}	...	...	0.79	8.0×10^{-9}
26	*TMEM97/VTN*	rs11080055	0.98	17:26,649,724	C/A	.463	.486	0.91	1.0×10^{-8}	...	...	0.92	1.5×10^{-5}
27	*NPLOC4/TSPAN10*	rs6565597	0.84	17:79,526,821	C/T	.400	.381	1.13	1.5×10^{-11}	...	...	1.12	2.1×10^{-7}
28.1	*C3*	rs2230199	-	19:6,718,387	C/G	.266	.208	1.43	3.8×10^{-69}	...	...	1.47	1.6×10^{-60}
28.2	*C3*	rs147859257	-	19:6,718,146	T/G	.012	.004	2.86	3.1×10^{-28}	3.14	6.0×10^{-33}	3.22	4.1×10^{-26}
28.3	*C3 (NRTN/FUT6)* [b]	rs12019136	0.92	19:5,835,677	G/A	.036	.048	0.71	2.4×10^{-15}	0.71	5.7×10^{-15}	0.74	4.0×10^{-9}
29	*CNN2*	rs67538026	0.83	19:1,031,438	C/T	.460	.498	0.90	2.6×10^{-8}	0.90	1.7×10^{-8}	0.90	1.4×10^{-6}
30.1	*APOE*	rs429358	0.99	19:45,411,941	T/C	.099	.135	0.70	2.4×10^{-42}	...	...	0.67	3.9×10^{-39}
30.2	*APOE(EXOC3L2/MARK4)*[b]	rs73036519	0.93	19:45,748,362	G/C	.284	.302	0.91	3.1×10^{-7}	0.90	4.6×10^{-8}	0.91	2.4×10^{-5}
31	*MMP9*	rs142450006	0.91	20:44,614,991	TTTTC/T	.124	.141	0.85	2.4×10^{-10}	...	...	0.84	5.3×10^{-9}
32	*C20orf85*	rs201459901	0.98	20:56,653,724	T/TA	.054	.070	0.76	3.1×10^{-16}	...	...	0.76	3.8×10^{-12}
33	*SYN3/TIMP3*	rs5754227	0.99	22:33,105,817	T/C	.109	.137	0.77	1.1×10^{-24}	...	...	0.79	5.7×10^{-16}
34	*SLC16A8*	rs8135665	-	22:38,476,276	C/T	.217	.195	1.14	5.5×10^{-11}	1.14	1.1×10^{-10}	1.14	1.4×10^{-8}

[a] Independent signals within loci that were detected by sequential forward selection are indexed by their corresponding signal number, jointly analyzed regions (10 Mb regions) were sorted by position, while the signals (within each of the 10 Mb regions) are sorted according to their discovery by the sequential forward selection; [b] The peak of this independent association signal is located closer to genes in parenthesis than to the locus-defining genes; [c] Step-wise conditional analysis per 10Mb region; [d] Fully conditioned regression model including all 52 variants. [e] dbSNPID of the signal index variant. [f] Imputation quality for imputed variants.

Supplementary Table 5: Regions of the 34 AMD loci. We defined AMD locus regions by the independently associated variants (**Supplementary Table 4**), their proxies, $r^2 \geq 0.5$ and added ± 500kb in order to catch far-reaching LD and potential rare variants missed by the r^2 criterion; overlapping regions were merged to define one joint locus with multiple signals.

AMD locus region Merged LD intervals after adding +/- 500kb region				Index variant			LD interval of farthermost variants with $r^2 \geq 0.5$ to index variants			
							Lower boundary		Upper boundary	
Locus name	**Chr**	**Start**	**End**	**Signal #**	**dbSNP ID**	**Position**	**dbSNP ID (r^2)**	**Position**	**dbSNP ID (r^2)**	**Position**
CFH	1	195,679,832	197,768,053	1.1	rs10922109	196,704,632	rs1831282 (0.500)	196,673,993	rs6675769 (0.503)	196,941,661
				1.2	rs570618	196,657,064	rs1329424 (0.990)	196,646,176	rs6428379 (0.557)	196,937,536
				1.3	rs121913059	196,716,375	(= index variant)		(= index variant)	
				1.4	rs148553336	196,613,173	rs192646611 (0.503)	196,179,832	rs139686286 (0.591)	197,268,053
				1.5	rs187328863	196,380,158	rs79524406 (0.962)	196,206,732	rs41314023 (0.691)	197,101,263
				1.6	rs61818925	196,815,450	rs6695321 (0.652)	196,675,861	rs139124820 (0.559)	196,905,991
				1.7	rs35292876	196,706,642	rs139452966 (0.631)	196,267,858	rs148050899 (0.844)	196,950,807
				1.8	rs191281603	196,958,651	rs187073701 (0.522)	196,395,310	rs148448529 (0.808)	197,114,403
COL4A3	2	227,573,015	228,592,110	2	rs11884770	228,086,920	rs13004766 (0.640)	228,073,015	rs13417485 (0.506)	228,092,110
ADAMTS9-AS	3	64,199,445	65,230,121	3	rs62247658	64,715,155	rs4611812 (0.972)	64,699,445	rs17676309 (0.959)	64,730,121
COL8A1	3	98,551,114	100,381,567	4.1	rs140647181	99,180,668	rs17823383 (0.591)	99,051,114	rs78245803 (0.535)	99,491,751
				4.2	rs55975637	99,419,853	rs142158920 (0.514)	99,323,745	rs13074029 (0.509)	99,881,567
CFI	4	110,126,506	111,185,820	5.1	rs10033900	110,659,067	rs7660005 (0.614)	110,626,506	(= index variant)	
				5.2	rs141853578	110,685,820	(= index variant)		(= index variant)	
C9	5	38,699,134	39,831,894	6	rs62358361	39,327,888	rs62358735 (0.776)	39,199,134	rs34882957 (0.998)	39,331,894
PRLR/SPEF2	5	34,769,332	36,493,378	7	rs114092250	35,494,448	rs141424971 (0.705)	35,269,332	rs144746304 (0.560)	35,993,378
C2/CFB/SKIV2L	6	30,505,490	33,238,589	8.1	rs116503776	31,930,462	rs149176277 (0.693)	31,894,355	(= index variant)	
				8.2	rs144629244	31,946,792	rs141397370 (0.501)	31,005,490	rs199858290 (0.539)	32,738,589
				8.3	rs114254831	32,155,581	rs116246398 (0.568)	32,147,696	rs115393945 (0.564)	32,190,406
				8.4	rs181705462	31,947,027	rs147277589 (0.505)	31,028,132	rs188309731 (0.512)	32,303,245
VEGFA	6	43,305,296	44,329,629	9	rs943080	43,826,627	rs2094197 (0.622)	43,805,296	rs7742835 (0.635)	43,829,629
KMT2E/SRPK2	7	104,081,402	105,563,372	10	rs1142	104,756,326	rs2470938 (0.620)	104,581,402	rs12535854 (0.681)	105,063,372
PILRB/PILRA	7	99,394,940	100,611,776	11	rs7803454	99,991,548	rs113667434 (0.755)	99,894,940	rs7778181 (0.592)	100,111,776

AMD locus region Merged LD intervals after adding +/- 500kb region				Index variant			LD interval of farthermost variants with $r^2 \geq 0.5$ to index variants			
							Lower boundary		Upper boundary	
Locus name	**Chr**	**Start**	**End**	**Signal #**	**dbSNP ID**	**Position**	**dbSNP ID (r^2)**	**Position**	**dbSNP ID (r^2)**	**Position**
TNFRSF10A	8	22,582,971	23,588,984	12	rs79037040	23,082,971	(= index variant)		rs11777697 (0.758)	23,088,984
MIR6130/RORB	9	75,935,160	77,189,752	13	rs10781182	76,617,720	rs1458496 (0.561)	76,435,160	rs1491451 (0.514)	76,689,752
TRPM3	9	72,938,605	73,946,180	14	rs71507014	73,438,605	(= index variant)		rs10780973 (0.582)	73,446,180
TGFBR1	9	101,358,102	102,431,769	15	rs1626340	101,923,372	rs11388918 (0.627)	101,858,102	rs334371 (0.503)	101,931,769
ABCA1	9	107,139,414	108,167,147	16	rs2740488	107,661,742	rs4149274 (0.568)	107,639,414	rs2254708 (0.541)	107,667,147
ARHGAP21	10	24,360,361	25,556,538	17	rs12357257	24,999,593	rs11014165 (0.886)	24,860,361	rs112276510 (0.507)	25,056,538
ARMS2/HTRA1	10	123,702,126	124,735,355	18	rs3750846	124,215,565	rs58649964 (0.58)	124,202,126	rs2672587 (0.664)	124,235,355
RDH5/CD63	12	55,615,585	56,713,297	19	rs3138141	56,115,778	rs3138142 (0.998)	56,115,585	rs56108400 (0.873)	56,213,297
ACAD10	12	110,919,995	113,502,935	20	rs61941274	112,132,610	rs61943028 (0.633)	111,419,995	rs73209657 (0.760)	113,002,935
B3GALTL	13	31,242,232	32,339,274	21	rs9564692	31,821,240	rs9539510 (0.531)	31,742,232	rs9572914 (0.878)	31,839,274
RAD51B	14	68,227,506	69,550,783	22.1	rs61985136	68,769,199	rs1957569 (0.564)	68,727,506	rs1028577 (0.956)	68,815,261
				22.2	rs2842339	68,986,999	rs61985769 (0.847)	68,973,917	rs56034765 (0.595)	69,050,783
LIPC	15	58,171,721	59,242,418	23.1	rs2043085	58,680,954	rs1601934 (0.587)	58,671,721	rs409668 (0.502)	58,706,926
				23.2	rs2070895	58,723,939	rs7170361 (0.507)	58,718,998	rs261336 (0.520)	58,742,418
CETP	16	56,485,514	57,506,829	24.1	rs5817082	56,997,349	rs1864163 (0.988)	56,997,233	rs289713 (0.571)	57,006,829
				24.2	rs17231506	56,994,528	rs72786786 (0.746)	56,985,514	rs1532624 (0.585)	57,005,479
CTRB2/CTRB1	16	74,732,528	76,017,115	25	rs72802342	75,234,872	rs72802340 (0.702)	75,232,528	rs72789426 (0.539)	75,517,115
TMEM97/VTN	17	26,092,946	27,240,139	26	rs11080055	26,649,724	rs241771 (0.522)	26,592,946	rs11869677 (0.674)	26,740,139
NPLOC4/TSPAN10	17	79,015,509	80,186,552	27	rs6565597	79,526,821	rs56737642 (0.585)	79,515,509	rs11868178 (0.687)	79,686,552
C3	19	5,311,717	7,224,340	28.1	rs2230199	6,718,387	rs2230203 (0.777)	6,710,782	rs163494 (0.749)	6,724,340
				28.2	rs147859257	6,718,146	(= index variant)		(= index variant)	
				28.3	rs12019136	5,835,677	rs201167147 (0.731)	5,811,717	rs150092813 (0.716)	5,846,277
CNN2	19	523,867	1,533,360	29	rs67538026	1,031,438	rs8102732 (0.766)	1,023,867	rs55725239 (0.544)	1,033,360
APOE	19	44,892,254	46,313,830	30.1	rs429358	45,411,941	rs6857 (0.683)	45,392,254	rs66626994 (0.603)	45,428,234
				30.2	rs73036519	45,748,362	rs346751 (0.511)	45,734,433	rs2377326 (0.596)	45,813,830
MMP9	20	44,114,991	45,160,699	31	rs142450006	44,614,991	(= index variant)		rs3859613 (0.517)	44,660,699
C20orf85	20	56,084,276	57,174,034	32	rs201459901	56,653,724	rs77125533 (0.503)	56,584,276	rs874430 (0.583)	56,674,034
SYN3/TIMP3	22	32,546,536	33,613,375	33	rs5754227	33,105,817	rs4373007 (0.559)	33,046,536	rs16991084 (0.553)	33,113,375
SLC16A8	22	37,795,271	39,003,972	34	rs8135665	38,476,276	rs2010472 (0.563)	38,295,271	rs56076229 (0.579)	38,503,972

Supplementary Table 6. Results of the 34 AMD lead variants in individuals of Non-European ancestry. We computed frequencies and advanced AMD association in an Asian (473 cases, 1,099 controls), African (52 cases, 361 controls), and "other ancestry" group (254 cases, 694 controls) for our 34 lead variants. Shown are association results (frequencies, Odds Ratios into the direction of the minor allele in European controls) in these three ancestry groups, re-stating association in the European analysis for comparability (16,144 cases, 17,832 controls). The "other ancestry" group is comprised of subjects with no distinct assignment to a specific ancestry group. Extended results can be found in **Supplementary File 2.**

Locus name	Lead variant	European major/ minor allele	Frequency of minor allele*						Odds Ratio				
			Non-European cases	Asian controls	African controls	"Other" controls	Non-European controls	European controls	Asian	African	"Other"	Non-European	European
CFH	rs10922109	C/A	0.344	0.583	0.540	0.510	0.552	0.426	0.43	0.80	0.44	0.45	0.38
COL4A3	rs11884770	C/T	0.273	0.260	0.468	0.352	0.324	0.278	0.83	1.01	1.00	0.90	0.90
ADAMTS9/AS2	rs62247658	T/C	0.752	0.823	0.848	0.748	0.803	0.433	0.88	0.76	0.80	0.81	1.14
COL8A1	rs140647181	T/C	0.008	0.001	0.001	0.014	0.006	0.016	4.36	0.62	1.00	1.16	1.59
CFI	rs10033900	C/T	0.593	0.684	0.373	0.478	0.565	0.477	0.91	0.80	1.29	0.95	1.15
C9	rs62358361	G/T	0.001	0	0.001	0.004	0.002	0.009	NA	2.81	0.69	0.70	1.80
PRLR/SPEF2	rs114092250	G/A	0.003	0.000	0.004	0.005	0.002	0.021	6.60	0.81	1.77	1.57	0.70
C2/CFB/SKIV2L	rs116503776	G/A	0.088	0.127	0.213	0.182	0.159	0.148	0.48	1.11	0.63	0.60	0.57
VEGFA	rs943080	T/C	0.286	0.326	0.161	0.321	0.297	0.497	0.87	1.10	0.78	0.84	0.88
KMT2E/SRPK2	rs1142	C/T	0.342	0.366	0.239	0.289	0.320	0.346	0.94	0.99	1.25	1.04	1.11
PILRB/PILRA	rs7803454	C/T	0.064	0.018	0.037	0.120	0.054	0.190	1.00	0.72	1.00	1.08	1.13
TNRSF10B	rs79037040	T/G	0.569	0.619	0.857	0.647	0.668	0.479	0.80	0.59	0.79	0.80	0.90
TRPM3	rs71507014	GC/G	0.558	0.581	0.554	0.524	0.558	0.306	1.04	0.92	1.00	0.99	1.10
MIR6130/RORB	rs10781182	G/T	0.581	0.544	0.829	0.575	0.602	0.405	1.09	1.03	1.19	1.12	1.11
TGFBR1	rs1626340	G/A	0.371	0.449	0.304	0.266	0.366	0.209	1.04	0.73	0.85	0.97	0.88
ABCA1	rs2740488	A/C	0.282	0.327	0.401	0.345	0.345	0.275	0.76	0.64	1.01	0.79	0.90
ARHGAP21	rs12357257	G/A	0.093	0.034	0.118	0.147	0.084	0.223	0.89	0.67	1.17	1.14	1.11
ARMS2/HTRA1	rs3750846	T/C	0.569	0.387	0.212	0.246	0.312	0.208	2.83	2.16	2.28	2.60	2.81
RDH5/CD63	rs3138141	C/A	0.110	0.103	0.025	0.110	0.092	0.207	0.79	1.11	1.57	1.05	1.16

Locus name	Lead variant	European major/ minor allele	Frequency of minor allele*						Odds Ratio				
			Non-European cases	Asian controls	African controls	"Other" controls	Non-European controls	European controls	Asian	African	"Other"	Non-European	European
ACAD10	rs61941274	G/A	0.006	0.002	0.003	0.007	0.004	0.018	-	1.91	2.94	2.46	1.51
B3GALTL	rs9564692	C/T	0.626	0.746	0.443	0.443	0.598	0.299	0.93	1.35	1.12	0.98	0.89
RAD51B	rs61985136	T/C	0.591	0.495	0.844	0.584	0.582	0.384	1.49	0.85	0.91	1.30	0.90
LIPC	rs2043085	T/C	0.500	0.505	0.645	0.516	0.532	0.381	1.02	0.99	1.07	0.99	0.87
CETP	rs5817082	C/CA	0.191	0.108	0.593	0.354	0.269	0.264	0.77	0.87	1.02	0.92	0.84
CTRB2	rs72802342	C/A	0.077	0.093	0.013	0.053	0.067	0.080	1.06	2.43	0.83	0.98	0.79
TMEM97/VTN	rs11080055	C/A	0.308	0.361	0.433	0.454	0.403	0.486	0.63	0.86	0.87	0.69	0.91
NPLOC4/TSPAN10	rs6565597	C/T	0.221	0.117	0.214	0.235	0.171	0.381	1.54	1.57	1.26	1.54	1.13
C3	rs2230199	C/G	0.060	0.009	0.056	0.107	0.048	0.208	0.65	1.90	1.22	1.38	1.43
CNN2	rs67538026	C/T	0.604	0.835	0.188	0.401	0.587	0.498	1.05	0.94	0.68	0.86	0.90
APOE	rs429358	T/C	0.094	0.135	0.219	0.167	0.159	0.135	0.57	0.59	0.88	0.61	0.70
MMP9	rs142450006	TTTTC/T	0.116	0.116	0.068	0.106	0.105	0.141	0.93	0.43	1.12	0.99	0.85
C20orf85	rs201459901	T/TA	0.023	0.007	0.014	0.040	0.019	0.070	0.41	2.66	1.13	1.17	0.76
SYN3/TIMP3	rs5754227	T/C	0.509	0.531	0.533	0.362	0.477	0.137	1.39	1.05	0.78	1.19	0.85
SLC16A8	rs8135665	C/T	0.166	0.148	0.343	0.274	0.222	0.195	0.71	0.78	1.08	0.84	1.14

* Major and minor allele in European Controls; NA = not applicable due to the variant being monomorph in this group.

Supplementary Table 7: Credible sets of variants per independent signal. Stated are the 95% credible sets of variants for those of the 52 independent signals that contain ≤ 5 variants (for a full list of 95% credible sets, see **Supplementary File 3**) using a Bayesian approach[21]; for loci with multiple signals, the association conditioned on the other signals in the loci was computed before estimating the Bayes Factor. Assuming that each signal has exactly one causal variant and this is contained in our data, such credible sets contain the causal variant with 95% probability. In parentheses, the posterior probability of each variant (add up to ≥ 95% per row by design) and the amino acid exchange (bold, if applicable).

Locus name	Signal number	Index variant	95% credible sets	
			# Variants	Variant (posterior probability; amino acid exchange, if applicable)
CFH	1.3*	rs121913059	3	rs121913059 (0.899, **R1210C**), rs7540032 (0.0272), rs7514261 (0.0267)
	1.4*	rs148553336	3	rs148553336 (0.634), rs139017763 (0.268, **G278S**), rs182780863 (0.0806)
CFI	5.1*	rs10033900	1	rs10033900 (0.999)
	5.2*	rs141853578	1	rs141853578 (0.996, **G119R**)
C9	6	rs62358361	2	rs62358361 (0.551), rs34882957 (0.447, **P167S**)
C2/CFB/SKIV2L	8.1*	rs116503776	1	rs116503776 (1)
	8.3*	rs114254831	4	rs114254831 (0.738), rs114764276 (0.107), rs116820884 (0.0975), rs115219661 (0.0443)
VEGFA	9	rs943080	4	rs943080 (0.424), rs7758685 (0.271), rs4711751 (0.22), rs1536304 (0.0556)
TNFRSF10A	12	rs79037040	1	rs79037040 (0.983)
ABCA1	16	rs2740488	3	rs2740488 (0.738), rs2575876 (0.158), rs1883025 (0.0666)
RDH5/CD63	19	rs3138141	3	rs3138141 (0.498), rs3138142 (0.401, **I141I**), rs56108400 (0.0943)
LIPC	23.1*	rs2043085	2	rs2414577 (0.49), rs2414578 (0.477)
	23.2*	rs2070895	5	rs2070895 (0.339), rs1800588 (0.303), rs1077835 (0.132), rs1077834 (0.131), rs35980001 (0.052)
C3	28.1*	rs2230199	1	rs2230199 (0.983, **R102G**)
	28.2*	rs147859257	1	rs147859257 (1, **K155Q**)
	28.3*	rs12019136	5	rs12019136 (0.538), rs17855739 (0.212, **E247K**), rs78060698 (0.0876), rs79744308 (0.0815, **A59T**), rs7255720 (0.0642, **R158R**)
APOE	30.1*	rs429358	1	rs429358 (1, **C130R**)
SYN3/TIMP3	33	rs5754227	4	rs5754227 (0.717), rs7289865 (0.183), rs5754222 (0.039), rs11379058 (0.0381)
SLC16A8	34	rs8135665	4	rs8135665 (0.643), rs11089861 (0.271), rs742398 (0.0248), rs144402192 (0.0207)

* For each locus with multiple signals (identified by sequential forward selection and LD analysis) we calculated their independent association signals by adding the other variants of that locus as covariate to the model before estimating the Bayes Factor

Supplementary Table 8: Non-synonymous variants in 95% credible sets. We derived 95% credible sets of variants for each of the 52 independent signals. Stated are the non-synonymous variants among these variants that reach ≥ 5% posterior probability (for a full list of variants in 95% credible sets, see **Supplementary File 3**). Shown are Odds Ratios (OR) and P-values from the locus-wide conditioned analysis (adjusting for the identified index variant(s) in the locus); these were used to compute the Bayes factor (not shown) and the posterior probability. References are given for studies that first described the association between of these non-synonymous variants and the risk for AMD.

Locus Name	Locus #	Variant in credible set	Chr:Position	Major / minor allele	Gene	Role	CADD Score*	MAF		OR	P	Post. prob	Ref.
								Cases	Controls				
CFH	1.3*	rs121913059	1:196,716,375	C/T	*CFH*	R1210C	15.48	0.00319	0.00014	31.8	3.2×10^{-31}	89.9%	[30]
	1.4*	rs139017763	1:196,965,193	G/A	*CFHR5*	G278S	16.16	0.00246	0.00858	0.32	1.9×10^{-18}	26.8%	[31]
CFI	5.2*	rs141853578	4:110,685,820	C/T	*CFI*	G119R	13.88	0.00288	0.000786	3.87	8.6×10^{-11}	99.6%	[31,32]
C9	6	rs34882957	5:39,331,894	G/A	*C9*	P167S	19.88	0.0156	0.00871	1.79	1.6×10^{-14}	44.7%	[31]
C2/CFB/SKIV2L	8.2*	rs114190211	6:31,929,737	C/T	*SKIV2L*	R324W	18.24	0.0163	0.012	2.59	3.9×10^{-37}	8.6%	new
ARMS2/HTRA1	18	rs10490924	10:124,214,448	G/T	*ARMS2*	A69S	8.897	0.436	0.208	2.81	1.9×10^{-734}	9.4%	[18,19]
C3	28.1*	rs2230199	19:6,718,387	C/G	*C3*	R102G	7.214	0.266	0.208	1.45	7.9×10^{-74}	98.3%	[8]
	28.2*	rs147859257	19:6,718,146	T/G	*C3*	K155Q	9.992	0.0124	0.00432	3.12	1.5×10^{-32}	100%	[22,31,33]
	28.3*	rs17855739	19:5,831,840	C/T	*FUT6*	E247K	16.35	0.0376	0.0494	0.72	1.6×10^{-14}	21.2%	new
		rs79744308	19:5,827,765	G/A	*NRTN*	A59T	15.14	0.0354	0.0463	0.71	4.3×10^{-14}	8.2%	new
APOE	30.1	rs429358	19:45,411,941	T/C	*APOE*	C130R	0.007	0.0993	0.135	0.70	3.7×10^{-43}	100%	[34,35]

MAF = minor allele frequency, Ref. = Reference; post. prob. = posterior probability; * CADD, Combined Annotation Dependent Depletion (CADD) scores combine the generality of conservation-based metrics with the specificity of subset-relevant functional metrics to estimate the variants' relative pathogenicity ([36], http://cadd.gs.washington.edu; ranging from 1 to 99).

Supplementary Table 9: Results of the rare variant burden analysis without and with conditioning on nearby identified variants. We computed the gene-based burden test based on rare protein altering variants comparing the 16,144 advanced AMD cases versus 17,832 controls using the Variable Threshold test. Stated are the 14 genes with significant burden ($P < 0.05/703 = 7.1 \times 10^{-5}$) among the 703 genes in the 34 AMD loci (defined by 52 identified variants, proxies, $r^2 > 0.5$, ± 500 kb); the 703 genes are those among the 17,044 coding RefSeq genes that overlap with the 34 loci and contain at least one rare protein altering variants. When conditioning on the identified index variant(s) in the same locus (locus-wide conditioning), many of the burden results disappear implying that the initial signal was a shadow effect of the common variant; for four genes, *CFI*, *TIMP3*, *CFH* and *SLC16A8*, the rare variant burden remains significant after locus-wide conditioning (bold). Rare variants are defined as MAF < 1% in each ancestry group (European, African, and Asian); protein altering variants are defined as non-synonymous variants (missense, stop loss, in-frame insertion/deletion, frameshift, premature stop codon) or splice sites. More details are given in **Supplementary File 4**.

Gene	# variants in analysis	Unconditioned				Locus-wide conditioned			
		Optimal RAC	# Variants with MAC ≤ optimal RAC	Effect direction[a]	*P*	Optimal RAC	# Variants with MAC ≤ optimal RAC	Effect direction[a]	*P*
CFI	**47**	**121**	**44**	**+**	**1.0×10^{-8}**	**46**	**43**	**+**	**1.0×10^{-8}**
TIMP3	**10**	**14**	**9**	**+**	**2.0×10^{-8}**	**14**	**9**	**+**	**9.0×10^{-8}**
CFH	**49**	**108**	**47**	**+**	**1.0×10^{-8}**	**10**	**37**	**+**	**1.2×10^{-6}**
SLC16A8	**9**	**648**	**9**	**+**	**2.3×10^{-5}**	**648**	**9**	**+**	**3.1×10^{-6}**
F13B	10	249	9	+	1.7×10^{-5}	33	8	+	0.0024
DXO	14	446	14	+	3.1×10^{-5}	446	14	+	0.012
CFHR4	12	293	12	-	1.0×10^{-8}	24	9	+	0.028
CFB	15	135	15	+	4.0×10^{-5}	135	15	+	0.07
KCNT2	9	207	9	-	6.0×10^{-7}	9	7	+	0.12
ATF6B	21	394	21	+	6.8×10^{-5}	21	21	+	0.13
CFHR2	6	413	6	-	2.0×10^{-8}	2	3	+	0.15
ASPM	68	634	68	-	1.0×10^{-8}	4	24	+	0.37
C3	52	553	52	+	1.0×10^{-8}	69	49	+	0.37
CFHR5	26	376	25	-	1.0×10^{-8}	1	10	+	0.41

RAC: risk allele count among cases and controls; MAC: minor allele count among cases and control; MAF = minor allele frequency
[a] "+" indicates an AMD risk increasing by the accumulated rare variants, and "–" an AMD risk decreasing burden.

Supplementary Table 10. Enriched pathways for genes in the 34 AMD loci. We queried pathway databases (Reactome, GO, KEGG) using INRICH for the 368 genes in the 34 narrow AMD locus regions (identified 52 variants and proxies, $r^2 \geq 0.5$, ±100kb, see **Methods**). Shown are the enriched pathways (corrected P<0.10, analyzing known and novel AMD loci); P-values, and the genes contributing to the enrichment (those in novel AMD loci in bold letters).

Data-base	Pathway	#genes in gene set	Known AMD loci			Known and novel AMD loci			Genes contributing to enrichment (those in novel AMD loci in bold)
			#loci in gene set	Empiri-cal P	Correc-ted P [a]	#loci in gene set	Empiri-cal P	Correc-ted P [a]	
Reac-tome	Complement Pathways								
	Regulation of Complement Cascade	22	5	1.0×10^{-6}	4.3×10^{-4}	6 (5+1)	1.0×10^{-6}	2.8×10^{-4}	*C2/C4A/C4B/CFB, C3, C9, CFH/CFHR3, CFI,* ***VTN***
	Complement Cascade	36	5	5.0×10^{-6}	0.0013	6 (5+1)	1.0×10^{-6}	2.8×10^{-4}	*C2/C4A/C4B/CFB, C3, C9, CFH/CFHR3, CFI,* ***VTN***
	Collagen Pathways								
	Assembly of Collagen Fibrils & Other Multimeric Structures	54	3	0.0011	0.11	6 (3+3)	1.0×10^{-6}	2.8×10^{-4}	*COL15A1, COL8A1, LOXL2,* ***COL4A3/COL4A4, MMP9, PCOLCE***
	Collagen Formation	84	3	0.0061	0.31	6 (3+3)	3.9×10^{-5}	0.0081	*COL15A1, COL8A1, LOXL2,* ***COL4A3/COL4A4, MMP9, PCOLCE***
	Lipid Pathways								
	Lipoprotein Metabolism	30	3	5.4×10^{-4}	0.065	5 (3+2)	2.0×10^{-5}	0.0041	*APOC2/APOE, CETP, LIPC,* ***ABCA1, PLTP***
	HDL-Mediated Lipid Transport	15	2	n/a [b]	n/a [b]	4 (2+2)	1.1×10^{4}	0.024	*APOC2/APOE, CETP,* ***ABCA1, PLTP***
	Lipid Digestion, Mobilization, and Transport	50	3	0.0016	0.14	5 (3+2)	2.3×10^{-4}	0.045	*APOC2/APOE, CETP, LIPC,* ***ABCA1, PLTP***
GO	Complement Pathways								
	Regulation of Complement Activation [a]	24	5	2.0×10^{-6}	0.0034	6 (5+1)	1.0×10^{-6}	0.0024	*C2/C4A/C4B/CFB, C3, C9, CFH, CFI,* ***VTN***
	Lipid Pathways								
	High-Density Lipoprotein Particle [b]	22	4	8.2×10^{-5}	0.067	5 (4+1)	4.4×10^{-5}	0.055	*APOC1/APOC4/APOE, APOM, CETP, LIPC,* ***ABCA1***
	Reverse Cholesterol Transport [c]	18	4	9.6×10^{-5}	0.074	5 (4+1)	3.8×10^{-5}	0.048	*APOC2/APOE, APOM, CETP, LIPC,* ***ABCA1***
	Phospholipid Transporter Activity [d]	7	1	n/a [b]	n/a [b]	3 (1+2)	8.7×10^{-5}	0.097	*CETB,* ***ABCA1, ABCA7***
	Extracellular Matrix Pathways								
	Extracellular Matrix [e]	185	7	7.9×10^{-5}	0.064	10 (7+3)	1.1×10^{-5}	0.015	*ADAMTS9, APOE, COL15A1, COL8A1, HTRA1, TIMP3, TNXB,* ***MMP19, PCOLCE, VTN***
	Others								
	Receptor-Mediated Endocytosis [f]	126	8	1.0×10^{-6}	0.0020	9 (8+1)	4.0×10^{-6}	0.0064	*APOE, CETP, CFI, DAB2, DMBT1, HSPH1, LOXL2, MICALL1,* ***VTN***
	Positive Regulation of Neuroblast Proliferation [g]	20	3	1.4×10^{-4}	0.10	4 (3+1)	2.2×10^{-5}	0.029	*ASPM, SOX10, VEGFA,* ***ZNF335***
	Endodermal Cell Differentiation [h]	27	1	n/a [b]	n/a [b]	4 (1+3)	4.0×10^{-5}	0.050	*COL8A1,* ***ITGA7, MMP9, VTN***
KEGG	Extracellular Matrix Pathways								
	Focal Adhesion [j]	197	3	0.21	0.89	9 (3+6)	3.0×10^{-4}	0.016	*TNXB, VAV1, VEGFA,* ***ACTG1, BCAR1, COL4A4, ITGA7, MYL2, VTN***

[a] GO:0030449, [b] GO:0034364, [c] GO:0043691, [d] GO:0005548, [f] GO:0031012, [g] GO:0006898, [h] GO:0002052, [j] GO:0035987

Supplementary Table 11: Biology of the 16 genes with highest priority per novel AMD locus. For the 16 genes with top gene score priority (see **Fig. 2A**) in 15 novel AMD loci (one novel locus with two genes of equal priority, one without any gene), manual literature search reveals interesting biology (evidence for relevant mouse phenotype in bold), except for *C20orf85* and *SRPK2*.

Gene Name	Description
COL4A3	Type 4 collagen, mutations in which cause a form of Alport's Syndrome, a basement membrane disorder. Approximately 85% of patients with Alport's also report a flecked retina, likely due to changes in the pigmentation of Bruchs' membrane. **Knockout mice reported for abnormal retinal morphology and aging (MGI:104688).**
SPEF2	Sperm flagellar protein 2 plays an important role in sperm motility and fertility. Chemically induced ENU mutant mice show aberrations in the immune, craniofacial, mortality/aging, nervous, reproductive, respiratory and skeletal systems, and this mouse serves as a model system for primary ciliary dyskinesia.
SRPK2	No link to eye phenotype for the Protein Kinase, serine/arginine-specific, a gene involved in pre-mRNA splicing, a critical step in the posttranscriptional regulation of gene expression.
PILRB	Paired immunoglobulin-like type 2 receptor β, related to the Siglec family of receptors, and plays a key activating role in immune function and inflammation. Gene expression increases with age.
TRPM3	Transient receptor potential cation channel, subfamily M, receptor 3, is a heat sensor in sensory neurons. Mutations in this gene cause inherited cataract and Glaucoma. ***TRPM3* and *TRPM1* mouse models show phenotypes associated with pupillary light response[37]**.
ABCA1	ATP Binding Cassette, Subfamily A, Member 1 is an ABC transporter that acts together with APOA1 in HDL efflux. It is involved in A2E catabolism in photoreceptors and was reported for association with intra-ocular pressure and open Glaucoma[38-40]. Mutations in the gene cause Tangier Disease (OMIM: 205400), an accumulation of cholesterol in tissues due to reduced plasma HDL levels. *ABCA1* is expressed in the macular ganglion cell layer, in the outer plexiform layer, and in RPE.
ARHGAP21	Rho GTPase activating protein 21 regulates the small GTPase CDC42 and is involved in golgi complex organization. May play an important role in retinal cilia, with retinopathy being a recurrent occurrence in ciliopathies.
MMP19	Matrix metalloproteinase 19 is a member of the endopeptidase family, which is involved in extracellular matrix degradation. Extracellular matrix turnover is mediated through a careful balance of tissue inhibitors of metalloproteinases (TIMPs) and MMPs. MMP19 has an important role in angiogenesis.
RDH5	The Retinol dehydrogenase 5 microsomal enzyme is abundant in RPE, where it has been proposed to catalyze the conversion of 11-cis retinol to 11-cis retinal[41] and involve RPE and Muller cells of retina. Mutations in this gene cause fundus albipunctatus[42], a rare form of night blindness characterized by a delay in generating cone and rod photopigments (NCBI). RDH5 is a drug target for NADH discussed as therapy for Parkinson's and Alzheimer disease. **An *RDH5* mouse model (MGI:1201412) shows impaired dark adaptation[43], is a disease model for fundus albipunctatus, and causes AMD-like pathology[44].**
PTPN11	The Protein-Tyrosin Phosphatase, Non-receptor11 gene exhibits numerous mutations causing the Noonan syndrome, a developmental disorder affecting multiple organs. *PTPN11* is widely expressed in most tissues and plays a regulatory role in various cell functions. **A *PTPN11* mouse model (MGI: 99511) shows a phenotype associated with the cardiovascular, immune, and hematopoietic system and mortality/aging.**
BCAR1	Breast Cancer Antiestrogen Resistance 1 is an Src family kinase substrate involved in various cellular events, including migration, survival, transformation, and invasion[45]. Insertion in *BCAR1* resulted in the tamoxifen-resistance for breast cancer therapy. **A *BCAR1* mouse model affects the formation of the retinal ganglion cell layer[46].**

VTN	Vitronectin is a tyrosine sulphated protein present in the RPE, among other cell types in the retina. It plays an important function in complement regulation, being an inhibitor of the complement cascade. Vitronectin and complement proteins are major components of drusen.
TSPAN10	Tetraspanin 10, also known as oculospanin, is expressed in the RPE, iris, ciliary body and retinal ganglion cells, but its function is not well described.
GPX4	Glutathione peroxidase 4 reduces phospholipid hydroperoxides within membranes and lipoproteins. Lipid peroxidation is implicated in inflammation and atherogenesis. Transgenic mice overexpressing human GPX4 show reduced oxidative injury after oxidative stress.
MMP9	Matrix metalloproteinase 9 is also a member of the endopeptidase family, and has an important role in extracellular matrix remodeling in angiogenesis. **Double knockouts of *MMP2* and *MMP9* show an important role in CNV[47], and provides a disease model for complement factor I deficiency and inflammatory/immunologic defect.**
C20orf85	Chromosome 20 open reading frame 85 is uncharacterized, has Gene Priority Score 0, but only gene in this region.

RPE: Retinal pigment epithelium

Supplementary Table 12: Four of the 34 lead variants show difference between advanced AMD subtypes. We compared the two advanced AMD subtypes, CNV and GA. Stated are Odds Ratios (OR) and P-values analyzing 10,749 CNV patients and 17,832 control subjects, 3,235 patients with GA compared to the same 17,832 control subjects, as well as the P-values testing the 10,749 CNV patients against the 3,235 GA patients. Four variants show a significant difference between disease subtypes (P_{diff} < 0.05/34 = 0.00147; rs3750846, rs5817082, rs14245000, rs5754227 in the loci *ARMS2/HTRA1*, *CETP*, *MMP9*, *SYN3/TIMP3*, marked bold). ORs are given into the direction of the minor allele in controls subjects.

Locus Name	Lead variant	Major/ minor allele	MAF			CNV vs Co.[a]		GA vs. Co[b]		CNV vs GA[c]
			CNV	GA	Co	OR	P	OR	P	P_{diff}
CFH	rs10922109	C/A	0.228	0.217	0.426	0.39	$4.6x10^{-457}$	0.37	$3.7x10^{-220}$	0.22
COL4A3	rs11884770	C/T	0.258	0.262	0.278	0.90	2.6×10^{-7}	0.93	1.9×10^{-2}	0.26
ADAMTS9-AS2	rs62247658	T/C	0.469	0.450	0.433	1.14	4.5×10^{-12}	1.09	2.1×10^{-3}	0.12
COL8A1	rs140647181	T/C	0.023	0.019	0.016	1.64	7.2×10^{-11}	1.32	2.0×10^{-2}	0.052
CFI	rs10033900	C/T	0.508	0.508	0.477	1.14	7.1×10^{-12}	1.13	7.2×10^{-6}	0.84
C9	rs62358361	G/T	0.016	0.015	0.009	1.73	3.1×10^{-10}	2.06	1.2×10^{-9}	0.10
SPEF2	rs114092250	G/A	0.015	0.018	0.022	0.69	6.8×10^{-7}	0.66	2.2×10^{-4}	0.76
C2/CFB/SKIV2L	rs116503776	G/A	0.092	0.092	0.148	0.59	1.4×10^{-72}	0.58	5.9×10^{-34}	0.90
VEGFA	rs943080	T/C	0.465	0.463	0.497	0.88	2.0×10^{-11}	0.87	7.0×10^{-7}	0.65
KMT2E/SRPK2	rs1142	C/T	0.211	0.207	0.346	1.13	1.4×10^{-9}	1.07	2.2×10^{-2}	0.12
PILRB/PILRA	rs7803454	C/T	0.373	0.362	0.190	1.14	4.1×10^{-8}	1.11	1.8×10^{-3}	0.62
TNFRSF10A	rs79037040	T/G	0.449	0.459	0.479	0.89	1.8×10^{-9}	0.92	2.0×10^{-3}	0.39
TRPM3	rs71507014	GC/G	0.429	0.421	0.405	1.11	8.3×10^{-8}	1.07	1.8×10^{-2}	0.29
MIR6130/RORB	rs10781182	G/T	0.330	0.319	0.306	1.12	5.5×10^{-8}	1.07	2.5×10^{-2}	0.19
TGFBR1	rs1626340	G/A	0.187	0.199	0.209	0.87	1.7×10^{-9}	0.94	5.4×10^{-2}	0.037
ABCA1	rs2740488	A/C	0.257	0.252	0.275	0.90	2.4×10^{-6}	0.89	1.9×10^{-4}	0.58
ARHGAP21	rs12357257	G/A	0.244	0.239	0.223	1.12	3.4×10^{-7}	1.11	2.4×10^{-3}	0.71
ARMS2/HTRA1	**rs3750846**	**T/C**	**0.447**	**0.384**	**0.208**	**2.95**	**$3.1x10^{-632}$**	**2.33**	**$3.6x10^{-173}$**	**1.0×10^{-17}**
RDH5/CD63	rs3138141	C/A	0.221	0.223	0.207	1.16	5.2×10^{-7}	1.16	4.1×10^{-4}	0.88
ACAD10	rs61941274	G/A	0.024	0.023	0.018	1.57	2.3×10^{-9}	1.42	1.8×10^{-3}	0.35
B3GALTL	rs9564692	C/T	0.279	0.267	0.299	0.90	2.5×10^{-7}	0.86	2.1×10^{-6}	0.23
RAD51B	rs61985136	T/C	0.361	0.361	0.384	0.89	1.0×10^{-8}	0.91	1.8×10^{-3}	0.37
LIPC	rs2043085	T/C	0.353	0.356	0.381	0.87	2.7×10^{-12}	0.89	2.7×10^{-5}	0.54
CETP	**rs5817082**	**C/CA**	**0.228**	**0.250**	**0.264**	**0.82**	**6.0×10^{-19}**	**0.93**	**2.3×10^{-2}**	**3.2×10^{-4}**
CTRB2/CTRB1	rs72802342	C/A	0.065	0.071	0.080	0.77	1.3×10^{-11}	0.85	4.4×10^{-3}	0.083
TMEM97/VTN	rs11080055	C/A	0.459	0.472	0.486	0.90	8.8×10^{-9}	0.94	3.3×10^{-2}	0.099
NPLOC4/TSPAN 10	rs6565597	C/T	0.404	0.385	0.381	1.15	2.1×10^{-11}	1.05	0.10	0.0080
C3	rs2230199	C/G	0.465	0.453	0.208	1.39	7.3×10^{-47}	1.53	7.2×10^{-37}	0.0047
CNN2	rs67538026	C/T	0.262	0.279	0.498	0.91	7.3×10^{-6}	0.90	8.2×10^{-4}	0.74
APOE	rs429358	T/C	0.102	0.099	0.135	0.73	7.2×10^{-28}	0.69	2.4×10^{-16}	0.38
MMP9	**rs142450006**	**TTTTC /T**	**0.116**	**0.144**	**0.141**	**0.78**	**8.4×10^{-17}**	**1.04**	**0.39**	**4.1×10^{-10}**
C20orf85	rs201459901	T/TA	0.054	0.057	0.070	0.74	2.5×10^{-14}	0.82	5.4×10^{-4}	0.12
SYN3/TIMP3	**rs5754227**	**T/C**	**0.106**	**0.121**	**0.137**	**0.75**	**2.5×10^{-24}**	**0.88**	**3.4×10^{-3}**	**3.8×10^{-4}**
SLC16A8	rs8135665	C/T	0.216	0.217	0.195	1.13	5.1×10^{-8}	1.14	9.3×10^{-5}	0.87

MAF = minor allele frequency, Co = control subjects; [a] Association in CNV versus controls; [b] Association in GA versus controls; [c] Association in CNV versus GA

Supplementary Table 13: Many but not all of the 34 lead variants show association with intermediate AMD. We compared the intermediate stage AMD patients with the advanced AMD patients. Stated are Odds Ratios (into the direction of the minor allele in controls) and P-values for the association of the 34 lead variants analysing 6,657 subjects with intermediate AMD and 17,832 control subjects of European ancestry, the re-stated Odds Ratios and P-value for advanced AMD and the P-value testing for difference of the variant association testing the 6,657 intermediate versus the 16,144 advanced AMD patients (P_{diff}).

Locus name	Lead variant	Major/ minor allele[a]	Allele frequency of control minor allele			Int. AMD vs. Co.[b]		Adv. AMD vs. Co.[c]		Int. vs. Adv. AMD[d]
			Int. AMD	Adv. AMD	Co	OR	P	OR	P	P_{diff}
CFH	rs10922109	C/A	0.315	0.223	0.426	0.62	$1.3x10^{-109}$	0.38	$9.6x10^{-618}$	1.36×10^{-84}
COL4A3	rs11884770	C/T	0.263	0.258	0.278	0.92	5.8×10^{-4}	0.90	2.9×10^{-8}	0.39
ADAMTS9/AS2	rs62247658	T/C	0.444	0.466	0.433	1.03	0.13	1.14	1.8×10^{-14}	1.0×10^{-5}
COL8A1	rs140647181	T/C	0.017	0.023	0.016	1.11	0.28	1.59	1.4×10^{-11}	3.1×10^{-5}
CFI	rs10033900	C/T	0.495	0.511	0.477	1.12	5.6×10^{-8}	1.15	5.4×10^{-17}	0.25
C9	rs62358361	G/T	0.013	0.016	0.009	1.52	2.8×10^{-5}	1.80	1.3×10^{-14}	0.058
PRLR/SPEF2	rs114092250	G/A	0.021	0.016	0.022	0.95	0.48	0.70	2.1×10^{-8}	3.1×10^{-4}
C2/CFB/SKIV2L	rs116503776	G/A	0.113	0.090	0.148	0.74	3.2×10^{-21}	0.57	$1.2x10^{-103}$	9.4×10^{-14}
VEGFA	rs943080	T/C	0.478	0.465	0.497	0.93	3.2×10^{-4}	0.88	1.1×10^{-14}	0.011
KMT2E/SRPK2	rs1142	C/T	0.358	0.370	0.346	1.06	8.5×10^{-3}	1.11	1.4×10^{-9}	0.031
PILRB/PILRA	rs7803454	C/T	0.200	0.209	0.190	1.06	0.020	1.13	4.8×10^{-9}	0.026
TNFRSF10A	rs79037040	T/G	0.458	0.451	0.479	0.92	1.2×10^{-4}	0.90	4.5×10^{-11}	0.16
TRPM3	rs71507014	GC/G	0.417	0.427	0.405	1.05	0.016	1.10	3.0×10^{-8}	0.046
MIR6130/RORB	rs10781182	G/T	0.312	0.328	0.306	1.03	0.26	1.11	2.6×10^{-9}	6.8×10^{-4}
TGFBR1	rs1626340	G/A	0.211	0.189	0.209	1.01	0.65	0.88	3.8×10^{-10}	2.2×10^{-7}
ABCA1	rs2740488	A/C	0.257	0.255	0.275	0.91	5.9×10^{-5}	0.90	1.2×10^{-8}	0.71
ARHGAP21	rs12357257	G/A	0.235	0.243	0.223	1.06	0.017	1.11	4.4×10^{-8}	0.061
ARMS2/HTRA1	rs3750846	T/C	0.293	0.436	0.208	1.57	4.1×10^{-80}	2.81	$6.5x10^{-735}$	$5.8x10^{-158}$
RDH5/CD63	rs3138141	C/A	0.213	0.222	0.207	1.06	0.070	1.16	4.3×10^{-9}	7.0×10^{-3}
ACAD10	rs61941274	G/A	0.019	0.024	0.018	1.12	0.20	1.51	1.1×10^{-9}	7.4×10^{-4}
B3GALTL	rs9564692	C/T	0.285	0.277	0.299	0.93	7.1×10^{-4}	0.89	3.3×10^{-10}	0.15
RAD51B	rs61985136	T/C	0.372	0.360	0.384	0.94	6.5×10^{-3}	0.90	1.6×10^{-10}	0.026
LIPC	rs2043085	T/C	0.366	0.350	0.381	0.93	$4. \times 10^{-4}$	0.87	4.3×10^{-15}	0.018
CETP	rs5817082	C/CA	0.235	0.232	0.264	0.86	1.7×10^{-10}	0.84	3.6×10^{-19}	0.48
CTRB2/CTRB1	rs72802342	C/A	0.079	0.067	0.080	0.98	0.63	0.79	5.0×10^{-12}	2.3×10^{-6}
TMEM97/VTN	rs11080055	C/A	0.490	0.463	0.486	1.02	0.42	0.91	1.0×10^{-8}	2.1×10^{-7}
NPLOC4/TSPAN10	rs6565597	C/T	0.391	0.400	0.381	1.05	0.038	1.13	1.5×10^{-11}	1.6×10^{-3}
C3	rs2230199	C/G	0.244	0.266	0.208	1.26	6.3×10^{-19}	1.43	3.8×10^{-69}	5.5×10^{-7}
CNN2	rs67538026	C/T	0.487	0.460	0.498	0.93	1.7×10^{-3}	0.90	2.6×10^{-8}	0.23
APOE	rs429358	T/C	0.113	0.099	0.135	0.82	1.7×10^{-10}	0.70	2.4×10^{-42}	2.4×10^{-5}
MMP9	rs142450006	TTTTC/T	0.141	0.124	0.141	1.00	0.88	0.85	2.4×10^{-10}	5.7×10^{-7}
C20orf85	rs201459901	T/TA	0.064	0.054	0.070	0.89	6.6×10^{-3}	0.76	3.1×10^{-16}	1.3×10^{-4}
SYN3/TIMP3	rs5754227	T/C	0.128	0.109	0.137	0.92	5.4×10^{-3}	0.77	1.1×10^{-24}	1.7×10^{-7}
SLC16A8	rs8135665	C/T	0.212	0.217	0.195	1.11	5.0×10^{-5}	1.14	5.5×10^{-11}	0.27

Co. = control subjects; OR = Odds Ratio; Int. AMD = intermediate AMD; Adv. AMD = advanced AMD
[a] Major and minor allele among controls [b] Association of intermediate AMD versus controls, [c] Association of advanced AMD versus controls, [d] Testing intermediate AMD versus advanced AMD.

Supplementary Table 14: Statistical power to detect intermediate AMD association. We computed the power to detect association in the analysis of 6,657 subjects with intermediate AMD compared to 17,832 controls of European ancestry for selected effect sizes (Odds Ratios) and frequencies that spanned the effect sizes and frequencies of the 34 lead variants. Also, different significance thresholds were assumed: $\alpha = 0.05$ (nominal significance), $0.05/34 = 0.0015$ (corrected for testing 34 variants), or genome-wide significance (5×10^{-8}). Since all common variant signals had a MAF between 10% and 50% with advanced AMD odds ratios > 1.1, and the rare variant signals (~1%, *C9, PRLR/SPEF2, ADAC10, C20orf85*) had ORs of 1.3 and higher, the power was ~80% to detect the signal also for intermediate AMD – if present - for all 34 lead variants at the nominal (5%) significance level.

		Significance level α		
MAF*	**Odds Ratio**	**0.05**	**0.0015**	**5×10^{-8}**
1%	1.10	16%	1%	0%
1%	1.30	78%	33%	0%
1%	1.50	99%	89%	14%
5%	1.10	55%	14%	0%
5%	1.30	100%	100%	70%
5%	1.50	100%	100%	100%
10%	1.10	82%	38%	0%
10%	1.30	100%	100%	100%
10%	1.50	100%	100%	100%
30%	1.10	99%	88%	13%
30%	1.30	100%	100%	100%
30%	1.50	100%	100%	100%
50%	1.10	100%	93%	22%
50%	1.30	100%	100%	100%
50%	1.50	100%	100%	100%

*MAF = minor allele frequency in controls

Supplementary Table 15: Enriched pathways for genes in the 10 AMD loci not associated with intermediate AMD versus the 24 loci associated with intermediate AMD. As before (see **Supplementary Table 10**), we queried pathway data bases (Reactome, GO, KEGG) using INRICH, obtaining P-values based on 50,000 Bootstrap rounds and 1,000,000 replication runs and 8,957,338 common variants (MAF ≥ 1% in cases and controls combined) and matching selected target regions in terms of gene count, variant density and total number of variants[76]. Only gene sets with ≥ 3 genes were analyzed. Shown are the enriched pathways (corrected P<0.10) and the genes contributing to the enrichment **A** for the 10 narrow AMD locus regions identified by the 11 of the 52 variants (variants and proxies, r^2≥0.5, ±100kb) that were not nominally associated with intermediate AMD ($P_{intermediate}$ > 0.05), and **B** the 24 narrow AMD locus regions identified by the 41 of the 52 that were nominally associated with intermediate AMD.

A for the 10 loci not associated with intermediate AMD

Database	Pathway	#genes in gene set	#AMD loci in gene set	Empirical P	Corrected P	Genes contributing to enrichment
Reactome	Collagen Pathways					
	Assembly of Collagen Fibrils and Other Multimeric Structures	54	3	2.6×10^{-5}	0.0034	*COL15A1, COL8A1, MMP9*
	Collagen Degradation	37	3	3.2×10^{-5}	0.0041	*COL15A1, MMP19, MMP9*
	Collagen Formation	84	3	2.2×10^{-4}	0.022	*COL15A1, COL8A1, MMP9*
	Extracellular Matrix Pathways					
	Degradation of the Extracellular Matrix	88	5	1.0×10^{-6}	4.0×10^{4}	*ADAMTS9, COL15A1, CTRB1/CTRB2, MMP19, MMP9*
GO	Collagen Pathway					
	Collagen Catabolic Process (GO:0030574)	73	4	1.0×10^{-6}	0.0014	*COL15A1, COL8A1, MMP19, MMP9*
	Extracellular Matrix Pathways					
	Extracellular Matrix Disassembly (GO:0022617)	116	5	4.0×10^{-6}	0.0037	*COL15A1, COL8A1, CTRB1/CTRB2, MMP19, MMP9*
	Extracellular Matrix (GO:0031012)	185	5	2.8×10^{-5}	0.017	*ADAMTS9, COL15A1, COL8A1, MMP19, VTN*
	Others					
	Endodermal Cell Differentiation (GO:0035987)	27	4	1.0×10^{-6}	0.0014	*COL8A1, ITGA7, MMP9, VTN*
KEGG	Extracellular Matrix pathways					
	Focal Adhesion	197	4	0.00336	0.080	*BCAR1, ITGA7, MYL2, VTN*

B for the 24 AMD loci associated with intermediate AMD

Database	Pathway	#genes in gene set	#AMD loci in gene set	Empirical P	Corrected P	Genes contributing to enrichment
Reactome	Complement Pathways					
	Regulation of Complement Cascade	22	5	2.0×10^{-6}	5.6×10^{-4}	*C2 /C4A/C4B/CFB, C3, C9, CFH/CFHR3, CFI*
	Complement Cascade	36	5	1.4×10^{-5}	0.0028	*C2 /C4A/C4B/CFB, C3, C9, CFH/CFHR3, CFI*
	Lipid Pathways					
	Lipoprotein Metabolism	30	4	1.6×10^{-4}	0.025	*ABCA1, APOC2/APOE, CETP, LIPC*
	Lipid Digestion, Mobilization, and Transport	50	4	4.9×10^{-4}	0.068	*ABCA1, APOC2/APOE, CETP, LIPC*
GO	Complement Pathways					
	Regulation of Complement Activation (GO:0030449)	24	5	6.0×10^{-6}	0.0083	*C2 /C4A/C4B/CFB, C3, C9, CFH, CFI*
	Complement Activation, Alternative Pathway (GO:0006957)	13	4	6.2×10^{-5}	0.059	*C3, C9, CFB, CFH/CFHR5*
	Lipid Pathways					
	High-Density Lipoprotein Particle (GO:0034364)	22	5	1.3×10^{-5}	0.016	*ABCA1, APOE/APOC1/APOC4, APOM, CETP, LIPC*
	Reverse Cholesterol Transport (GO:0043691)	18	5	1.4×10^{-5}	0.017	*ABCA1, APOE/APOC2, APOM, CETP, LIPC*
	Phospholipid Transporter Activity (GO:0005548)	7	3	5.5×10^{-5}	0.054	*ABCA1, ABCA7, CETP*
	High-Density Lipoprotein Particle Assembly (GO:0034380)	8	4	9.3×10^{-5}	0.084	*ABCA1, ABCA7, APOE, APOM*
	Others					
	Receptor-Mediated Endocytosis (GO:0006898)	126	8	1.0×10^{-6}	0.0025	*APOE, CETP, CFI, DAB2, DMBT1, HSPH1, LOXL2, MICALL1*
KEGG	Complement Pathways					
	Complement and Coagulation Cascades	68	5	7.6×10^{-4}	0.039	*C2/C4A/C4B, C3, C9, CFB, CFH/F13B, CFI*

Supplementary Table 16: Relative and absolute risk for advanced AMD by genetic risk score deciles based on the 52 signals. We derived the weighted genetic risk score (GRS) based on the 52 independent index variants (using the fully conditioned log Odds Ratio per variant as weight) and obtained a realistic GRS distribution by modeling a general population with different prevalence assumptions using different weights for advanced AMD cases and controls. Shown here are the GRS intervals assuming 5% prevalence (weight for cases = 0.105, for controls = 1.81; see **Material and Methods**). We computed approximate relative risk (Odds Ratio) for advanced AMD per GRS decile with the lowest decile as reference and absolute risk of advanced AMD (= % affected, again utilizing the weighted counts).

GRS decile	GRS interval	Observed				Weighted		
		# cases	# Controls	% affected	Odds ratio	# cases	# controls	% affected
1	[-3.829;-1.439]	214	1,865	10.3%	reference	22.5	3,375.8	0.7%
2	(-1.439;-0.957]	270	1,861	12.7%	1.3	28.4	3,368.5	0.8%
3	(-0.957;-0.590]	364	1,856	16.4%	1.7	38.3	3,359.5	1.1%
4	(-0.589;-0.270]	529	1,847	22.3%	2.5	55.7	3,343.2	1.6%
5	(-0.270;0.032]	664	1,838	26.5%	3.2	69.9	3,326.9	2.1%
6	(0.032;0.331]	891	1,825	32.8%	4.2	93.8	3,303.4	2.8%
7	(0.332;0.672]	1,266	1,804	41.2%	6.1	133.2	3,265.4	3.9%
8	(0.672;1.077]	1,813	1,771	50.6%	8.9	190.8	3,205.6	5.6%
9	(1.077;1.635]	2,796	1,715	62.0%	14.0	294.2	3,104.3	8.7%
10	(1.635;7.186]	7,337	1,450	83.5%	44.0	772.1	2,624.6	22.7%

GRS = genetic risk score; Adv. AMD = advanced AMD

Supplementary Table 17: Heterogeneity of risk variants previously published. We repeated the analysis of our joint data set by meta-analysing study-specific statistics, computed the I^2 measure of heterogeneity[48], and compared these results for the 19 index variants from the previous study[11]. Stated are the variants and their I^2 measures in the previous and the current study, which indicates that heterogeneity of the current study tended to lower than that of the previous study (median across the 19 variants being 12% versus 23%, respectively).

Locus name	Chr[(c)]	Previous Study		r^2[(f)]	Current Study	
		dbSNP ID[(d)]	I^2 (%)[(e)]		dbSNP ID[(g)]	I^2 (%)[(h)]
CFH	1	rs10737680	84%	-	rs10737680	59%
ADAMTS9-AS2	3	rs6795735	23%	-	rs6795735	47%
COL8A1/FILIP1L	3	rs13081855	0%	-	rs13081855	29%
CFI	4	rs4698775	46%	-	rs4698775	0%
C2-CFB	6	rs429608	56%	-	rs429608	46%
VEGFA	6	rs943080	34%	-	rs943080	12%
COL10A1 [(b)]	6	rs3812111	34%	-	rs3812111	0%
IER3/DDR1 [(c)]	6	rs3130783	0%	-	rs3130783	27%
TNFRSF10A	8	rs13278062	0%	-	rs13278062	0%
TGFBR1	9	rs334353	0%	-	rs334353	0%
ARMS2/HTRA1	10	rs10490924	77%	-	rs10490924	64%
B3GALTL	13	rs9542236	57%	-	rs9542236	23%
RAD51B	14	rs8017304	6%	-	rs8017304	22%
LIPC	15	rs920915	0%	-	rs920915	0%
CETP	16	rs1864163	0%	-	rs1864163	0%
C3	19	rs2230199	25%	-	rs2230199	0%
APOE	19	rs4420638	9%	-	rs4420638	22%
TIMP3	22	rs5749482	0%	0.82	rs5754227[(a)]	0%
SLC16A8	22	rs8135665	39%	-	rs8135665	30%

Chr = Chromosome; (a) proxy variant used; (b) *COL10A1* locus was not genome-wide significant in current study (rs3812111: $P=1.2 \times 10^{-4}$), (c) *IER3/DDR1* was merged with C2/CFB in current study because the signal disappeared after adjusting for the C2/CFB index variant, (d) dbSNP identifier of previous study lead variant, (e) I^2 in European subjects of discovery cohort of the previous study, (f) correlation coefficient, r^2, between previous lead variant and proxy variant in the current study, if applicable, (g) dbSNP identifier of (proxy) variant in current study, (h) I^2 of (proxy) variant in current study.

Supplementary Table 18: Heterogeneity of the 34 identified lead variants. I^2 as measure of heterogeneity and the P-value testing for significant heterogeneity based on the Q statistics were computed. Results of a random effects model are provided, which show very similar Odds Ratio estimates as the our original analysis and a tendency towards larger P-values as expected given the known lower power of a random effects model.

Locus name	Lead variant	European major/ minor allele	Heterogeneity		Random effects model		Original analysis
			I^2	P	P	Odds Ratio	Odds Ratio
KNOWN (previously reported with genome-wide significance, $P < 5x10^{-8}$)							
CFH	rs10922109	C/A	59%	$1.20x10^{-4}$	$2.21x10^{-175}$	0.40	0.38
ADAMTS9/AS2	rs62247658	T/C	47%	$5.27x10^{-3}$	$1.05x10^{-06}$	1.13	1.14
COL8A1	rs140647181	T/C	0%	0.74	$4.90x10^{-08}$	1.49	1.59
CFI	rs10033900	C/T	0%	0.55	$3.02x10^{-13}$	1.13	1.15
C9	rs62358361	G/T	0%	0.47	$2.18x10^{-09}$	1.64	1.80
C2/CFB/SKIV2L	rs116503776	G/A	46%	$6.53x10^{-3}$	$2.41x10^{-47}$	0.57	0.57
VEGFA	rs943080	T/C	0%	0.63	$2.29x10^{-13}$	0.88	0.88
TNFRSF10A	rs79037040	T/G	0%	0.85	$1.56x10^{-10}$	0.90	0.90
TGFBR1	rs1626340	G/A	0%	0.72	$7.95x10^{-11}$	0.87	0.88
ARMS2/HTRA1	rs3750846	T/C	64%	$8.30x10^{-6}$	$2.17x10^{-189}$	2.75	2.81
B3GALTL	rs9564692	C/T	23%	0.15	$5.77x10^{-07}$	0.89	0.89
RAD51B	rs61985136	T/C	22%	0.16	$8.43\ x10^{-06}$	0.91	0.90
LIPC	rs2043085	T/C	0%	0.74	$7.54x10^{-12}$	0.88	0.87
CETP	rs5817082	C/CA	0%	0.66	$4.00x10^{-18}$	0.84	0.84
C3	rs2230199	C/G	0%	0.90	$3.18x10^{-50}$	1.37	1.43
APOE	rs429358	T/C	22%	0.15	$3.91x10^{-24}$	0.72	0.70
SYN3/TIMP3	rs5754227	T/C	0%	0.69	$1.17x10^{-22}$	0.77	0.77
SLC16A8	rs8135665	C/T	30%	0.08	$1.27x10^{-06}$	1.14	1.14
NOVEL (reported with genome-wide significance, $P < 5x10^{-8}$, for the first time)							
COL4A3	rs11884770	C/T	13%	0.27	$2.89x10^{-06}$	0.90	0.90
PRLR/SPEF2	rs114092250	G/A	0%	0.81	$4.13\ x10^{-05}$	0.75	0.70
PILRB/PILRA	rs7803454	C/T	11%	0.31	$3.27\ x10^{-08}$	1.14	1.13
KMT2E/SRPK2	rs1142	C/T	19%	0.19	$7.40\ x10^{-06}$	1.10	1.11
TRPM3	rs71507014	GC/G	2%	0.44	$2.55\ x10^{-06}$	1.09	1.10
MIR6130/RORB	rs10781182	G/T	0%	0.94	$1.26\ x10^{-06}$	1.09	1.11
ABCA1	rs2740488	A/C	51%	$1.74x10^{-3}$	$3.16\ x10^{-06}$	0.87	0.90
ARHGAP21	rs12357257	G/A	11%	0.31	$6.05\ x10^{-06}$	1.11	1.11
RDH5/CD63	rs3138141	C/A	0%	0.54	$5.24\ x10^{-08}$	1.16	1.16
ACAD10	rs61941274	G/A	28%	0.10	$3.61x10^{-05}$	1.46	1.51
CTRB2/CTRB1	rs72802342	C/A	0%	0.52	$8.12\ x10^{-08}$	0.83	0.79
TMEM97/VTN	rs11080055	C/A	0%	0.93	$3.19\ x10^{-08}$	0.91	0.91
NPLOC4/TSPAN10	rs6565597	C/T	9%	0.34	$2.84\ x10^{-08}$	1.12	1.13
CNN2	rs67538026	C/T	0%	0.86	$3.66\ x10^{-06}$	0.91	0.90
MMP9	rs142450006	TTTTC/T	0%	0.73	$1.29\ x10^{-10}$	0.84	0.85
C20orf85	rs201459901	T/TA	0%	0.71	$2.88\ x10^{-14}$	0.76	0.76

Supplementary Table 19: Sensitivity analyses results for the rare variants burden test with locus-wide conditioning. We repeated the rare variant burden analysis (a) excluding the overlap of our association analysis data set with the previously sequenced subjects (excluding 858 advanced AMD cases and 650 control persons, thus comparing 15,286 advanced AMD subjects and 17,182 controls) and (b) prioritizing variants with high functionality by restricting to variants with a CADD score ≥ 20 using the original 16,144 advanced AMD patients and 17,832 control subjects of European ancestry.

Gene	# variants in analysis	Locus-wide conditioned			
		Optimal RAC	# Variants with MAC ≤ optimal RAC	Effect direction[a]	*P*
Sensitivity analysis excluding previously sequenced subjects					
CFI	47	44	43	+	1.0×10^{-8}
TIMP3	10	14	9	+	4.0×10^{-8}
CFH	49	8	35	+	2.8×10^{-7}
SLC16A8	9	615	9	+	3.5×10^{-6}
Sensitivity analysis restricting to variants with CADD ≥ 20					
CFI	12	46	12	+	2.0×10^{-8}
TIMP3	3	14	3	+	3.5×10^{-6}
SERPINA1 [b]	6	1,055	5	-	3.4×10^{-7}
CFH	1 [c]	NA	NA	NA	NA
SLC16A8	7	648	7	+	1.5×10^{-6}

RAC: risk allele count; MAC: minor allele count; MAF = minor allele frequency; NA = not available

[a] "+" and "-" indicates AMD risk increase or decrease by the accumulated rare variants

[b] Outside of 34 risk loci, and thus no locus wide conditioning

[c] Burden test results not available for CFH, only genes with > 1 variant with CADD ≥ 20 were analyzed.

Supplementary Table 20: Overview of information in gene priority score table. The number of genes per locus and the number of genes per locus with the respective characteristics (e.g. the number of genes with identified expression etc.) are provided.

Locus name	Chromosome	Total # genes in locus	Min Gene Priority Score	Max Gene Priority Score	Expression in retina/RPE/choroid	Eye phenotype in knock-out mouse	≥ 1 variant in 95% credible sets	Rare variant burden	Protein altering	5' UTR/3' UTR	Other exonic	Promoter region	Enriched molecular pathway	Drug target
CFH	1	12	1	7	6	2	12	1	4	2	2	1	3	0
COL4A3	2	3	1	4	3	1	2	0	0	0	0	0	2	0
ADAMTS9-AS2	3	3	2	2	2	0	3	0	0	0	0	1	0	0
COL8A1	3	7	0	3	3	0	5	0	0	0	0	0	1	0
CFI	4	7	0	5	6	1	3	1	1	0	0	0	1	0
C9	5	3	1	3	2	0	2	0	1	0	0	0	1	0
PRLR/SPEF2	5	6	0	3	4	0	1	0	0	0	0	1	0	1
C2/CFB/SKIV2L	6	110	0	7	68	7	54	0	2	3	4	3	5	7
VEGFA	6	1	4	4	1	1	0	0	0	0	0	0	1	1
KMT2E/SRPK2	7	7	1	3	3	0	5	0	0	1	2	0	0	0
PILRB/PILRA	7	23	0	5	18	0	15	0	1	5	5	7	1	0
TNFRSF10A	8	5	1	3	5	0	2	0	0	0	0	1	1	0
MIR6130/RORB	9	0	0	0	0	0	0	0	0	0	0	0	0	0
TRPM3	9	2	2	4	1	1	2	0	0	1	0	1	0	0
TGFBR1	9	4	1	3	4	0	3	0	0	1	0	0	1	0
ABCA1	9	1	4	4	1	0	1	0	0	0	0	0	1	1
ARHGAP21	10	3	1	3	3	0	2	0	0	1	0	0	0	0
ARMS2/HTRA1	10	5	0	3	2	1	4	0	1	0	0	0	0	0
RDH5/CD63	12	13	0	4	11	2	11	0	0	0	0	0	3	2
ACAD10	12	19	0	4	12	1	12	0	0	1	0	0	1	2
B3GALTL	13	2	1	2	2	0	1	0	0	0	0	0	0	0
RAD51B	14	1	2	2	1	0	1	0	0	0	0	0	0	0
LIPC	15	1	3	3	0	0	1	0	0	0	0	1	1	0
CETP	16	5	0	2	2	0	4	0	0	0	0	0	1	1
CTRB2/CTRB1	16	12	1	4	10	1	4	0	0	0	0	0	3	1
TMEM97/VTN	17	15	1	5	11	1	13	0	1	1	1	0	1	2
NPLOC4/TSPAN10	17	16	0	5	12	3	10	0	1	1	0	1	1	4
C3	19	18	1	6	16	2	12	0	3	0	1	0	2	2
CNN2	19	10	2	4	8	1	10	0	0	2	1	0	0	2
APOE	19	24	0	5	20	3	14	0	1	1	0	1	3	0
MMP9	20	10	1	5	10	2	5	0	1	0	0	1	2	1
C20orf85	20	1	0	0	0	0	0	0	0	0	0	0	0	0
SYN3/TIMP3	22	2	2	5	2	1	1	1	0	0	0	0	1	1
SLC16A8	22	17	0	5	13	1	4	1	0	0	0	0	1	3
Total		**368**	**33**	**127**	**262**	**32**	**219**	**4**	**17**	**20**	**16**	**19**	**38**	**31**

Supplementary Note 1: Comparison to previously published loci

Among the 21 previously reported loci, there are three that are not supported in our analysis. First, rs3130783 near *IER3/DDR1* lost evidence of association in our analysis after conditioning on variants near the *CFB/C2* genes (approximately 2 Mb distant; P = 2.31 x 10^{-7} before and P = 0.0134 after conditioning on rs116503776). This indicates that this locus was an artifact from linkage disequilibrium in the previous study that we were able to solve in our investigation by creating wider locus regions and by identifying independent variants through a sequential conditioning approach. Second, rs3812111 near *COL10A1* exhibited only relatively modest evidence for association in our data (P = 1.2 x 10^{-4}), without genome-wide significance. Third, rs1713985 around *REST/C4orf14/POLR2B/IGFBP7* on chromosome 4q12, which was previously associated with advanced AMD in Japanese individuals[49], was not associated in our data with advanced AMD (OR=1.05, p=0.081) in our European subjects.

Effect size estimates at our known loci variants were comparable with previous reports, except for *CFI* G119R (rs141853578) with our odds ratio of 3.64 being substantially smaller than the previously published odds ratio of 20.2, possibly pointing to a inflated effect size in the identifying study (winner's curse). When we applied the same meta-analysis approach as in our previous analysis that identified 19 lead variants[11,31], we found that between-study heterogeneity was smaller in our current data (median heterogeneity across the 19 variants = 12%) than in our previous data (median 23%) including a decrease from 84% to 60% for the top CFH variant (**Supplementary Table 17**).

For our 34 lead variants, we observed significant heterogeneity (P of Q statistics ≤ 0.05) for the loci *CHF*, *ARMS/HTRA*, *C2/CFB/SKIV2L*, and *ADAMTS9* (I^2 from 47% to 64%) as observed previously; for the novel loci, we find low heterogeneity, except for one locus (*ABCA1*, I^2 =51%, **Supplementary Table 18**). When applying a model accounting for between-study heterogeneity (random effects model[50]), we find very similar Odds Ratio estimates and a tendency towards large P-values, as expected due to the known lower power of the random effects model (**Supplementary Table 18**).

Supplementary Note 2: We evaluated interaction of each of the 52 variants with the four main lead variants (at the loci *ARMS2*, *CFH*, *C2*/*CFB* and *C3*) by logistic regression models including the respective variants and the two-way interaction term additionally to the other covariates (two principal components and DNA source). This corresponded to 198 (=51+50+49+48) pairwise interaction tests, since the 4 main lead variants were also among the 52 variants. Accounting for these 198 models tested (P< 0.05/198 = 2.52 x 10x-4) , two signals passed the Bonferroni significance threshold: an interaction between CFH: rs10922109 and CFH: rs61818925 (P for interaction = 1.8 x 10^{-29}), which could be the result

of LD or a haplotype effect, and an interaction between CFH: rs570618 with *C2/CFB/SKIV2L:* rs116503776 (P for interaction = 6.4×10^{-6}).

Supplementary Note 3: Sensitivity analyses for burden test

There was an overlap of our subjects that we had sequenced to detect rare variants with the subjects in our association analysis (see **Methods**). When we conducted a sensitivity analysis excluding the 650 controls and 858 advanced AMD cases in the overlap, we found virtually the same results without (data not shown) and with conditioning on the other identified variants (*CFH*, *CFI*, *TIMP3* and *SLC16A8*, **Supplementary Table 19**).

We were also interested whether focusing on functionality would improve the burden signal. We thus restricted the rare protein-altering variants in the variable threshold test to variants with CADD Score[36] ≥ 20; we found similar results for three of the four genes with significant burden conditioned on the other variants in the locus as in the original analysis (*CFI*, *SLC16A8*, and *TIMP3*). CFH was not included in these analyses, as there was no variant with CADD score ≥ 20. Additionally, we found a new gene with genome-wide significant burden in *SERPINA1* (**Supplementary Table 19**); this burden is in fact driven by a single missense variant, rs28928474.

Supplementary Note 4: Motivated by recent work on the short *CFH* isoform (factor H-like protein 1)[51], we dissected the variants in the *CFH* between those unique to the long isoform coding regions (21 variants) and the others (15 in the overlap, one unique to the short isoform) and found a predominant burden from the latter (OR = 13.5, $P = 9.9 \times 10^{-7}$, 50 rare alleles in cases, 4 in controls versus OR 1.6, P= 0.08, 38 rare alleles in cases, 34 in controls). It should be noted that this analysis eliminated the strong effect of R1210C (rs121913059), a variant unique to the long isoform, by conditioning the burden analyses on the eight *CFH* variants among the 52 identified variants.

Supplementary Note 5: The gene priority table is a summary of the evidence from expression, association, bio-informatics, pathway enrichment, and drugability. The score can be customized by each researcher depending on his/her own priorities or research questions by putting a specific weight to each column. A summary of the total number of genes per locus and the number of genes with the respective characteristics per locus is given in **Supplementary Table 20**.

Supplementary Note 6: One additional subtype-specific locus by CNV-specific genome-wide single variant association analysis

This signal specific to choroidal neovascular disease (CNV) near *MMP9* was markedly stronger in an analysis restricted to CNV cases compared to the primary genome-wide association analysis (P_{CNV} = 8.4 x 10^{-17} vs. P_{GWAS} = 2.4 x 10^{-10}) in-line with larger power of stratified analyses for signals only present in one stratum[52]. Encouraged by the possibility of additional subtype-specific loci, we repeated our genome-wide analysis stratified by disease subtype and identified one additional locus, near *LINC00461 (aka Eyelinc1)*[53], specific to neovascular disease (rs17421410, CAF = 6.92%, OR = 1.23, $P_{GWAS/CNV}$ = 5.0 x 10^{-9}) without association for GA (OR_{GA} = 1.03; $P_{GWAS/GA}$ = 0.59; P_{diff} = 1.2 x 10^{-3})[54,55]. This is a biologically interesting locus for CNV-specific genetic association due to previous association with retinal vessel caliber.

Supplementary File Legends:

Supplementary File 1: Locuszoom Plots for each of the 52 identified signals. Region plots showing single variant association P-values of variants around each of the 52 index variants. Shown are also the location/direction of underlying genes and the location of the variants in the 95% credible sets.

Supplementary File 2: Extended results of the 34 lead variants in Non-European subjects. We analysed the association of advanced AMD compared to control subjects in an Asian (473 cases, 1,099 controls), African (52 cases, 361 controls), and "other ancestry" group (254 cases, 694 controls) for our 34 lead variants. Shown are frequencies, Odds ratios, and P-values from the Firth-corrected logistic regression for all analyses.

Supplementary File 3: Variants in 95% credible sets and their annotation. For each of the 52 index variants, the 95% credible sets contain the minimal set of variants that add up to > 95% posterior probability.

Supplementary File 4: Details about the identified rare protein-altering variants in CFH, CFI, TIMP3 and SLC18A8 that we found to be enriched in AMD cases (see also Table 2). Here we show the variants of each gene below the optimal risk allele frequency that contributed to the observed significant burden.

Supplementary File 5: Genes in the 34 identified AMD locus regions. Stated are all genes that overlap with the 34 AMD locus regions (defined by the 52 identified variants, their proxies, $r^2 \geq 0.5$, ±500kb) as well as an indicator whether this gene was also among the 368 genes in the narrow AMD locus regions (defined by 52 identified variants, proxies, $r^2 \geq 0.5$, ±100kb).

Supplementary File 6: Gene expression in retina and RPE/choroid for genes in 34 narrow AMD regions. Gene expression in human retina tissue as well as retina pigment epithelium (RPE) or human choroid tissue for the 368 genes in 34 narrow AMD locus regions have been provided by two laboratories, the Weber lab and the Stambolian lab (see **Material and Methods**). A consensus rating was obtained by defining the gene as "expressed", if it was expressed in both data sets. It was defined as "not expressed", if it was found as not expressed in at least one laboratory, and as "missing" otherwise.

Supplementary File 7: Relevant eye phenotypes in genetic mouse models in 33 genes in the 34 narrow AMD regions. We queried data bases and conducted a literature search (see **Material and Methods**) for the 368 genes in the 34 narrow AMD regions and found relevant eye phenotypes for 33 of these genes.

Supplementary File 8: Approved and experimental drug targets among 368 genes in narrow AMD region. We queried the DrugBank data base (Version 4.1, Web Resources) to obtain overlap of the 368 genes in our 34 identified AMD regions with drug target list. We found 31 of these genes to be a current drug target.

Supplementary File 9: Summary of biological and statistical evidence for genes in narrow AMD regions. For all genes in the narrow AMD loci (**Supplementary File 5**), we gathered evidence whether the gene (i) was expressed in retina or RPE/choroid (**Supplementary File 6**), (ii) had a retina or RPE/choroid phenotype in genetic mouse models (**Supplementary File 7**), (iii) contained ≥1 variant in a 95% credible set by extending to ± 50 kb around the gene (**Supplementary File 3**), or (iv) had a significant rare variant burden (**Table 2, Supplementary File 4**). Furthermore, we derived whether the credible set variants in the gene (± 50 kb) contained (v) a protein altering, (vi) a 5'/3'UTR, (vii) another exonic coding, or (viii) a putative promoter variant (**Supplementary File 3**), or, whether the gene (ix) was in an enriched molecular pathway (**Supplementary Table 10**), or (x) linked to an approved or experimental drug (**Supplementary File 8**).

References – Supplementary Information:

1. Age-Related Eye Disease Study Research, G. 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* **132**, 668-81 (2001).
2. Seddon, J.M., Sharma, S. & Adelman, R.A. Evaluation of the clinical age-related maculopathy staging system. *Ophthalmology* **113**, 260-6 (2006).
3. Klein, R. *et al.* The Wisconsin age-related maculopathy grading system. *Ophthalmology* **98**, 1128-34 (1991).
4. Bird, A.C. *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* **39**, 367-74 (1995).
5. Ristau, T. *et al.* Allergy is a protective factor against age-related macular degeneration. *Invest Ophthalmol Vis Sci* **55**, 210-4 (2014).
6. Age-Related Eye Disease Study Research, G. A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and E and beta carotene for age-related cataract and vision loss: AREDS report no. 9. *Arch Ophthalmol* **119**, 1439-52 (2001).
7. Klein, R. *et al.* Risk alleles in CFH and ARMS2 and the long-term natural history of age-related macular degeneration: the Beaver Dam Eye Study. *JAMA Ophthalmol* **131**, 383-92 (2013).
8. Yates, J.R. *et al.* Complement C3 variant and the risk of age-related macular degeneration. *N Engl J Med* **357**, 553-61 (2007).
9. Hageman, G.S. *et al.* A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc Natl Acad Sci U S A* **102**, 7227-32 (2005).
10. Pauer, G.J., Sturgill, G.M., Peachey, N.S. & Hagstrom, S.A. Protective effect of paraoxonase 1 gene variant Gln192Arg in age-related macular degeneration. *Am J Ophthalmol* **149**, 513-22 (2010).
11. Fritsche, L.G. *et al.* Seven new loci associated with age-related macular degeneration. *Nat Genet* **45**, 433-9, 439e1-2 (2013).
12. Chowers, I. *et al.* Association of complement factor H Y402H polymorphism with phenotype of neovascular age related macular degeneration in Israel. *Mol Vis* **14**, 1829-34 (2008).
13. Chowers, I. *et al.* Sequence variants in HTRA1 and LOC387715/ARMS2 and phenotype and response to photodynamic therapy in neovascular age-related macular degeneration in populations from Israel. *Mol Vis* **14**, 2263-71 (2008).
14. McCarty, C.A., Wilke, R.A., Giampietro, P.F., Wesbrook, S.D. & Caldwell, M.D. Marshfield Clinic Personalized Medicine Research Project (PMRP): design, methods and recruitment for a large population-based biobank. *Personalized Medicine* **2**, 49-79 (2005).
15. Schmidt, S. *et al.* Association of the apolipoprotein E gene with age-related macular degeneration: possible effect modification by family history, age, and gender. *Mol Vis* **6**, 287-93 (2000).
16. Chen, W. *et al.* Genetic variants near TIMP3 and high-density lipoprotein-associated loci influence susceptibility to age-related macular degeneration. *Proc Natl Acad Sci U S A* **107**, 7401-6 (2010).
17. Schaumberg, D.A. *et al.* Prospective study of common variants in the retinoic acid receptor-related orphan receptor alpha gene and risk of neovascular age-related macular degeneration. *Arch Ophthalmol* **128**, 1462-71 (2010).
18. Jakobsdottir, J. *et al.* Susceptibility genes for age-related maculopathy on chromosome 10q26. *Am J Hum Genet* **77**, 389-407 (2005).
19. Rivera, A. *et al.* Hypothetical LOC387715 is a second major susceptibility gene for age-related macular degeneration, contributing independently of complement factor H to disease risk. *Hum Mol Genet* **14**, 3227-36 (2005).
20. Lotery, A., Xu, X., Zlatava, G. & Loftus, J. Burden of illness, visual impairment and health resource utilisation of patients with neovascular age-related macular degeneration: results from the UK cohort of a five-country cross-sectional study. *Br J Ophthalmol* **91**, 1303-7 (2007).
21. Chen, Y. *et al.* Assessing susceptibility to age-related macular degeneration with genetic markers and environmental factors. *Arch Ophthalmol* **129**, 344-51 (2011).
22. Helgason, H. *et al.* A rare nonsynonymous sequence variant in C3 is associated with high risk of age-related macular degeneration. *Nat Genet* **45**, 1371-4 (2013).

23. Owen, L.A. *et al.* FLT1 genetic variation predisposes to neovascular AMD in ethnically diverse populations and alters systemic FLT1 expression. *Invest Ophthalmol Vis Sci* **55**, 3543-54 (2014).
24. Feng, X. *et al.* Complement factor H Y402H and C-reactive protein polymorphism and photodynamic therapy response in age-related macular degeneration. *Ophthalmology* **116**, 1908-12 e1 (2009).
25. Craig, J.E. *et al.* Rapid inexpensive genome-wide association using pooled whole blood. *Genome Res* **19**, 2075-80 (2009).
26. Oliver, V.F. *et al.* Hypomethylation of the IL17RC promoter in peripheral blood leukocytes is not a hallmark of age-related macular degeneration. *Cell Rep* **5**, 1527-35 (2013).
27. Mitchell, P., Smith, W., Attebo, K. & Wang, J.J. Prevalence of age-related maculopathy in Australia. The Blue Mountains Eye Study. *Ophthalmology* **102**, 1450-60 (1995).
28. Wang, J.J. *et al.* Risk of age-related macular degeneration 3 years after cataract surgery: paired eye comparisons. *Ophthalmology* **119**, 2298-303 (2012).
29. Cugati, S. *et al.* Australian prospective study of cataract surgery and age-related macular degeneration: rationale and methodology. *Ophthalmic Epidemiol* **14**, 408-14 (2007).
30. Raychaudhuri, S. *et al.* A rare penetrant mutation in CFH confers high risk of age-related macular degeneration. *Nat Genet* **43**, 1232-6 (2011).
31. Seddon, J.M. *et al.* Rare variants in CFI, C3 and C9 are associated with high risk of advanced age-related macular degeneration. *Nat Genet* **45**, 1366-70 (2013).
32. van de Ven, J.P. *et al.* A functional variant in the CFI gene confers a high risk of age-related macular degeneration. *Nat Genet* **45**, 813-7 (2013).
33. Zhan, X. *et al.* Identification of a rare coding variant in complement 3 associated with age-related macular degeneration. *Nat Genet* **45**, 1375-9 (2013).
34. Klaver, C.C. *et al.* Genetic association of apolipoprotein E with age-related macular degeneration. *Am J Hum Genet* **63**, 200-6 (1998).
35. Souied, E.H. *et al.* The epsilon4 allele of the apolipoprotein E gene as a potential protective factor for exudative age-related macular degeneration. *Am J Ophthalmol* **125**, 353-9 (1998).
36. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).
37. Hughes, S. *et al.* Profound defects in pupillary responses to light in TRPM-channel null mice: a role for TRPM channels in non-image-forming photoreception. *Eur J Neurosci* **35**, 34-43 (2012).
38. Chen, Y. *et al.* Common variants near ABCA1 and in PMM2 are associated with primary open-angle glaucoma. *Nat Genet* **46**, 1115-9 (2014).
39. Hysi, P.G. *et al.* Genome-wide analysis of multi-ancestry cohorts identifies new loci influencing intraocular pressure and susceptibility to glaucoma. *Nat Genet* **46**, 1126-30 (2014).
40. Gharahkhani, P. *et al.* Common variants near ABCA1, AFAP1 and GMDS confer risk of primary open-angle glaucoma. *Nat Genet* **46**, 1120-5 (2014).
41. Yamamoto, H. *et al.* Mutations in the gene encoding 11-cis retinol dehydrogenase cause delayed dark adaptation and fundus albipunctatus. *Nat Genet* **22**, 188-91 (1999).
42. Jang, G.F. *et al.* Characterization of a dehydrogenase activity responsible for oxidation of 11-cis-retinol in the retinal pigment epithelium of mice with a disrupted RDH5 gene. A model for the human hereditary disease fundus albipunctatus. *J Biol Chem* **276**, 32456-65 (2001).
43. Shang, E. *et al.* Targeted disruption of the mouse cis-retinol dehydrogenase gene: visual and nonvisual functions. *J Lipid Res* **43**, 590-7 (2002).
44. Fogerty, J. & Besharse, J.C. Subretinal infiltration of monocyte derived cells and complement misregulation in mice with AMD-like pathology. *Adv Exp Med Biol* **801**, 355-63 (2014).
45. Sawada, Y. *et al.* Force sensing by mechanical extension of the Src family kinase substrate p130Cas. *Cell* **127**, 1015-26 (2006).
46. Riccomagno, M.M. *et al.* Cas adaptor proteins organize the retinal ganglion cell layer downstream of integrin signaling. *Neuron* **81**, 779-86 (2014).
47. Lambert, V. *et al.* MMP-2 and MMP-9 synergize in promoting choroidal neovascularization. *FASEB J* **17**, 2290-2 (2003).
48. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat Med* **21**, 1539-58 (2002).
49. Arakawa, S. *et al.* Genome-wide association study identifies two susceptibility loci for exudative age-related macular degeneration in the Japanese population. *Nat Genet* **43**, 1001-4 (2011).
50. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control Clin Trials* **7**, 177-88 (1986).

51. Clark, S.J. *et al.* Identification of factor H-like protein 1 as the predominant complement regulator in Bruch's membrane: implications for age-related macular degeneration. *J Immunol* **193**, 4962-70 (2014).
52. Behrens, G., Winkler, T.W., Gorski, M., Leitzmann, M.F. & Heid, I.M. To stratify or not to stratify: power considerations for population-based genome-wide association studies of quantitative traits. *Genet Epidemiol* **35**, 867-79 (2011).
53. Mustafi, D. *et al.* Evolutionarily conserved long intergenic non-coding RNAs in the eye. *Hum Mol Genet* **22**, 2992-3002 (2013).
54. Ikram, M.K. *et al.* Four novel Loci (19q13, 6q24, 12q24, and 5q14) influence the microcirculation in vivo. *PLoS Genet* **6**, e1001184 (2010).
55. Sim, X. *et al.* Genetic loci for retinal arteriolar microcirculation. *PLoS One* **8**, e65804 (2013).