nature genetics

A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants

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, mostly rare, protein-altering variants. Analyzing 16,144 patients and 17,832 controls, we identify 52 independently associated common and rare variants ($P < 5 \times 10^{-8}$) distributed across 34 loci. Although wet and dry AMD subtypes exhibit predominantly shared genetics, we identify the first genetic association signal specific to wet AMD, near *MMP9* (difference *P* value = 4.1×10^{-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 AMD is a neurodegenerative disease and the leading cause of vision loss among the elderly, affecting 5% of those >75 years of age[1,2]. The disease is characterized by reduced retinal pigment epithelium (RPE) function and photoreceptor loss in the macula. Advanced AMD is classified as wet (choroidal neovascularization; when accompanied by angiogenesis) or dry (geographic atrophy; when angiogenesis is absent). These advanced stages of disease are typically preceded by clinically asymptomatic earlier stages[3]. Advanced AMD is estimated to affect 10 million patients worldwide, with the number of affected individuals reaching >150 million for earlier stages[4]. At present, understanding of disease biology and therapies remains limited[5].

Identification of associated genetic variants can help uncover disease mechanisms and provide entry points into therapy. Analyses of common variation have identified numerous risk-associated loci for many complex diseases (GWAS Catalog; see URLs), including 21 loci for AMD[6–12]. However translation of these loci into biological insights remains a challenge, as the functional consequences of disease-associated common variants are typically subtle[13] and hard to decipher.

With advances in sequencing technology, genetic analyses are gradually extending to rare variants, which often have more obvious functional consequences[14,15] and can thus accelerate translation into biological understanding[14,16]. For example, identifying multiple disease-associated coding variants (particularly knockout alleles) in the same gene provides strong evidence that disrupting gene function leads to disease[17]. So far, studies that implicate specific rare variants in complex diseases 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 about the sample sizes and study designs that will enable systematic identification of association with these variants[16].

Here we set out to systematically examine common and rare variation of AMD in the International AMD Genomics Consortium (IAMDGC). 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 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.

RESULTS

Study data and genomic heritability

We gathered advanced AMD cases with geographic atrophy and/or choroidal neovascularization, intermediate AMD cases and control subjects across 26 studies (**Supplementary Table 1**). Whereas recruitment and ascertainment strategies varied (**Supplementary Table 2**), DNA samples were collected and genotyped centrally. Making maximal use of genotyping technologies, we used a chip with (i) the usual genome-wide variant content, (ii) exome content comparable to an exome chip (adding protein-altering variants from across all exons), and a specific customization to add (iii) protein-altering variants detected by our previous sequencing of known AMD-associated loci (Online 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 primarily noncoding (93%) common

A full list of authors and affiliations appears at the end of the paper.

Received 22 May; accepted 22 October; published online 21 December 2015; doi:10.1038/ng.3448

variants (frequency among controls >1%) and 163,714 protein-altering variants (including 8,290 from known AMD-associated loci), which were mostly rare (88% with frequency among controls ≤1%). Imputation to the 1000 Genomes Project reference panel enabled examination of a total of 12,023,830 variants (**Supplementary Table 3**). Our final data set included a total of 43,566 subjects, consisting of 16,144 patients with advanced AMD 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 Fig. 1** and **Supplementary Table 3**).

Altogether, our genotyped markers accounted for 46.7% (ref. 28) of variability in advanced AMD risk in the European-ancestry subjects (95% confidence interval (CI) = 44.5–48.8%). With respect to AMD subtypes, heritability estimates for choroidal neovascularization (h^2 = 44.3%, CI = 42.2–46.5%) and geographic atrophy (h^2 = 52.3%, CI = 47.2–57.4%) were similar; bivariate analyses[29] showed a genetic correlation of 0.85 (CI = 0.78–0.92) between these 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 λ = 1.13) comparing the 16,144 patients with advanced AMD and 17,832 controls of European ancestry (full results available online; see URLs). We obtained >7,000 variants with genome-wide significant association ($P \leq 5 \times 10^{-8}$; **Supplementary Fig. 2**). Sequential forward selection (**Supplementary Fig. 3**) identified 52 independently associated variants at $P \leq 5 \times 10^{-8}$ (**Supplementary Table 4** and **Supplementary Data Set 1**). These variants are distributed across 34 'locus regions' (**Fig. 1a**), each extending across the identified and correlated variants, including variants with $r^2 \geq 0.5$ and ±500 kb with respect to the index SNP (**Supplementary Table 5**). The 34 loci include 16 loci that reached genome-wide significance for the first time (new loci; **Table 1**) and overlap genes with compelling biology, such as extracellular matrix genes (*COL4A3*, *MMP19* and *MMP9*), an ABC transporter linked to high-density lipoprotein (HDL) cholesterol (*ABCA1*) and a key activator in immune function (*PILRB*). Also included are 18 of the 21 AMD susceptibility loci that reached genome-wide significance previously[6,9] (known loci; **Table 1**); between-study heterogeneity was low, particularly for the new loci (**Supplementary Tables 6** and **7**, and **Supplementary Note**).

Most (45/52) associated variants are common, showing fully conditioned odds ratios (ORs; from one model including all 52 variants) from 1.1 to 2.9 (**Fig. 1b** and **Supplementary Table 4**) with two interacting variants (**Supplementary Note**). We also observed seven rare variants with frequencies between 0.01 and 1% and ORs between 1.5 and 47.6 (**Fig. 1b** and **Supplementary Table 4**). All of these variants were also rare in individuals of non-European ancestry (**Supplementary Table 8**; see extended association results for non-Europeans in **Supplementary Data Set 2**). All seven rare variants are located in or near complement genes; they comprise four previously described nonsynonymous variants (*CFH*, p.Arg1210Cys; *CFI*, p.Gly119Arg; *C9*, p.Pro167Ser; *C3*, p.Lys155Gln)[7–11] and three others (rs148553336, rs191281603 and rs35292876 in *CFH*) described here for the first time, including two for which the rare allele decreases disease risk. To ensure the validity of our results, we verified associations of lead variants in sensitivity analyses that relied on alternate association tests, adjusted for age, sex or ten ancestry principal components, or that were restricted to population-based controls or controls ≥50 years of age (data not shown). Altogether, our genome-wide single-variant analysis nearly doubles the number of AMD-associated loci and variants.

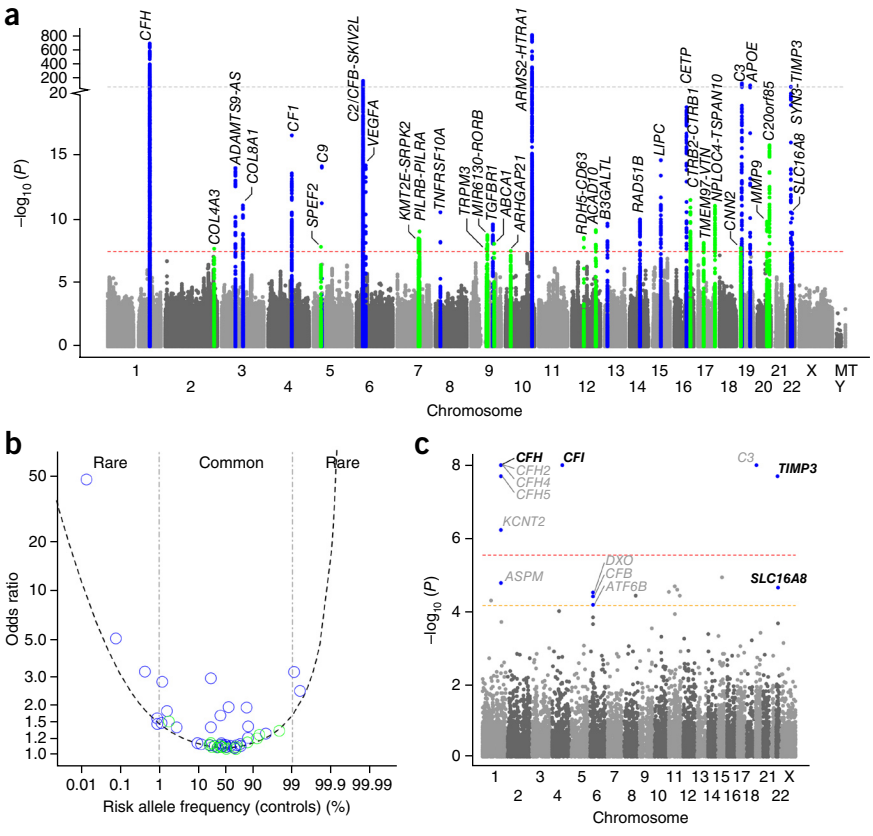

Figure 1 Genome-wide search identifies 34 loci and genes with rare variant burden for AMD. (**a**) We conducted a genome-wide single-variant association analysis for >12 million variants in 16,144 patients with advanced AMD versus 17,832 controls. Shown is the Manhattan plot exhibiting *P* values for association, highlighting new ($P < 5 \times 10^{-8}$ for the first time; green) and known (blue) AMD loci (**Table 1**). The horizontal lines indicate genome-wide significance ($P = 5 \times 10^{-8}$; red) or the *y*-axis break ($P = 1 \times 10^{-20}$; gray). MT, mitochondrial DNA. (**b**) We computed the independent effect size (odds ratio) 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 an 80% power curve (dashed line). (**c**) We conducted a genome-wide gene-based test for disease burden based on protein-altering variants, testing 17,044 RefSeq genes by the variable-threshold test[51]. Shown is the Manhattan plot with *P* values; the red dashed line indicates genome-wide significance ($P \leq 0.05/17,044 = 2.9 \times 10^{-6}$), and the yellow dashed line indicates significance across an AMD-associated locus (given 703 genes in the 34 AMD loci, $P \leq 0.05/703 = 7.1 \times 10^{-5}$). No gene outside the 34 loci is genome-wide significant; 14 genes are significant across the AMD loci (blue), four of which remain significant after locus-wide conditioning (highlighted in bold; **Supplementary Table 11**).

Table 1 Thirty-four loci for age-related macular degeneration

Lead variant	Chr.	Position[a]	Major/ minor allele	Locus name[b]	Number of signals[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}
New (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}

Our genome-wide single-variant association analysis identified 34 loci for advanced AMD with genome-wide significance ($P < 5 \times 10^{-8}$) on the basis of logistic regression in 16,144 cases and 17,832 controls of European ancestry. Shown for each locus are the *P* value and effect size (odds ratio) for the variant with the smallest *P* value (lead variant) and the number of independent signals at the locus (**Supplementary Table 4**). Chr., chromosome; MAF, minor allele frequency; OR, odds ratio.
[a]Chromosomal position is given according to the NCBI RefSeq hg19 human genome reference assembly. [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 independently associated variants in this locus.

Prioritizing variants within 52 association signals

It is often challenging to translate common variant association signals into a mechanistic understanding of biology; two key challenges are (i) variants with similar signals because of linkage disequilibrium (LD) and (ii) variants with subtle functional consequences. Without narrowing lists of candidate variants, follow-up functional experiments are complicated. To prioritize variants near index SNPs for follow-up, 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]. These 52 credible sets each included from one to >100 variants (total of 1,345 variants; **Supplementary Data Set 3**). Twenty-seven of the 52 sets were small, with ≤10 variants (19 sets had ≤5 variants; **Supplementary Table 9**); seven sets included only one variant. Among the 205 variants with >5% probability of being causal, we observe 11 protein-altering (all nonsynonymous) variants (versus two expected assuming that 1% of variants are protein altering; *P* value for enrichment = 8.7×10^{-6}; **Supplementary Table 10**). We recognize that the analysis has limitations (for example, when the signal is due to a combination of multiple variants, as in the counterexample in **Supplementary Fig. 4**).

Rare variant association signals

Analysis of rare variants that alter peptide sequences (nonsynonymous), truncate proteins (stop gain) or affect RNA splicing (splice site) can help to identify causal mechanisms—particularly when multiple associated variants reside in the same gene[16,32]. We examined the cumulative effect of rare protein-altering variants in each ancestry group. Across the genome, no signal was detected with $P \leq 0.05/17{,}044 = 2.9 \times 10^{-6}$ outside the 34 AMD-associated loci (**Fig. 1c**). Within the 34 loci, we found 14 genes with significant disease burden ($P < 0.05/703$ genes = 7.1×10^{-5}; **Supplementary Table 11**). To eliminate scenarios in which a finding of rare variant burden is an LD shadow of association for a nearby common variant, we reevaluated each signal of burden with conditioning on nearby single variants (from **Supplementary Table 4**). Four of the 14 genes

Table 2 Four genes within the 34 AMD loci with a significant rare variant burden independent of identified index variants

	Optimal RAC threshold	Number of variants below optimal RAC	Summed RAC (frequency as %)			
Gene	Count (%)	Total (exome chip base + custom)	Cases n = 16,144	Controls n = 17,832	P^a	OR
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

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 (from up to 100 million permutations) and odds ratios from the collapsed burden test, both adjusted for the other variants identified in each 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 Data Set 4**. Results for the 14 genes that show significant burden within the 34 AMD loci without locus-wide conditioning are shown in **Supplementary Table 11**. Rare variants were defined here as variants with MAF in cases and controls <1% in each of the ancestry groups—European, Asian and African.
RAC, rare allele count.
[a]P values are from the variable-threshold test with conditioning on other variants identified in the locus (locus-wide conditioned).

retained burden at $P < 0.05/703 = 7.1 \times 10^{-5}$ in this analysis (*CFH*, *CFI*, *TIMP3* and *SLC16A8*; conditioned $P = 1.2 \times 10^{-6}$, 1.0×10^{-8}, 9.0×10^{-8} and 3.1×10^{-6}, respectively; **Table 2**). Sensitivity analyses provided similar (excluding previously sequenced subjects) and extended results (prioritizing variants with high predicted functionality; **Supplementary Table 12** and **Supplementary Note**).

Several interesting patterns emerge, many of which we owe to our chip design. First, three of the four rare variant burden signals (*CFH*, *CFI* and *TIMP3*) are due to variants with frequency <0.1%, all genotyped (**Supplementary Data Set 4**). Many human genetic studies have used frequency thresholds of 1 to 5% as a working definition of 'rare', but our data suggest that trait-associated variants with clear function may often be much rarer—necessitating very large sample sizes for analysis. In two genes (*CFH* and *CFI*), the rare burden was detected because we enriched the array with variants from previous sequencing of AMD-associated loci[10] (54 of 80 variants). The burden findings in *CFH* (new locus; **Supplementary Note**) and *CFI*[9], together with the *CFH* variant encoding p.Arg1210Cys and the *CFI* variant encoding p.Gly119Arg (refs. 7,9), corroborate a causal role for these genes in AMD etiology.

a

Gene	Locus number	Locus name	Gene priority score	Expression in retina/RPE/choroid	Eye phenotype in knockout mouse	≥1 variant in 95% credible sets	Rare variant burden	Protein altering	5′ or 3′ UTR	Other exonic	Promoter region	Enriched molecular pathway	Drug target
		Level of evidence		Biology		Statistics		Annotation				Pathways	
				1	1	1	1	1	1	1	1	1	1
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

Gene	Locus number	Locus name	Gene priority score	Expression in retina/RPE/choroid	Eye phenotype in knockout mouse	≥1 variant in 95% credible sets	Rare variant burden	Protein altering	5′ or 3′ UTR	Other exonic	Promoter region	Enriched molecular pathway	Drug target
		Level of evidence		Biology		Statistics		Annotation				Pathways	
				1	1	1	1	1	1	1	1	1	1
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	■	■	■	■						■

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, $r^2 \geq 0.5$ and ±100 kb) for biological evidence (red; gene expression in retina, RPE and/or choroid (**Supplementary Data Set 6**) and ocular mouse phenotypes (**Supplementary Data Set 7**)), statistical evidence (blue; ≥1 credible set variant in the gene or within 50 kb of it (**Supplementary Data Set 3**) and rare variant burden (**Table 2**)), evidence of putative function (green; ≥1 credible set variant that is in the gene or within 50 kb of it that is protein altering or is in the 5′ or 3′ UTR, otherwise exonic or in a putative promoter (**Supplementary Data Set 3**)) and molecular evidence (magenta; gene in an enriched molecular pathway or a drug target). We here focus on the gene(s) with the highest GPS per locus (full list of genes in **Supplementary Data Set 9**). (**a**,**b**) Shown are the 16 genes with the highest GPS values in the 15 new AMD-associated loci (one new locus did not have any gene) (**a**) and the 25 genes with the highest GPS values in the 18 known AMD-associated loci. Colored fields indicate the presence of evidence; the GPS is the total number of colored fields per row.

The third signal (*TIMP3*) was in a gene previously associated with Sorsby's fundus dystrophy, a rare monogenic 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-associated 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-associated locus[5,35] targeted in previous sequencing efforts[32,35] that were too small to evaluate rare variation on this scale (one variant in 17,832 controls versus 29 variants in 16,144 cases). Interestingly, although Sorsby's-associated *TIMP3* variants typically occur in exon 5, four of the unpaired cysteine residues we observed map to other exons—perhaps because unpaired cysteine residues in different locations impair protein folding in different ways. AMD cases with these rare *TIMP3* risk alleles still exhibited 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 other risk-associated 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 signal of rare variant burden in *SLC16A8* was primarily driven by a putative splice-site variant (c.214+1G>C, rs77968014, minor allele frequency among controls, CAF = 0.81%, OR = 1.5, imputed with $R^2 = 0.87$; **Supplementary Data Set 4**). This is not a burden from multiple rare variants but from a single variant emerging as significant owing to the reduced multiple-testing burden from gene-wide testing (single-variant association $P = 9.1 \times 10^{-6}$, P value with conditioning on rs8135665 = 1.3×10^{-6}). This variant is interesting because it is predicted to disrupt processing of the encoded transcript (as a variant at the +1 position; Human Splicing Finder 3.0). *SLC16A8* encodes a cell membrane transporter involved in the 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 mice with knockout of *Slc16a8* have changes in visual function and scotopic electroretinograms but not overt retinal pathology[40]. Interestingly, progressive loss of *SLC16A8* expression in eyes affected with geographic atrophy 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-associated loci to biological insights

Many analyses can further narrow the list of candidate genes in our loci. We annotated the 368 genes closest to our 52 association signals (including the index variant and proxies with $r^2 \geq 0.5$ and ±100 kb relative to index variants; **Supplementary Data Set 5**), noting among these genes those that contained associated credible set variants (**Supplementary Data Set 3**) or a rare variant burden (**Table 2**)—these genes represent 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 Data Set 6**). We sought relevant eye phenotypes in genetically modified mice (observed for 32 of the 368 queried genes; **Supplementary Data Set 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 13**),

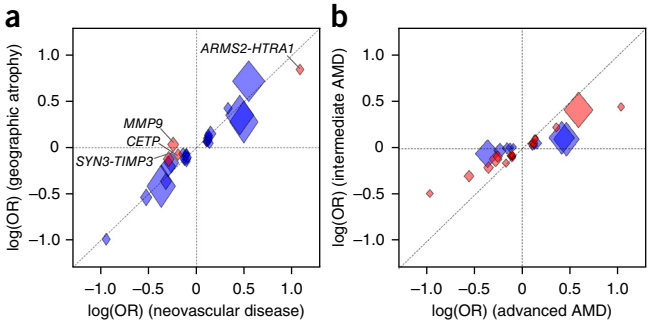

Figure 3 Comparison of advanced AMD subtypes and intermediate versus advanced AMD. We compared associations of the 34 lead variants across different AMD phenotypes. Shown are effect sizes (log(OR)) per minor allele in controls, as well as 95% confidence intervals (widths and heights of diamonds). (**a**) Comparison of neovascular disease (10,749 cases versus 17,832 controls) and geographic atrophy (3,235 cases versus 17,832 controls) identifies four variants (in the *MMP9*, *ARMS2-HTRA1*, *CETP* and *SYN3-TIMP3* loci) with significantly different association in the two subtypes ($P_{difference} < 0.05/34$; red diamonds; **Supplementary Table 16**). (**b**) Comparison of intermediate AMD (6,657 cases versus 17,832 controls) with advanced AMD (16,144 cases versus 17,832 controls) identifies 24 variants with nominally significant ($P < 0.05$; red diamonds) association with intermediate AMD ($P_{binomial} = 4.8 \times 10^{-24}$), all of which have the same direction of effect and less extreme effect sizes in intermediate compared to advanced AMD (**Supplementary Table 17**).

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 Data Set 8**). Finally, we prioritized genes where at least one of the credible set variants (**Supplementary Data Set 3**) was protein altering or located in a putative functional region (promoter, 3′ UTR or 5′ UTR).

All this information is summarized in the gene priority score (GPS) table (**Supplementary Table 14**, **Supplementary Note** and **Supplementary Data Set 9**), which uses a simple customizable scoring scheme to assign priority: the scheme, using equal weights for each column, assigns highest scores (**Fig. 2a** and **Supplementary Table 15**) to genes such as master regulators of immune function (*PILRB*), matrix metalloproteinase genes (*MMP9* and *MMP19*), genes involved in lipid metabolism (*ABCA1* and *GPX4*), an inhibitor of the complement cascade (*VTN*), another collagen gene (*COL4A3*), a gene causing a developmental monogenic disorder (*PTPN11*) and a retinol dehydrogenase (*RDH5*). Six of these are current drug targets (*ABCA1*, *MMP19*, *RDH5*, *PTPN11*, *VTN* and *GPX4*). Among the known AMD-associated loci, the highest scores per locus were found in the 'usual suspects' (*CFH*, *CFI*, *CFB*, *C3* and *APOE*) as well as *TIMP3* and *SLC16A8* (**Fig. 2b**). This summary of evidence is not amenable to formal statistical enrichment testing but may help prioritize genes for follow-up functional experiments.

Commonalities and differences of advanced AMD subtypes

Previously identified risk variants all contribute to the two advanced AMD subtypes—choroidal neovascularization and geographic atrophy. We compared association signals for the 10,749 cases with choroidal neovascularization and the 3,235 cases with geographic atrophy. Four of the 34 lead variants showed a significant difference (difference between subtypes, $P_{difference} < 0.05/34 = 0.00147$) between disease subtypes (in the *ARMS2-HTRA1*, *CETP*, *MMP9* and *SYN3-TIMP3* loci; **Fig. 3a** and **Supplementary Table 16**). Variant rs42450006

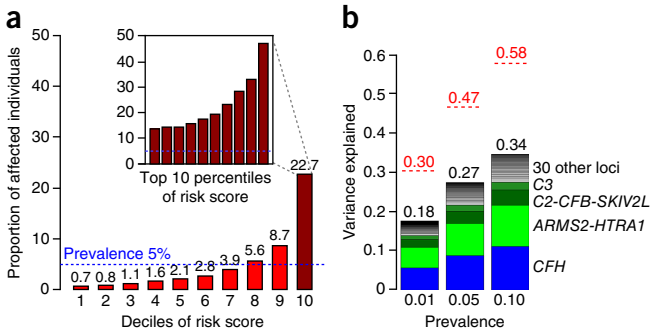

Figure 4 Variance explained and absolute risk of disease based on the 52 identified variants. (**a**) Absolute disease risk (proportion of affected individuals) by genetic risk score (GRS) intervals (deciles and top ten percentiles in embedded bar plot) based on our case-control data weighted to model a general population with disease prevalence of 5% (**Supplementary Table 20**). (**b**) Shown is the 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%.

upstream of *MMP9* was the only one that was specific to one subtype, being exclusively associated with choroidal neovascularization (frequency in controls = 14.1%, $P = 8.4 \times 10^{-17}$, OR = 0.78) but not with geographic atrophy ($P = 0.39$, OR = 1.04; $P_{difference} = 4.1 \times 10^{-10}$; **Supplementary Note**). The *MMP9* signal for neovascular disease fits well with previous evidence: upregulation of *MMP9* appears to induce neovascularization[43] and interacts with vascular endothelial growth factor (VEGF) signaling in the RPE[44]. VEGF currently provides an effective therapy target for patients with choroidal neovascularization, but the struggle to keep vision continues. Beyond confirming shared genetic predisposition for the two subtypes, our data identify for the first time, to our knowledge, a variant that is specific to one subtype.

Commonalities and differences of advanced and early AMD

We evaluated our association signals in 6,657 individuals with intermediate AMD, defined as having more than five macular drusen with a diameter of greater than 63 µm and/or pigmentary changes in the RPE. Examining all genotyped variants[28], we found a correlation of 0.78 (95% CI = 0.69–0.87) between intermediate and advanced AMD, indicating substantial overlap between genetic determinants of advanced and intermediate AMD. Of our 34 index variants, 24 showed nominally significant association ($P_{intermediate} \leq 0.05$) with intermediate AMD (two expected, $P_{binomial} = 4.8 \times 10^{-24}$); all had odds ratios in the same direction but smaller in magnitude (**Fig. 3b** and **Supplementary Table 17**). The other ten variants showed no association with intermediate AMD ($P_{intermediate} > 0.05$), despite sufficient power to detect association (**Supplementary Table 18**). Interestingly, these ten variants point to seven extracellular matrix genes (*COL15A1*, *COL8A1*, *MMP9*, *PCOLCE*, *MMP19*, *CTRB1-CTRB2* and *ITGA7*; **Supplementary Table 19**), on the basis of which one may hypothesize that extracellular matrix variation represents a disease subtype without early-stage manifestation or with extremely rapid progression. If this hypothesis is confirmed, a group of rapidly progressing patients or patients 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 population risk score distribution (Online Methods). Individuals in the highest decile of genetic risk have 44-fold increased risk of developing advanced AMD in comparison to those in the lowest decile; of those in the highest decile, 22.7% are predicted to have AMD in an elderly general population above 75 years of age with ~5% disease prevalence (**Fig. 4a**, **Supplementary Table 20**). Altogether, the 52 variants explain 27.2% of disease variability (**Fig. 4b**, also showing results based on other prevalence assumptions), including a contribution of 1.4% from rare variants. The 52 identified variants thus explain more than half of the genomic heritability; 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.

DISCUSSION

We set out to improve understanding of rare and common genetic variation for macular degeneration biology, to thereby guide the development of therapeutic interventions and facilitate early diagnosis, monitoring and prevention of disease. 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 a frequency >0.5% and >1.8-fold increased disease risk). Our study provides a 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 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 across the genome, significant signals can be assigned to single rare variants as well as to individual genes with rare variant disease burden.

Our study also demonstrates the challenges of these analyses. For three of the genes for which we identified rare variant burden, the accumulated evidence was spread across very rare variants with frequencies <0.1% in controls. Most of these variants were derived from sequencing patients with AMD. This emphasizes the value of a hybrid approach with direct targeted sequencing of patient samples for variant discovery followed by genotyping in larger samples for association analysis. Another conclusion is related to required sample sizes: although rare variants such as these 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 a combination of more sequencing and even larger sample sizes. Although our findings of rare variant burden are predominantly from targeted enrichment, knowledge of the effect sizes and frequencies of contributing variants illustrates that applying the approach across the genome to detect new loci requires extremely large sample sizes. In our view, a recent estimate that sequencing of 25,000 patients will be needed to identify genes where rare variants have a substantial impact on disease risk is likely to be optimistic, 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 splicing variant in *SLC16A8* can greatly increase the risk of AMD, providing strong evidence that this gene is directly involved in disease biology. *SLC16A8* encodes a lactate transporter expressed[39] specifically by the RPE; a deficit in lactate transport results in acidification of the retina and photoreceptor dysfunction in *Slc16a8*-knockout 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.

Although 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 do not support this notion. Our study and others successfully identify many low-frequency disease risk alleles, and these provide clues about disease biology; however, our results also show that common variants make a much larger contribution to disease risk. Common variants also suggest interesting leads and pathways for future analysis (**Fig. 2a** and **Supplementary Table 15**), 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* and *GPX4*). In a literature search, we identified no previous candidate gene association studies targeting our newly associated loci, although several model organism, cellular and functional studies evaluated potential links between genes in these loci and AMD (highlights from this search appear in **Supplementary Table 15**), and a few loci were nominally associated and proposed as candidates in previous genome-wide searches[46,47]. As richer functional annotations of the genome[48] 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. Although our results show that the majority of genetic risk is shared by geographic atrophy and choroidal neovascularization, we also identify for the first time, to our knowledge, a variant that is specific to one advanced AMD subtype: a genetic variant near *MMP9* specific to choroidal neovascularization. *MMP9* is a candidate gene also supported by previous gene expression analyses in the Bruch's membrane of patients with neovascular disease[49]. Future efforts extending to longitudinal data might help improve the dissection of pure choroidal neovascularization and pure geographic atrophy and their genetic make-up even further. If substantiated, the fact that nearly all disease-associated variants modulate risk of both choroidal neovascularization and geographic atrophy has potentially major therapeutic consequences. It implies that individuals at high risk of choroidal neovascularization are also at high risk of geographic atrophy. This suggests that therapeutic strategies that mitigate choroidal neovascularization but not geographic atrophy will only provide temporary relief to patients, who are likely to remain at high risk of developing geographic atrophy and may still require future interventions to prevent this.

Therefore, our findings have several key 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 tens of thousands 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. Because all the large-effect rare variants we identify reside in or near loci identified by genome-wide association study (GWAS), as is the case for most complex trait–associated rare variants[7–11,20,21,23,50], studies focused on regions around GWAS loci may continue to be a cost-effective compromise. Third, our analysis of cysteine variants encoded in *TIMP3* illustrates not only the potential for targeted variant discovery but also the critical need to understand the consequences of rare variants when analyzing them together. Although 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 understanding of disease biology.

URLs. GWAS Catalog, http://www.ebi.ac.uk/gwas/home; Exome Variant Server, National Heart, Lung, and Blood Institute (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 Project reference panel, http://www.sph.umich.edu/csg/abecasis/MACH/download/1000G.2013-09.html; Human Genome Diversity Project (HGDP) 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/; Human Splicing Finder 3.0, http://www.umd.be/HSF3/index.html; 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/index.html; 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 12 January 2015), http://www.reactome.org/; Gene Ontology (GO) Consortium (downloaded 12 January 2015), http://geneontology.org/; DrugBank (downloaded 4 June 2014), http://www.drugbank.ca/; GCTA, http://www.complextraitgenomics.com/software/gcta/; variance (or heritability) explained by genetic variants, https://sites.google.com/site/honcheongso/software/varexp.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Data permitted for sharing by respective institutional review boards and summary statistics reported in the paper have been deposited in the database of Genotypes and Phenotypes (dbGaP) under accession phs001039.v1.p1. Full GWAS summary statistics are available at http://amdgenetics.org/.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank all participants of all the studies included for enabling this research by their participation in 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. Group-specific acknowledgments can be found in the **Supplementary Note**. The Center for Inherited Diseases Research (CIDR) Program contract number is HHSN268201200008I. This and the main consortium work were predominantly funded by 1X01HG006934-01 to G.R.A. and R01 EY022310 to J.L.H.

AUTHOR CONTRIBUTIONS

Clinical ascertainment, contribution of samples, study coordination and data analysis. G.R.A., A.A., J.A., R.A., I.A., A. Brucker, P.N.B., E.B., M. Benchaboune,

H.B., J.B., F.B., A. Boleda, C.B., K.E.B., M.H.B., K.P.B., M.S.C., P.C., A.C., D. Chen, D. Cho, I.C., I.J.C., J.E.C., A.J.C., C.A.C., M.D., J.-F.D., A.I.d.H., B.D., L.E., L.A.F., S.F., H.F., K.F., J.R.F., L.G.F., L.G., B.G., M.B.G., S.V.G., R.H.G., S.H.-L., S.A.H., J.L.H., J.H., M.A.H., C.H., S.J.H., J.R.H., I.M.H., A.W.H., J.D.H., F.G.H., C.B.H., D.J.H., T.I., S.K.I., M.P.J., N.K., J.C.K., I.K.K., T.E.K., C.C.W.K., B.E.K.K., M.L.K., R.K., J.L.K., A.M.K., S.L., T. Langmann, R.L., Y.T.E.L., K.E.L., T. Léveillard, M.L., H.H.L., G.L., D.L., A.J.L., H.L., D.A.M., G.M., T.M.M., I.L.M., J.A.M., J.E.M., J.C.M., S.M.M., P.M., S.M.-S., A.T.M., E.L.M., C.E.M., A.O., M.I.O., H.O., K.H.P., N.S.P., M.A.P.-V., E.A.P., C.A.R., A.J.R., G.R., J.-A.S., N.T.M.S., D.A.S., T.S., H.P.N.S., S.G.S., W.K.S., S.S., H.S., G.S., R.T.S., E. Souied, E. Souzeau, D.S., Z.S., A.S., A.G.T., B.T., E.E.T., C.M.v.D., C.N.v.S., B.J.V., J.J.W., B.H.F.W., D.E.W., C.W., A.W., Z.Y., J.R.W.Y., D.Z. and K.Z.

Phenotype committee. I.K.K. (lead), S.K.I. (lead), M.D. (lead), G.H.S.B., E.Y.C., I.C., A.I.d.H., S.F., M.B.G., J.L.H., I.M.H., A.W.H., C.C.W.K., B.E.K.K., M.L.K., R.K., T. Léveillard, A.J.L., K.H.P., J.J.W. and K.Z.

Data analysis. *Team 1: quality control of data:* J.L.B.-G., M.D., L.G.F., M. Gorski, W.I. and I.K.K. *Team 2: single-variant analysis:* L.G.F. (lead), I.M.H. (lead), G.R.A. (lead), W.I. (lead), J.L.B.-G., G.H.S.B., V.C., M.D., M. Gorski, F.G., M. Grunin, J.L.H., R.P.I., S.K.I., C.C.W.K., M.O., K.S. and X.Z. *Team 3: pathway and rare variant burden analysis:* L.G.F. (lead), J.N.C.B. (lead), M.S. (lead), G.R.A., M.A.B., M. Brooks, G.H.S.B., M.D.C., M.D., E.K.d.J., A.I.d.H., L.A.F., F.G., J.L.H., I.M.H., J.D.H., W.I., R.P.I., S.K.I., Y.J., M.A.M., M.O., M.A.P.-V., R.J.S., W.K.S., K.S., A.S., B.H.F.W., D.E.W. and X.Z. *Team 4: analysis of non-SNP variation:* R.P.I. (lead), S.K.I. (lead), P.N.B. (lead), G.R.A., M.D.C., L.G.F. and J.L.H. *Team 5: functional data analysis:* D.S. (lead), B.H.F.W. (lead), M.D. (lead), S.K.I. (lead), V.C., J.N.C.B., M.D.C., E.K.d.J., A.I.d.H., S.F., L.G.F., F.G., J.L.H., C.H., I.M.H., W.I., D.J.M., M.A.M., R.R., C.M.S., A.S. and X.Z.

Design of overall experiment. G.R.A., M.D., L.G.F., J.L.H., I.M.H., S.K.I., M.A.P.-V. and B.H.F.W.

Genotyping and quality control. K.F.D. (lead), J.R. (lead), L.G.F. (lead), M. Gorski (lead), G.R.A., J.L.B.-G., M.D.C., F.G., J.L.H., I.M.H., J.D.H., W.I., M.O. and X.Z.

Writing team. L.G.F. (lead), I.M.H. (lead), G.R.A., J.N.C.B., M.D., J.L.H., W.I., S.K.I., I.K.K., D.S. and B.H.F.W.

Critical review of manuscript. G.R.A., R.A., P.N.B., M.H.B., I.C., J.N.C.B., M.D., S.F., A.I.d.H., L.A.F., L.G.F., M.B.G., S.A.H., J.L.H., C.H., I.M.H., A.W.H., W.I., S.K.I., I.K.K., C.C.W.K., B.E.K.K., M.L.K., R.K., T. Léveillard, A.J.L., P.M., A.T.M., K.H.P., N.S.P., M.A.P.-V., D.A.S., D.S., A.S., J.J.W., B.H.F.W., D.E.W., J.R.W.Y. and K.Z.

Steering committee of IAMDGC. A.S., G.R.A., A.W.H., M.H.B., K.Z., B.H.F.W., I.M.H., M.D., L.A.F., K.H.P., I.K.K., D.S., T. Léveillard, A.J.L., I.C., S.K.I., S.A.H., N.S.P., B.E.K.K., R.K., D.A.S., M.A.P.-V., P.M., J.J.W., R.A., A.T.M., J.R.W.Y., J.L.H., S.F., A.I.d.H., P.N.B., M.L.K., M.B.G., D.E.W., C.H. and C.C.W.K.

Senior executive committee of IAMDGC. G.R.A., M.D., J.L.H., S.K.I., M.A.P.-V. and B.H.F.W.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

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. *Br. Med. J.* **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**, 1570–1574 (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–e116 (2014).
5. Fritsche, L.G. *et al.* Age-related macular degeneration: genetics and biology coming together. *Annu. Rev. Genomics Hum. Genet.* **15**, 151–171 (2014).
6. Fritsche, L.G. *et al.* Seven new loci associated with age-related macular degeneration. *Nat. Genet.* **45**, 433–439 (2013).
7. Raychaudhuri, S. *et al.* A rare penetrant mutation in *CFH* confers high risk of age-related macular degeneration. *Nat. Genet.* **43**, 1232–1236 (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**, 1371–1374 (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**, 1366–1370 (2013).
10. Zhan, X. *et al.* Identification of a rare coding variant in complement 3 associated with age-related macular degeneration. *Nat. Genet.* **45**, 1375–1379 (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**, 813–817 (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**, 1001–1004 (2011).
13. Gibson, G. Rare and common variants: twenty arguments. *Nat. Rev. Genet.* **13**, 135–145 (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–R9 (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**, 100–104 (2012).
16. Zuk, O. *et al.* Searching for missing heritability: designing rare variant association studies. *Proc. Natl. Acad. Sci. USA* **111**, E455–E464 (2014).
17. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546–1558 (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**, 517–520 (2013).
20. Rivas, M.A. *et al.* Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat. Genet.* **43**, 1066–1073 (2011).
21. Flannick, J. *et al.* Loss-of-function mutations in *SLC30A8* protect against type 2 diabetes. *Nat. Genet.* **46**, 357–363 (2014).
22. Cruchaga, C. *et al.* Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature* **505**, 550–554 (2014).
23. Do, R. *et al.* Exome sequencing identifies rare *LDLR* and *APOA5* alleles conferring risk for myocardial infarction. *Nature* **518**, 102–106 (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**, 233–245 (2014).
25. Walters, R.G. *et al.* A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* **463**, 671–675 (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**, 187–196 (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**, 2540–2542 (2012).
30. Wellcome Trust Case Control Consortium. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44**, 1294–1301 (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**, 387–389 (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**, 158–161 (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**, 482–494 (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**, 937–948 (2010).
37. Köttgen, A. *et al.* New loci associated with kidney function and chronic kidney disease. *Nat. Genet.* **42**, 376–384 (2010).
38. Allikmets, R. *et al.* Mutation of the Stargardt disease gene (*ABCR*) in age-related macular degeneration. *Science* **277**, 1805–1807 (1997).
39. Halestrap, A.P. The *SLC16* gene family–structure, role and regulation in health and disease. *Mol. Aspects Med.* **34**, 337–349 (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–C457 (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. USA* **78**, 4828–4832 (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**, 960–968 (2014).
44. Hollborn, M. *et al.* Positive feedback regulation between MMP-9 and VEGF in human RPE cells. *Invest. Ophthalmol. Vis. Sci.* **48**, 4360–4367 (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**, 571–580 (2012).

46. 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. USA* **107**, 7401–7406 (2010).
47. Logue, M.W. *et al.* A search for age-related macular degeneration risk variants in Alzheimer disease genes and pathways. *Neurobiol. Aging* **35**, 1510.e7–1510.e18 (2014).
48. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
49. 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**, 4459–4466 (2011).
50. Johansen, C.T. *et al.* Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat. Genet.* **42**, 684–687 (2010).
51. Price, A.L. *et al.* Pooled association tests for rare variants in exon-resequencing studies. *Am. J. Hum. Genet.* **86**, 832–838 (2010).

Lars G Fritsche[1,100], Wilmar Igl[2,100], Jessica N Cooke Bailey[3,100], Felix Grassmann[4,100], Sebanti Sengupta[1,100], 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[11–15], 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[38], 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], 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[49–53], Jeeyun Ahn[54], Giuliana Silvestri[55], Evangelia E Tsironi[56], Kyu Hyung Park[57], Lindsay A Farrer[49–53], Anton Orlin[58], Alexander Brucker[59], Mingyao Li[60], Christine A Curcio[61], Saddek Mohand-Saïd[62–65], José-Alain Sahel[25,62–67], Isabelle Audo[62–64,68], Mustapha Benchaboune[65], Angela J Cree[69], Christina A Rennie[70], Srinivas V Goverdhan[69], Michelle Grunin[71], Shira Hagbi-Levi[71], Peter Campochiaro[11,13], Nicholas Katsanis[72–74], Frank G Holz[17], Frédéric Blond[62–64], Hélène Blanché[75], Jean-François Deleuze[75,76], Robert P Igo Jr[3], Barbara Truitt[3], Neal S Peachey[18,77], Stacy M Meuer[19], Chelsea E Myers[19], Emily L Moore[19], Ronald Klein[19], Michael A Hauser[78–80], Eric A Postel[78], Monique D Courtenay[22], Stephen G Schwartz[81], Jaclyn L Kovach[81], William K Scott[22], Gerald Liew[23], Ava G Tan[23], Bamini Gopinath[23], John C Merriam[24], R Theodore Smith[24,82], Jane C Khan[41,83,84], Humma Shahid[84,85], Anthony T Moore[25,26,86], J Allie McGrath[27], Reneé Laux[3], Milam A Brantley Jr[87], Anita Agarwal[87], 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[88], John Blangero[31], Daniel E Weeks[32,89], Bal Dhillon[90], Cornelia M van Duijn[35], Kimberly F Doheny[91], Jane Romm[91], Caroline C W Klaver[34,35], Caroline Hayward[33], Michael B Gorin[92,93], Michael L Klein[88], Paul N Baird[30], Anneke I den Hollander[29,94], Sascha Fauser[28], John R W Yates[25,26,84], Rando Allikmets[24,95], Jie Jin Wang[23], Debra A Schaumberg[20,96,97], Barbara E K Klein[19], Stephanie A Hagstrom[77], Itay Chowers[71], Andrew J Lotery[69], Thierry Léveillard[62–64], Kang Zhang[8,44], Murray H Brilliant[7], Alex W Hewitt[6,30,41], Anand Swaroop[38], Emily Y Chew[98], Margaret A Pericak-Vance[22,101], Margaret DeAngelis[48,101], Dwight Stambolian[10,101], Jonathan L Haines[3,99,101], Sudha K Iyengar[3,101], Bernhard H F Weber[4,101], Gonçalo R Abecasis[1,101] & Iris M Heid[2,101]

[1]Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA. [2]Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany. [3]Department of Epidemiology and Biostatistics, Case Western Reserve University School of Medicine, Cleveland, Ohio, USA. [4]Institute of Human Genetics, University of Regensburg, Regensburg, Germany. [5]Kidney Epidemiology and Cost Center, Department of Internal Medicine–Nephrology, University of Michigan, Ann Arbor, Michigan, USA. [6]School of Medicine, Menzies Research Institute Tasmania, University of Tasmania, Hobart, Tasmania, Australia. [7]Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, Wisconsin, USA. [8]Department of Ophthalmology, University of California, San Diego and Veterans Affairs San Diego Health System, La Jolla, California, USA. [9]Retina Service, Massachusetts Eye and Ear, Department of Ophthalmology, Harvard Medical School, Boston, Massachusetts, USA. [10]Department of Ophthalmology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [11]Department of Ophthalmology, Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [12]Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [13]Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [14]Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [15]Institue de la Vision, Université Pierre et Marie Curie, Paris, France. [16]Hôpital Intercommunal de Créteil, Hôpital Henri Mondor, Université Paris Est Créteil, Créteil, France. [17]Department of Ophthalmology, University of Bonn, Bonn, Germany. [18]Louis Stokes Cleveland Veterans Affairs Medical Center, Cleveland, Ohio, USA. [19]Department of Ophthalmology and Visual Sciences, University of Wisconsin, Madison, Wisconsin, USA. [20]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. [21]Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA. [22]John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, Florida, USA. [23]Centre for Vision Research, Department of Ophthalmology and Westmead Millennium Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia. [24]Department of Ophthalmology, Columbia University, New York, New York, USA. [25]University College London Institute of Ophthalmology, University College London, London, UK. [26]Moorfields Eye Hospital, London, UK. [27]Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, Tennessee, USA. [28]Department of Ophthalmology, University Hospital of Cologne, 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, Australia. [31]South Texas Diabetes and Obesity Institute, School of Medicine,

University of Texas Rio Grande Valley, Brownsville, Texas, USA. [32]Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [33]Medical Research Council (MRC) Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [34]Department of Ophthalmology, Erasmus Medical Center, Rotterdam, the Netherlands. [35]Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands. [36]Quantitative Biomedical Research Center, Department of Clinical Science, University of Texas Southwestern Medical Center, Dallas, Texas, USA. [37]Center for the Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, Texas, USA. [38]Neurobiology, Neurodegeneration and Repair Laboratory (N-NRL), National Eye Institute, US National Institutes of Health, Bethesda, Maryland, USA. [39]Department of Ophthalmology and Visual Sciences, University of Michigan, Kellogg Eye Center, Ann Arbor, Michigan, 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, Chengdu, China. [45]EyeCentre Southwest, Stuttgart, Germany. [46]University Eye Clinic, Ludwig Maximilians University, Munich, Germany. [47]Department of Ophthalmology, University Hospital Regensburg, Regensburg, Germany. [48]Department of Ophthalmology and Visual Sciences, University of Utah, Salt Lake City, Utah, USA. [49]Department of Medicine (Biomedical Genetics), Boston University Schools of Medicine and Public Health, Boston, Massachusetts, USA. [50]Department of Ophthalmology, Boston University Schools of Medicine and Public Health, Boston, Massachusetts, USA. [51]Department of Neurology, Boston University Schools of Medicine and Public Health, Boston, Massachusetts, USA. [52]Department of Epidemiology, Boston University Schools of Medicine and Public Health, Boston, Massachusetts, USA. [53]Department of Biostatistics, Boston University Schools of Medicine and Public Health, Boston, Massachusetts, 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, New York, USA. [59]Scheie Eye Institute, Department of Ophthalmology, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA. [60]Department of Biostatistics and Epidemiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA. [61]Department of Ophthalmology, University of Alabama at Birmingham, Birmingham, Alabama, USA. [62]INSERM, Paris, France. [63]Institut de la Vision, Department of Genetics, Paris, France. [64]Centre National de la Recherche Scientifique (CNRS), Paris, France. [65]Centre Hospitalier National d'Ophtalmologie des Quinze-Vingts, Paris, France. [66]Fondation Ophtalmologique Adolphe de Rothschild, Paris, France. [67]Académie des Sciences–Institut de France, Paris, France. [68]Department of Molecular Genetics, Institute of Ophthalmology, London, UK. [69]Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK. [70]University Hospital Southampton, Southampton, UK. [71]Department of Ophthalmology, Hadassah Hebrew University Medical Center, Jerusalem, Israel. [72]Center for Human Disease Modeling, Duke University, Durham, North Carolina, USA. [73]Department of Cell Biology, Duke University, Durham, North Carolina, USA. [74]Department of Pediatrics, Duke University, Durham, North Carolina, USA. [75]Centre d'Etude du Polymorphisme Humain (CEPH) Fondation Jean Dausset, Paris, France. [76]Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Institut de Génomique, Centre National de Génotypage, Evry, France. [77]Cole Eye Institute, Cleveland Clinic, Cleveland, Ohio, USA. [78]Department of Ophthalmology, Duke University Medical Center, Durham, North Carolina, USA. [79]Department of Medicine, Duke University Medical Center, Durham, North Carolina, USA. [80]Duke Molecular Physiology Institute, Duke University Medical Center, Durham, North Carolina, USA. [81]Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Naples, Florida, USA. [82]Department of Ophthalmology, New York University School of Medicine, New York, New York, USA. [83]Department of Ophthalmology, Royal Perth Hospital, Perth, Western Australia, Australia. [84]Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK. [85]Department of Ophthalmology, Cambridge University Hospitals National Health Service (NHS) Foundation Trust, Cambridge, UK. [86]Department of Ophthalmology, University of California San Francisco Medical School, San Francisco, California, USA. [87]Department of Ophthalmology and Visual Sciences, Vanderbilt University, Nashville, Tennessee, USA. [88]Casey Eye Institute, Oregon Health and Science University, Portland, Oregon, USA. [89]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [90]School of Clinical Sciences, University of Edinburgh, Edinburgh, UK. [91]Center for Inherited Disease Research (CIDR) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [92]Department of Ophthalmology, David Geffen School of Medicine, Stein Eye Institute, University of California, Los Angeles, Los Angeles, California, USA. [93]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA. [94]Department of Human Genetics, Radboud University Medical Centre, Nijmegen, the Netherlands. [95]Department of Pathology and Cell Biology, Columbia University, New York, New York, USA. [96]Center for Translational Medicine, Moran Eye Center, University of Utah School of Medicine, Salt Lake City, Utah, USA. [97]Division of Preventive Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. [98]Division of Epidemiology and Clinical Applications, Clinical Trials Branch, National Eye Institute, US National Institutes of Health, Bethesda, Maryland, USA. [99]Institute for Computational Biology, Case Western Reserve University School of Medicine, Cleveland, Ohio, USA. [100]These authors contributed equally to this work. [101]These authors jointly supervised this work. Correspondence should be addressed to I.M.H. (iris.heid@klinik.uni-regensburg.de), G.R.A. (goncalo@umich.edu) or S.K.I. (ski@case.edu).

ONLINE METHODS

Study data and phenotype. In IAMDGC, we gathered 26 studies with each including (i) advanced AMD cases with geographic atrophy and/or choroidal neovascularization 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 63 μm in diameter 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**, and **Supplementary Note**). All groups collected data according to 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 from more than 50,000 individuals. Groups with very limited amounts of available DNA contributed aliquots after whole-genome amplification (8% of subjects).

We used a custom-modified HumanCoreExome array by Illumina, which includes (i) tagging variants across the genome (genome chip content) and (ii) a catalog 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], including 19 index variants with genome-wide significance and three variants with consistent direction of effect in the replication stage and $4 \times 10^{-7} \leq P \leq 2 \times 10^{-6}$), selecting (iii) tagging variants (pairwise tagging $r^2 < 0.8$) from Phase I 1000 Genomes Project or HapMap[52,53] common variants (MAF ≥1% in European or East Asian individuals) using Tagger implemented in Haploview[54] within 100 kb 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 databases (dbSNP[55], the NHLBI Exome Sequencing Project[56] and Phase I of the 1000 Genomes Project; see URLs), and (v) protein-altering variants within 500 kb of the 22 index variants identified by resequencing AMD case-control study data (targeted resequencing of 2,335 AMD cases and 789 controls[10,57] and whole-genome sequencing of 60 AMD cases and 60 controls; A.M.K., X.Z., L.G.F., G.R.A. and A.S. *et al.*, unpublished data). The customization further included (vi) the 1,000 top independent (>2 Mb distant) variants from the previous analysis and 100 additional top variants each from the previous choroidal neovascularization only and geographic atrophy only analyses, (vii) and 375 variants in *ABCA4*, including known variants causing Stargardt disease[58], benign variants and those with unknown consequences, as well as ten known and 44 predicted cysteine-altering mutations in *TIMP3*, whose selection was motivated by the known variants causing Sorsby's fundus dystrophy[33,34].

Annotation. Variant identifiers were based on NCBI dbSNP v137. Chromosomal position and functional annotation of the variants was based on the NCBI RefSeq hg19 reference genome[59] (downloaded December 2012) and SeattleSeq Annotation 138 (ref. 60; see URLs). We particularly focus on protein-altering variants, including nonsynonymous coding variants (missense, stop loss, in-frame insertion or deletion, frameshift and stop gain) and splice-site variants. We converted the description of splice-site variants to HGVS nomenclature using Mutalyzer version 2.0.beta-33 (ref. 61; see URLs).

Genotypes. We genotyped all subjects centrally at the Center for Inherited Diseases Research (CIDR) at Johns Hopkins University School of Medicine. From the 569,645 genotyped variants, our quality control excluded variants with poor genotyping as evidenced by genotype call rates <98.5% (5.8%), deviation from Hardy-Weinberg equilibrium with $P < 1 \times 10^{-6}$ (0.34%), mapping to multiple genome locations (0.25%) or failure of other criteria, resulting in 521,950 (91.6%) variants passing all quality criteria. After excluding monomorphic variants (15.8%), we retained 264,655 common variants distributed across autosomes, sex chromosomes and mitochondrial DNA, 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 whole genome–amplified DNA).

We phased the autosomal and X-chromosomal genotype data using SHAPEIT (200 states, 2.5-Mb windows)[62] and then imputed genotypes on the basis of the 1000 Genomes Project[63] reference panel (1000 Genomes Project Phase I, version 3, SHAPEIT2 Reference) using Minimac[64] (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 poor 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 filtering 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 $\phi \geq 0.0884$, indicating third-degree relatives or closer)[65], subjects with discrepancies between reported sex and sex chromosomal information or with atypical sex chromosome configurations[66], and subjects with genotyping call rates <98.5%; we derived ancestry from the first two principal components using autosomal genotyped variants together with genotype information for the samples from the Human Genome Diversity Project (HGDP)[67]. 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 ancestry (**Supplementary Table 3b**).

Genomic heritability and genomic correlation. Combined contribution of genotyped variants to disease was evaluated using a variance component–based heritability analysis[68]. This analysis used genotypes to build a similarity matrix, summarizing the overall genetic kinship for each pair of individuals, and then examined the correspondence between genetic and phenotypic similarity. We estimated the variance explained by all genotyped autosomal variants using restricted maximum likelihood (REML) analysis implemented in GCTA[28] (see URLs). We jointly estimated the contributions of rare (CAF <1%) and common (CAF ≥1%) genotyped variants by first separately calculating their genetic relationship matrices (GRMs) 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[68].

We estimated the genomic correlation between disease subphenotypes using bivariate REML analyses implemented in GCTA and only included common (CAF ≥1%) genotyped variants[29]. We compared 10,749 cases with choroidal neovascularization to 3,325 cases with geographic atrophy (excluding the 2,070 cases with mixed choroidal neovascularization and geographic atrophy), and we compared 6,657 intermediate AMD cases with 16,144 advanced AMD cases. For both analyses, we used the control subjects as a reference and avoided shared controls between traits by randomly splitting the 17,832 unrelated European control individuals into two subsamples of 8,916 individuals each.

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[69], which is recommended for genetic association studies that include rare variants[70], as implemented in EPACTS (see URLs). Analyses were adjusted for two principal components and source of DNA (whole-blood or whole genome–amplified DNA). Allele dosages were used for the imputed data. Sensitivity analyses were conducted to evaluate the influence of alternative association tests and alternative covariate adjustment including with age or sex or with up to ten 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[71] 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 GWAS, we considered $P \leq 5 \times 10^{-8}$ to indicate genome-wide significance.

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, for each significantly associated locus, we selected the variant with the smallest *P* value and the 10-Mb region centered on this variant, repeating the process until no genome-wide significant variant ($P \leq 5 \times 10^{-8}$) was left, thereby identifying a number of 10-Mb regions.

Within each of these large regions, we reanalyzed 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 the region spanned by the correlated variants ($r^2 \geq 0.5$) adding a further 500 kb to both sides; overlapping locus regions were merged into one locus, and some loci therefore contained more than one index variant (details in **Supplementary Fig. 3**).

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

Bayesian approach to prioritize variants. To summarize the statistical evidence of association strength, for each variant, we computed the Bayes factor, which is a measure of the strength of the association that is comparable between variants irrespective of variant frequency or study sample size. The Bayes factor 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 for each variant[72]. The posterior probability of each variant is then computed as the Bayes factor for the variant relative to the sum of the Bayes factors for all variants 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 approach 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 for each SNP obtained by conditioning on the other independent variants at that locus to compute the Bayes factor.

For each signal, we derived the 95% credible set of variants, constituting the minimal set of variants for which the sum of the posterior probabilities exceeds 95%. This approach was recommended for fine mapping of association signals and for prioritizing variants[30]. Assuming that there is only one causal variant for each 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 the functionality of the variants in each of the 95% credible sets.

Gene-based burden analysis. Single-variant analyses have limited power to detect 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[51] as implemented in EPACTS. These analyses 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, as we assumed that other variants (not protein altering) would far outnumber these predicted deleterious variants and would thus dilute disease burden due to the deleterious variants. Assuming negative selection against such deleterious variants that causes their frequency to be low across ancestry groups, 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 used 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 with the highest imputation quality ($R^2 \geq 0.8$).

We assessed statistical significance by adaptive permutation testing with variable thresholds (up to 100 million permutations; minimal P value = 1×10^{-8}) (ref. 51). If 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 signal for the already-identified variant. Therefore, we repeated the variable-threshold test with conditioning 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 across the genome, applying a genome-wide Bonferroni-corrected significance threshold of $0.05/17{,}044 = 2.9 \times 10^{-6}$ (17,044 genes across the genome with at least one variant included in the analysis, that is, 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 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. As there was an overlap of the sequenced subjects with the subjects for whom chip data were used, we conducted a sensitivity analysis for the burden test excluding overlapping subjects (**Supplementary Note**).

Follow-up queries for genes overlapping the association signals. To derive information for all genes overlapping our 52 identified association signals (spread across the 34 AMD loci), we built a gene list containing all genes that overlapped a more narrow definition of locus regions. We used a particularly comprehensive definition of the locus region during the signal identification step (index variants and proxies, $r^2 \geq 0.5$, ±500 kb), to avoid far-reaching LD that may generate shadow signals (particularly in 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 broad 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 multiple-testing corrections for the AMD locus-wide burden test search. However, this broad definition is less adequate when prioritizing genes around the identified signals under the assumption that most protein-altering or regulatory 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, $r^2 \geq 0.5$, ±100 kb) and yield 368 overlapping RefSeq genes (**Supplementary Data Set 5**).

Gene expression. For the 368 genes in our gene list, we sought to determine gene expression in relevant tissues—retina, RPE and choroid—in two independent data sets (see details in the **Supplementary Note**). A consensus rating of gene expression observed in the two laboratories performing the analysis was derived as follows. Expression of a gene in one set of tissues (retina or RPE/choroid) was inferred if both laboratories detected expression in the respective set of tissues; if at least one of the laboratories did not observe expression, the gene was considered to be not expressed. Gene expression of all other genes (with one laboratory observing expression and the other with missing data or where both laboratories had missing data) was regarded as missing.

Mouse model phenotypes. For the 368 genes in our gene list, we queried the Mouse Genome Informatics (MGI)[73] and International Mouse Phenotyping Consortium (IPMC)[74] databases (see URLs) and manually curated results by information from published literature. We determined whether a gene exhibited a relevant eye phenotype (retina, RPE or choroid phenotype) in established genetic mouse models (knockout, knock-in or transgenic mice).

Enrichment for molecular pathways. For the 368 overlapping genes, we performed functional enrichment analysis using INRICH[75] with default settings unless stated otherwise (**Supplementary Note**). The target intervals for this analysis were the narrow AMD locus regions (index variants and proxies, $r^2 \geq 0.5$ and ±100 kb; **Supplementary Table 5**). As there is no consensus approach to pathway analysis, we queried multiple databases: (i) the Kyoto Encyclopedia of Genes and Genomes (KEGG)[76], (ii) Reactome[77] and (iii) the Gene Ontology (GO) Consortium[78] (see URLs). For example, whereas KEGG is a manually curated database on metabolic pathways, the GO database also includes automatic annotations and a more comprehensive set of cellular processes and molecular functions.

Drug pathways and targets. To derive information on whether the product of a gene among the 368 genes in our gene list was a direct drug target, we

doi:10.1038/ng.3448

searched the DrugBank database (version 4.1), which contains 4,207 drug targets (genes) and 7,740 drugs[79] (see URLs).

Explained variability in disease liability. On the basis of the 52 identified AMD variants, we estimated the variance in disease liability explained by these variants (see URLs)[80] using the log-transformed 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.

Genetic risk score and relative and absolute genetic risk of AMD. For each individual, we computed a 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. To derive a realistic GRS distribution, we modeled a general population on the basis of our case-control data, with this model requiring an assumption of the prevalence of advanced AMD (**Supplementary Note**). For this modeled general population, we derived the GRS distribution and its deciles. For weighting, the log-transformed odds ratio for each of the 52 variants was derived from the fully adjusted model (including all 52 variants) to assure independence of the effect sizes.

We derived relative risk estimates (as odds ratios) for each GRS decile, with the first decile as a reference. This relative risk estimate is independent of disease prevalence, except that the decile used to form the genetic risk groups was based on the GRS distribution as expected in a general population (which requires a prevalence assumption). We also computed absolute risk estimates for each GRS decile as the proportion of advanced AMD cases, applying weights and prevalence assumptions as described above.

52. International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
53. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
54. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
55. Sherry, S.T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
56. Tennessen, J.A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012).
57. Age-Related Eye Disease Study Research Group. A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and E and β carotene for age-related cataract and vision loss: AREDS report no. 9. *Arch. Ophthalmol.* **119**, 1439–1452 (2001).
58. 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**, 2112–2118 (2012).
59. Pruitt, K.D. *et al.* RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.* **42**, D756–D763 (2014).
60. Ng, S.B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276 (2009).
61. 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).
62. Delaneau, O., Marchini, J. & Zagury, J.F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2012).
63. Ristau, T. *et al.* Allergy is a protective factor against age-related macular degeneration. *Invest. Ophthalmol. Vis. Sci.* **55**, 210–214 (2014).
64. 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**, 955–959 (2012).
65. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
66. Turner, S. *et al.* Quality control procedures for genome-wide association studies. *Curr. Protoc. Hum. Genet.* Chapter 1, Unit1.19 (2011).
67. Cavalli-Sforza, L.L. The Human Genome Diversity Project: past, present and future. *Nat. Rev. Genet.* **6**, 333–340 (2005).
68. 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).
69. Firth, D. Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27–38 (1993).
70. 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**, 539–550 (2013).
71. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
72. Stephens, M. & Balding, D.J. Bayesian statistical methods for genetic association studies. *Nat. Rev. Genet.* **10**, 681–690 (2009).
73. 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–D817 (2014).
74. Brown, S.D. & Moore, M.W. Towards an encyclopaedia of mammalian gene function: the International Mouse Phenotyping Consortium. *Dis. Model. Mech.* **5**, 289–292 (2012).
75. Lee, P.H., O'Dushlaine, C., Thomas, B. & Purcell, S.M. INRICH: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics* **28**, 1797–1799 (2012).
76. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
77. Croft, D. *et al.* The Reactome pathway knowledgebase. *Nucleic Acids Res.* **42**, D472–D477 (2014).
78. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
79. Law, V. *et al.* DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* **42**, D1091–D1097 (2014).
80. 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**, 310–317 (2011).