OXFORD

Human Molecular Genetics, 2024, Vol. 33, 24, 2085–2093
https://doi.org/10.1093/hmg/ddae141
Advance access publication date 26 September 2024
Original Article

Identifying X-chromosome variants associated with age-related macular degeneration

Michelle Grunin[1,2], Robert P. Igo Jr[1,2,†], Yeunjoo E. Song[1,2], Susan H. Blanton[3,4], Margaret A. Pericak-Vance[3,4], Jonathan L. Haines[1,2,*], International Age-related Macular Degeneration Genomics Consortium

[1]Population and Quantitative Health Sciences, Case Western Reserve University, 2103 Cornell Road, Cleveland, OH 44106, United States
[2]Cleveland Institute for Computational Biology, Case Western Reserve University, 2103 Cornell Road, Cleveland, OH 44106 United States
[3]Dr. John T Macdonald Department of Human Genetics, University of Miami School of Medicine, 1600 NW 10th Ave, Miami, FL 33136, United States
[4]The John P Hussman Institute for Human Genomics, University of Miami School of Medicine, Miami, 1600 NW 10th Ave, FL 33136, United States
*Corresponding author. Population and Quantitative Health Sciences, 2529 Wolstein Research Building, 2103 Cornell Road, Case Western Reserve University, Cleveland, OH 44106, United States. E-mail: jlh213@case.edu
†Robert P. Igo Jr is deceased.

Abstract

Purpose: In genome-wide association studies (GWAS), X chromosome (ChrX) variants are often not investigated. Sex-specific effects and ChrX-specific quality control (QC) are needed to examine these effects. Previous GWAS identified 52 autosomal variants associated with age-related macular degeneration (AMD) via the International AMD Genomics Consortium (IAMDGC), but did not analyze ChrX. Therefore¸ our goal was to investigate ChrX variants for association with AMD. Methods: We genotyped 29 629 non-Hispanic White (NHW) individuals (M/F:10404/18865; AMD12,087/14723) via a custom chip and imputed after ChrX-specific QC (XWAS 3.0) using the Michigan Imputation Server. Imputation generated 1 221 623 variants on ChrX. Age, informative PCs, and subphenotypes were covariates for logistic association analyses with Fisher's correction. Gene/pathway analyses were performed with VEGAS, GSEASNP, ICSNPathway, DAVID, and mirPath. Results: Logistic association on NHW individuals with sex correction identified variants in/near the genes *SLITRK4*, *ARHGAP6*, *FGF13* and *DMD* associated with AMD ($P < 1 \times 10^{-6}$,Fisher's combined-corrected). Association testing of the subphenotypes of choroidal neovascularization and geographic atrophy (GA), identified variants in *DMD* associated with GA ($P < 1 \times 10^{-6}$, Fisher's combined-corrected). Via gene-based analysis with VEGAS, several genes were associated with AMD ($P < 0.05$, both truncated tail strength/truncated product P) including *SLITRK4* and *BHLHB9*. Pathway analysis using GSEASNP and DAVID identified genes associated with nervous system development (FDR: P:0.02), and blood coagulation (FDR: P:0.03). Variants in the region of a microRNA (miR) were associated with AMD ($P < 0.05$, truncated tail strength/truncated product P). Via DIANA mirPath analysis, downstream targets of miRs showed association with brain disorders and fatty acid elongation ($P < 0.05$). A long noncoding RNA on ChrX near the *DMD* locus was also associated with AMD ($P = 4 \times 10^{-7}$). Epistatic analysis (t-statistic) for a quantitative trait of AMD vs control including covariates found a suggestive association in the XG gene ($P = 2 \times 10\text{^}-5$). Conclusions: Analysis of ChrX variation identifies several potential new locifor AMD risk and these variants nominate novel AMD pathways. Further analysis is needed to refine these results and to understand their biological significance and relationship with AMD development in worldwide populations.

Keywords: X chromosome; genome wide association study; age-related macular degeneration; non-coding

Introduction

Age-related macular degeneration (AMD) is the leading cause of blindness in the western world for those over the age of 60 [1]. The world's aging population is due to double in the next 40 years [2–4]. Currently, the global cost of AMD is around $343 billion, with yearly direct costs from the United States medical system at $575 million dollars and $38 665 per person [5]. Current treatments for AMD focus on late stage symptoms, particularly the use of anti-VEGF antibodies to slow and control the growth of pathological new blood vessels through the choroid into the retina [6–8].

The etiology of AMD is not fully understood, but age, smoking, and genetic variation, along with disorders such as hypertension, obesity, and diabetes, are contributing factors to this multifactorial condition [9–14]. The strongest known genetic risk factors include the *ARMS2/HTRA1* locus and genes in the complement system, such as *CFH* [15], *C2*, *C3*, *CFB* [16–18], *CFI* and *C9* [19]. Other genes and pathways are involved such as the *LIPC* gene in the lipid metabolism pathway, the VEGF pathway and receptor genes, and rare variants in multiple genes [20]. Currently, the largest dataset of AMD participants has been collected by the International Age-Related Macular Degeneration Genomics Consortium (IAMDGC), comprised of 24 centers worldwide who have collectively gathered 16 144 advanced AMD NHW participants and 17 832 NHW controls for genetic and phenotypic analysis Analyses of this dataset identified 52 common and rare variants in 34 loci across the autosomes associated with AMD pathogenesis [20]. However, only ∼60% of the heritability of AMD is accounted for by the known risk variants, so investigation into the "missing

Received: May 13, 2024. **Revised:** August 14, 2024. **Accepted:** September 16, 2024

heritability" of AMD will contribute greatly to the understanding of AMD pathogenesis.

Other studies have investigated transcription wide association (TWAS), protein expression in blood and retina, and analysis of microRNAs in AMD. However, the GWAS, TWAS, and other analyses performed did not include the entire genome, since the X chromosome (X) was excluded. Unfortunately, only a small subset of the thousands of GWAS performed across different traits include X [21–23], and most exclude X before analyzing, even if originally included in pre-quality control (PreQC) data [23, 24]. Those studies that did include X sometimes use methods that did not account for the uniqueness of X [25–27], including the obviously smaller male and female specific sample sizes, the impact of X inactivation in females vs males, reduced diversity on X, and sex-specific population structures, and of course association testing for autosomes incorrectly applied to X [21]. Fortunately, reanalysis and reinvestigation of X can be performed [28] now that methods are available to handle quality control (QC) of X, such as sex stratified effects, clarifying the sample size of both men and women in the study, and the different variants between sexes on control samples can be done [22, 27, 29, 30].

X, at 155 million base pairs, represents a significant portion (5%) of the nuclear genome, and harbors 10% of the causes of Mendelian disorders [21]. Studies have highlighted the role of sexual dimorphism in many diseases such as autoimmune disorders, cancer, and psychiatric disorders. There is possible enrichment for sexually antagonistic alleles that can contribute to disease risk, and a difference between males and females on the genes that have unique functions on X [21, 31–33]. X includes seven percent of the known microRNA (miRNA) and non-coding variants in the genome, which include long non-coding RNAs (lincRNA)s [34–36]. Approximately 16% of X-linked miRNA are implicated in immunity including those linked to the immune response and auto immune disease triggers. Dysregulation of the miRNA on X can contribute to some of these triggers [34–38]. Some autosomal miRNAs are associated with AMD. However, X-linked miRNAs have never been investigated [39–41]. Interaction between X-linked miRNAs and genes located on and off X exist for multiple complex diseases [37]. Shared areas between intronic miRNA and known protein coding genes on X can lead to clusters working in tandem for biological function [37, 38].

Previous studies investigated 150 AREDS patients and 1804 SNPs on X, and found a haplotype on X related with AMD [42]. This was a protective haplotype encompassing a 272 kb region and covered the gene *DIAPH2*. The hypothesis at the time was an instability in X chromosome inactivation due to Primary Ovarian Failure, with a connection to *DIAPH2* [43]. Another case study identified a recombinant X mutation in the *MECP2* gene due to a duplicated Xq28 terminal portion of the X chromosome, that caused a neurological phenotype and macular degeneration [44]. Winkler et al [45] reanalyzed the IAMDGC dataset and tested for sex-stratified effects on the autosomes, without finding differences amongst the lead variants by sex, despite having 80% power in each sex to detect variants of interest. AMD is more frequent in women than men. 65% of prevalent AMD cases in 2010 were female [46], and in 2019, the prevalence for late stage AMD was 0.88/million [0.53–1.35] for females and 0.6/million [0.36–0.91] for males [47]. As such, sex-specific analysis and chromosome-specific genetic analysis especially on large datasets are needed to fully unravel the genetic underpinnings of this disorder.

The IAMDGC GWAS [20] used data imputed with the available reference panel at the time, but was unable to impute X due to software limitations. Imputing the European ancestry participants/non-Hispanic white (NHW; originally defined as EUR) with a more comprehensive Haplotype Reference Consortium (HRC) reference panel, including X, significantly increases the power of the IAMDGC dataset as a resource and allows for greater identification of variants contributing to AMD risk. We focused on the NHW ancestry to be consistent with the IAMDGC GWAS [20]. Difficulties in X-specific analyses include the issue of X inactivation, quality control with X to prevent removal of X, sex-specific confounding, sex-specific population structures on X, and wrongfully applied autosomal testing methods to the X chromosome, which is present weighted by 2 in females as opposed to males [21, 27, 28]. All of these have never been applied to the IAMDGC dataset before this paper.

Therefore, we reanalyzed the IAMDGC dataset by first reimputing with the HRC to boost our power and coverage and then to analyze the X chromosome specifically to identify novel risk variants that could contribute to our understanding of AMD. We investigated X not only with methodology specific to the X chromosome utilizing the new imputation, but also more comprehensively by performing multiple analyses that incorporated X-specific methodology, including X-specific GWAS (XWAS) via variant and gene analysis, expression and eQTL analysis, non-coding RNA investigation including microRNA and lncRNA that comprise so much of the X chromosome, pathway and interaction analysis, and epistasis. This represents an exhaustive genetic analysis of the X chromosome in AMD.

Results

Single variant and gene-based analyses

Using logistic regression with sex correction, variants in or near the genes *SLITRK4*, *ARHGAP6*, *FGF13* and *DMD* along with intergenic regions nearby those genes, were nominally associated with AMD (all single variant testing in those genes were below a threshold of $P < 1 \times 10^{-6}$; Fisher's combined-corrected, gene-based testing < 0.05). One gene, *SLITRK4*, was Bonferroni-corrected significant. *ARHGAP6* and *DMD* were not significant in gene-based testing via VEGAS2 (Table 1). Two more genes (*MAGED1* and *BHLHB9*) showed significance only in gene based testing ($P < 0.05$) but not in single variant based testing (Table 1). All genes identified as significant ($P < 1 \times 10^{-6}$) were expressed in eye and may be involved in photoreceptor transduction or neurological disorders (Table 1). Spectacle data [48] was utilized to investigate expression in single cell datasets. Expression of each gene in different ocular cell types can be found in Supplementary Table 1. Pathway analysis using multiple programs (GSEASNP (GSEA4GWAS), DAVID, and ICSNPathway) after gene-based analysis on significant genes identified by VEGAS2 across all associated variants showed genes associated with nervous system development (FDR-P:0.02), and blood coagulation (FDR-P:0.03), both already known important pathways in AMD. In a paper published using IPA analysis for AMD, the nervous system, embryonic and organ development were the top three pathways found [49] and these pathways were found by our analysis as well. In addition, we also identified variants in the blood coagulation pathway, which was previously associated with response to VEGF in AMD patients [50]. DAVID pathway analysis of location based pathways (i.e. pathways that cluster in one tissue or location in the body, for example, eye, or blood cells) indicated clusters of significant genes located in brain or involved in locations involved in brain cancer. According to ICSNPathway analysis, the candidate causal SNP for the pathways identified was located in the gene GLUD2 (rs9697983) indicating the strength of that gene in the pathway found of electronic

Table 1. Variants and genes identified as associated with AMD ($P < 1 \times 10^{-6}$, Fisher's combined corrected) in chromosome X.

Gene	Chr:BP (hg37)	Number of SNPs included per gene for gene based testing (full set, including those in linkage disequalibrum)	Fisher's' corrected P-value (most significant SNP)	VEGAS P-value: gene-based (total SNP P-value)
SLITRK4	X:142723657	40	2.5 × 10–7	0.001
ARHGAP6	X:11563246	2363	1 × 10–6	$P > 0.05$
FGF13	X:137852982	3085	6.5 × 10–7	0.03
DMD	X: 31217428	11 712	2 × 10–6	$P > 0.05$
MAGED1	Gene based only	202	$P > 1 \times 10$–6	0.0004
BHLHB9	Gene based only	120	$P > 1 \times 10$–6	0.0009

Table 2. Pathway analysis for associated genes using three different methods.

Pathway Analysis	Cluster Name	FDR
DAVID	Cytoplasm	0.04
GSEA4GWAS	Endoplasmic reticulum network	0.01
	Nervous System Development	0.02
	Blood Coagulation	0.03
	Wound Healing/Platelets	0.03
ICSNPathway	Electron Transport (GO:0006118)	0.02
	Metabolism of Organic Acids, Amino Acids, Carboxylic Acids (GO:0006520, GO:0019752).	0.03

transport and metabolism (Table 2). This gene is also part of important metabolic pathways along with ABCA1 that were previously found in AMD [20]. GLUD2 is responsible for glutamate oxidative deamination, and connects to the same Lipoprotein metabolism pathway that was highlighted in genes like ABCA1, LIPS, and APOC2/APOE [51]. GLUD2 acts on Glu and conversts to alpha ketoglutarate. Since lipids are one of the main energy sources used, GLUD2 remains an important part of that pathway [52].

Heritabilty via GCTA was estimated at 40%(SE = 0.01), while when X was added, heritability was estimated at 41%(SE = 0.03). This suggests that the X chromosome may add modestly to overall heritability of AMD.

Epistatic analyses

Every single variant not in linkage disequilibrium on the X chromosome was tested (n = 164 976) for epistatic interactions. A suggestive epistatic interaction was found within X between a variant in the XG gene (X:2699555) and an intergenic variant (X:11868036). This epistatic interaction was differential between AMD patients and controls ($P = 2 \times 10^{-5}$, test according to [53, 54]). This suggestive epistatic interaction was detected in the absence of main effects. Data from the Harvard Transcriptome project on expression of RNA from AMD patients and controls [55] indicates that RNA in the intergenic region and the XG gene is expressed in AMD retina but not in control. To visualize eQTL or other expression traits or data, we examined the gTEX and Spectacle datasets to see if this interaction or expression was found in retina. According to the eye gTEX dataset [56], the area between the genes *MSL3* (X:11776278-11793872) and *FRMPD4* (X:12156585-12742642) (near *ARHGAP6*), which contains the intergenic region found above via epistasic analysis (lead variant at X: 11868036, rs186195192), have some expression in retina (general, not cell specific), but it is not highly enriched (n = 101 MGS1 retinas). The highest ranked variant does not have non-coding RNA nearby. XG was not mapped in that dataset at all, but would be between *ZBED1* and *CD99P1*, which both had some expression in retina but were not highly enriched. Investigation of all Spectacle [48] datasets for human retina indicate expression in retina in various cell types for the XG gene (Table 3).

Subphenotype gene based analyses

Sub-phenotype analysis found several genes with $P < 1 \times 10^{-6}$, Fisher's corrected in single variant testing. The same procedure was followed for sub-phenotype analyses as was in the original analysis. Analysis on CNV alone identified an association with the gene *PHEX* (4916 cases/7926 controls). When including the mixed samples (CNV + mixed) an association was found with *ARHGAP6* and *PTCHD1*-AS (5828 cases/7926 controls). GA alone was associated with *DMD*, and one variant in *DACH2* (1482 cases/7926 controls). For the GA + mixed, all significant SNPs found were intronic (2424 cases/7926 controls). The results reaffirm the general X results and indicate a trend of genes that are driving the respective XWAS significant p-values amongst the subphenotypes. (Supplementary Table 2 for exonic hits).

Non-coding RNA analyses

One long non-coding RNA (lncRNA), lnc-KDM6A1:2, was significantly associated ($P = 4 \times 10^{-7}$, Fisher's-corrected P-value). Lnc-KDM6A1:2 is 717 bp, and functions as antisense. It has been validated by both the ENCODE and HAVANA datasets. According to the HARVARD transcriptome [55] it is present in normal human retina via SAGE data. Mir584 was associated in gene based testing and is located near the *DMD* locus, which may be its target (RF0106, also known as AC090632.13/miRNA584-f), although the two loci are not in linkage disequilibrium. This miRNA still needs to be functionally validated in retina, but is present in normal human retina [55]. MiRNA5845p has a tumor suppressive function [60], whereas 584-3p inhibits gastric cancer progression [61]. Low levels of this miRNA are linked to poor prognosis in T-cell

Table 3. Spectacle datasets for human retina for the XG gene expression.

Dataset	Species	Cell Type	Expression Level
Yan et al. [57]	Human	Endothelium	Highly Expressed
Yan et al. [57]	Human	Microglia	Partially Expressed
Lu et al. [58]	Human	Bipolar Cells	Highly Expressed
Lu et al. [58]	Human	Muller glia, retinal ganglion cells, amacrine cells	Partially Expressed
Orozco et al. [59]	Human	Vascular cells	Highly Expressed

Table 4. microRNA regions in chromosome X associated with AMD (VEGAS2, P < 0.05, truncated tail strength and truncated product *P*).

miRNA	Chr:BP (hg37)	P-value
miR-220	X:122695945-122696055	0.03
miR-652	X:10939960-10940057	0.003
miR-548	X:22191615-22191699	0.001
miR-584	X:31356817-31356938	0.04
miR-934	X:135633036-135633119	0.004
miR-2114	X:149396238-149369318	0.04
miR-509-1	X:146340277-146341244	0.05

Table 5. Pathway analysis of downstream microRNA targets using DIANA mirPath.

KEGG Pathway	P-value	Num. of Genes	Num. of miRs
Prion Diseases	3.7310^{-31}	2	3
Huntington's Disease	0.003	13	4
Hippo signaling	0.004	13	5
Splicesome	0.01	15	6
Endoplasmic reticulum processing	0.01	17	6
Fatty acid elongation	0.03	3	2
mRNA surveillance	0.03	13	5
Proteoglycans in cancer	0.04	17	5
Parkinson's Disease	0.04	7	4

lymphoma [62]. All miRNA and lncRNA were checked via BLAST to confirm location and stem-loop sequence. It is part of the stringent set of the LNCpedia database and can be found across Ensembl as well.

Several variants in the region of a microRNA (miRNA/miR) were associated with AMD (VEGAS2, $P < 0.05$, truncated tail strength and truncated product P, Table 4). Only the variants with the strongest association signals ($P < 1 \times 10^{-6}$) are included in the gene-based test, allowing a lesser p-value for gene-based associations. The locations of these miRNA are near *PHEX* and *DMD*. These miRNAs may directly influence nearby genes. miRNA were scanned by both the 3p and the 5p to determine their sequence and location. Via ENCODE data, these miRNAs were near some regulatory elements that regulate areas in the cerebrum and embryonic cell development (location: X:1427222).

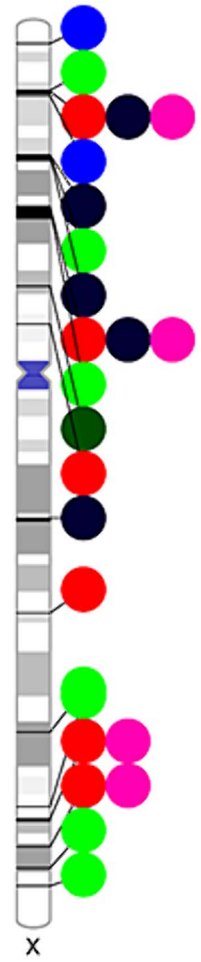
To help in understanding and summarizing of the X-chromosome results above as to where the genes are located for the main, sub-phenotype, and non-coding, and epistatic hits, we provide an explanatory figure with their locations across chromosome X (Fig. 1).

To determine the predicted function for the miRNAs initially identified via VEGAS2, DIANA mirPath analysis was utilized to identify downstream targets of these miRs using KEGG pathways (Table 5). Results show association with brain disorders and fatty acid elongation ($P < 0.05$), particularly for prion diseases (gene targets *MAP2K2* and *PRNP*) and Huntington's disease (gene targets *HTT*, *CLTA*, *NDUFS1*, *COX4I1*, *EP300*, *SOD2*, *TBP*, and *UQCRB*).

Transcription factor motif analysis was performed using the MEME suite of programs. Motifs were investigated in genes that were associated with AMD via VEGAS2 gene-based testing. This was performed with an unbiased analysis, one or more motifs found in sequences, and a maximum of 5 motifs per sequence with $P < 1 \times 10^{-7}$. DREME motif analysis for short motifs was also performed to investigate an E-value < 0.01 (considered significant as the significance threshold for motifs is E = 0.05), but none were significant. GOMO analysis (part of the MEME suite of analysis) on promoters was performed on GO terms where the q-values were < 0.05. Once motifs were identified, the HOCOMOCO database for human-specific motifs associated with transcription factors was utilized with TOMTOM to identify motifs that were associated with transcription factors found in multiple genes associated with AMD. This was compared to random with over 10 permutations to clarify the null hypothesis. An association to a motif found in multiple sequences from genes associated with AMD was found for the transcription factors IRF3 (interferon), PRDM6 (involved in histone activity), and ZFP82 (involved in DNA binding and zinc-finger activity), along with others of lesser significance (Table 6).

Discussion

In previous analyses of AMD, the X chromosome was not included despite its many known roles in human biology, although its companion, the Y chromosome and its association with age and



SNP	chr	pos	endpos	Phenotype
XG	X	2699555	2699555	Epistatic
miR-652	X	10939960	10940057	miRNA
ARHGAP6	X	11136239	11683821	GeneBased
ARHGAP6	X	11136239	11683821	Subphenotype
ARHGAP6	X	11563246	11563246	TopSNP
Intergenic	X	11868036	11868036	Epistatic
PHEX	X	22050559	22269427	Subphenotype
miR-548	X	22191615	22191699	miRNA
PTCHD1-AS	X	22277914	23311263	Subphenotype
DMD	X	31115794	33357558	GeneBased
DMD	X	31115794	33357558	Subphenotype
DMD	X	31217428	31217428	TopSNP
miR-584	X	31356817	31356938	miRNA
Lnc-KDM6A1:2	X	45042496	45193162	lncRNA
MAGED1	X	51546103	51645453	GeneBased
DACH2	X	85403462	86087607	Subphenotype
BHLHB9	X	101975616	102008468	GeneBased
miR-220	X	122695945	122696055	miRNA
miR-934	X	135633036	135633119	miRNA
FGF13	X	137713735	138304939	GeneBased
FGF13	X	137852982	137852982	TopSNP
SLIRTK4	X	142710595	142723926	GeneBased
SLIRTK4	X	142723657	142723657	TopSNP
miR-509-1	X	146340277	146341244	miRNA
miR-2114	X	149396238	149369318	miRNA

Figure 1. An explanatory figure looking at the locations of chromosomal, subphenotype, non-coding, and epistatic hits across chromosome X.

Table 6. Transcription factor motifs found to be present in genes associated with AMD on the X chromosome.

Transcription Factor	P-value	E-value	Motif Overlap
IRF3	$9.59 \times 10^{\wedge -7}$	3.86×10^{-4}	20
PRDM6	$3.12 \times 10^{\wedge -6}$	1.25×10^{-3}	13
ZFP82	$9.02 \times 10^{\wedge -6}$	3.63×10^{-3}	24
IRF1	$1.55 \times 10^{\wedge} -5$	6.25×10^{-3}	20
BC11A	$1.73 \times 10^{\wedge -5}$	6.97×10^{-3}	17
NFAC1	$2.18 \times 10^{\wedge -5}$	8.75×10^{-3}	15
SPIB	$5.01 \times 10^{\wedge -5}$	2.02×10^{-2}	17
SPI1	$7.21 \times 10^{\wedge -5}$	2.90×10^{-2}	17
IRF2	$7.75 \times 10^{\wedge -5}$	3.11×10^{-2}	20
STAT2	8.67×10^{-05}	3.49×10^{-02}	19
IRF8	9.38×10^{-05}	3.77×10^{-02}	20
ZN394	2.10×10^{-04}	8.43×10^{-02}	20
GATA3	2.60×10^{-04}	1.05×10^{-01}	11
ETS2	2.93×10^{-04}	1.18×10^{-01}	13
SOX2	3.43×10^{-04}	1.38×10^{-01}	13
FLI1	4.55×10^{-04}	1.83×10^{-01}	18
ANDR	6.56×10^{-04}	2.64×10^{-01}	18

AMD has already been studied [63]. We hypothesized that X chromosome variants are associated with AMD pathogenesis, and that these variants may identify to novel pathways. We found multiple novel loci that associate with AMD status on the X chromosome, including some found in the non-coding X genome. Therefore, although the X chromosome has been consistently overlooked in GWAS, its importance cannot be.

We further explored the potential role of miRNAs and epistatic interactions between loci on the X chromosome and other loci that may contribute to AMD that have not yet been identified. The genes identified include some that are associated with other retinal disorders or are regulators of other pathways like the non-coding RNA identified.

Pathways identified include both neuronal and eye/retina pathways, as well as lesser studied pathways like wound healing and platelet function [64]. These analyses, both gene and non-coding RNA based, as well as incorporating both pathway and epistatic interactions, provide a novel and important piece in the genetic puzzle of AMD that has not been examined until now. Many of these genes may represent novel targets, and the interactions of these miRNA and lncRNA with the genes on the X chromosome, as well as other genes across the genome has not been studied fully.

The genes identified were all expressed in retina. *DMD* is the protein for dystrophin, mutations in which cause Duchenne muscular dystrophy (DMD); in the retina it is involved in signaling and is located in photoreceptor terminals, retinal neurons, cone/rod synapses, and other parts of the photoreceptor synaptic complexes [65]. It is involved in the optical neuron-bipolar pathways as well as the plasma membrane in Muller cells [66]. Most patients with DMD have some abnormal ERG results, including "negative" scotopic ERGs, especially those with deletion of the *DMD* gene. There are also differential isoforms based on the various mutations/deletions, which associate with phenotype. This indicates that dystrophin continues to play a role in phototransduction [65, 67]. In addition, *HTRA1*, one of the major genetic risk factors for AMD, is upregulated in DMD and may play a role in cell growth, influencing abnormal growth in the disease or affected cell growth or tolerance for repair.

ARHGAP6 is a rhoGAP family member. The activation of the enzyme GTPAse protein with a specificity for rhoA and a cytoskeletal protein for actin remodeling. *SLITRK4* is an axonal growth controlling protein. *FGF13* is a fibroblast growth factor, which is involved in mitogenic and cell survival, tissue repair, morphogenesis and cell growth. All of these genes were expressed in retina according to the HARVARD Retinal Transcriptome and could be novel targets for either drug therapy or in how they regulate the other AMD known loci. No known association was found with DIAPH2 in our current study, although the original study found it only in a small proportion of patients.

There was an association found with miRNA targeting genes involved in prion diseases in the pathway analysis (hsa05020). Prion diseases have a subset of "pathogenic" non-coding RNAs that are involved in the disorder, and these same miRNA are involved in AMD and Alzheimer disease (AD) [68, 69]. In addition, the protein aggregation processes in AMD and AD have a similarity to prion disease, and some classify them as 'prion- like' diseases. All prion-like diseases cause retinal damage in humans, like the plaques of amyloid-beta found in AD, but also in AMD [70], and therefore, they can be thought of as related to each other, not just in prion disease itself which has its own photoreceptor degradation, but in the pathogenesis involved in these diseases [71].

This study represents the first study performed on the full X chromosome in AMD and includes a novel lncRNA association as well as genomic variants nominally associated with AMD pathogenesis. Further investigation is needed to clarify the roles of these identified genes and loci. This could involve via silencing/knockdown or overexpression of miRNA/lncRNA in RPE cell lines to evaluate function and survival [39, 72, 73], or utilizing the MERFISH technique for localization in situ inside the cell for postranscriptional regulation [74, 75]. Further analysis is needed to confirm these results and to understand their biological significance and relationship with development of AMD in worldwide populations.

Methods

The IAMDGC SNP array data has ~250 K tagging and ~250 K rare/-common variants, resulting in a starting point for pre-imputation of 569 645 variants genome wide. These were filtered according to previous methodology [20]. Briefly, only the individuals with a known phenotype (geographic atrophy (GA), neovascular AMD (nvAMD) as well as ophthalmalogically examined controls, along with NHW status based on the previous population stratification analysis [20] were used in this analysis. Pre/post imputation pipelines were developed for an X-specific GWAS, and analysis was performed with Plink 1.9 and XWAS 3.0, along with R and Bioconductor scripts. 18 865 female and 10 404 male IAMDGC samples were used (12 087 control/14273 AMD), which included CNV/GA/Mixed as 4916/1482/942. Samples were removed if they did not pass QC via the XWAS standard pipeline. In addition the Beaver Dam samples were removed from the IAMDGC original, as well as whole genome amplified samples, which did not impute well enough via the HRC imputation panel. All samples included were of European descent, identified in the original IAMDGC analysis as EUR (Supplementary Table 3).

The Haplotype Reference Consortium panel (HRC) 1.1 dataset was used for the imputation reference panel. It contained 64 976 haplotypes including X. The Michigan Imputation Server (MIS) was utilized taking into account X-specific imputation. ShapeIT was used for pre-phasing, data on X was split by non-pseudoautosomal (non-PAR) and PAR regions. QC was performed via sex, for non-PAR and we excluded samples that did not match (n = 56). Heterozygote genotypes were evaluated. Three files were developed: the non-PAR female, PAR, and non-PAR male. Imputation was performed using minimac3.

The final variant count across the genome was 569 645 before imputation; only 6411 variants were present on X. The QC pre-imputation included a new identity by descent calculation (IBD), removal of the heterozygote haploid genotypes, Hardy–Weinberg equilibrium (HWE) was calculated separately for males and females in controls and the variants with a threshold of $P < 5.9 \times 10^{-8}$ were removed. Missing and frequency testing for removing variants was performed according to standard approaches (GENO > 0.1, MAF < 0.0005) and variants removed, along with removal of samples with missing phenotypes. Three SNPs failed the sex frequency testing at a threshold for sex differences of $P = 7.56 \times 10^{-6}$, and were removed.

X was then imputed to 1 221 623 variants via the MIS and HRC 1.1.

Post-imputation processing sensitive to sex occurred in similar steps to pre-imputation processing with similar cutoffs including HWE, missingness, frequency, and sex-specific details, along with the QUAL threshold. There were still some SNPs with a different frequency amongst sexes in controls with $P < 5.9 \times 10^{-8}$. After

post-imputation processing the final number of variants on X was 597 585, and the final number of samples for the X analysis was 29 269 (M/F:10404/18865).

For analysis, informative PCs, age, and sub-phenotypes were taken into account for all downstream analyses. Informative PCs were determined by a threshold above eigenval 19 (PC 1,2,3,4,5,6,7), calculated with Plink 1.9. The population structure on X captures a 1:2 male/female contribution, which is different than the autosomes. However, Chang [21] evaluated both PCs taken from X and from autosomes, and determined that correction is more appropriate via the autosomes because of the small number of SNPs on X relative to the autosomes. Stouffers and Fisher's correction were both used for analysis; results were similar.

Logistic regression was performed for a standard XWAS (X-wide association study), including covariates, and was sex-stratified. Logistic association was performed with sex correction, using both Fisher's and Stouffers correction, including age and informative PCs as covariates. The significance threshold was set at 0.05/the number of independent tests (pruned variants to account for linkage disequilibrium, final n = 164 976 variants) for a Bonferroni correction = 3.03×10^{-7}. To look at X interaction, epistatic analysis was performed. The test used was a modified epistatic test that uses a t-statistic for a quantitative trait (XWAS 3.0).

VEGAS2, a part of the XWAS suite, was utilized for gene-based testing using both the truncated tail strength and the truncated product p-value testing. The suggestive P-value threshold was set at < 0.05 for gene based testing, whereas the full significance value was set at 6.2×10^{-5} (genes in the X chromosome n = 804, $0.05/804 = 6.2 \times 10^{-5}$). The threshold for significant SNPs to be included was set at 0.0001 for gene-based analysis. We modeled males to be as equivalent to female homozygotes. In addition, we performed a sub-phenotype analysis, with GA alone, CNV alone, and a mixed analysis, where the mixed phenotype samples were added to GA and CNV separately.

Heritability analyses were performed with GCTA. Briefly, the 34 loci with boundaries were taken as described in Waksmunski et al [76], and then a genetic relationship matrix was calculated for the loci themselves, and then the loci that we identified in the X chromosome was added to the relationship matrix. This allowed us to understand heritability with and without the X chromosome.

Pathway analysis programs were utilized to identify pathways of interest involving the most significant genes. Pathway analysis programs GSEASNP [77] and ICSNPathway [78] (http://icsnpathway.psych.ac.cn/) were used to test if a disease phenotype is influenced by genes enriched in a signaling pathway. GSEA-SNP modifies the original GSEA program that only took into account genes rather than SNPs of interest. We also utilized DAVID [79], which takes into account genes that are found in clusters incorporating both GO and KEGG terms, as well as pathway and disease connections (e.g. OMIM). These multiple programs were used to ascertain if genes/SNPs would overlap utilizing multiple methods, after performing pathway analysis.

Transcription factor motif analysis was performed using MEME, and the MEME suite of related programs [80].

Author contributions

M.G., R.P.I. Jr, Y.S. analyzed data. M.G. and J.L.H. drafted and critically reviewed the manuscript. Y.S., S.B., M.P.-V., J.L.H. developed the ideas, advised on results, and critically reviewed the manuscript.

Supplementary data

Supplementary data is available at *HMG Journal* online.

Funding

IAMDGC: NIH 1X01HG006934-01 and RO1 EY022310. Michelle Grunin: M2021006F from the Bright Focus Fellowship for Macular Degeneration.

Conflict of interest statement: There are no other competing interests to declare.

Data availability

The genotype data analyzed during the current study were generated by the IAMDGC and are available through the database of Genotypes and Phenotypes (dbGAP; Accession: phs001039.v1.p1). Summary statistics for the IAMDGC data is available currently at http://amdgenetics.org.

Ethical approval

The IRB of Case Western Reserve University—University Hospitals (IRB Number EM-14-04) gave ethical approval for this work.

The study participants were previously ascertained by IAMDGC cohorts as described in Fritsche et al. 2016, Nature Genetics. All participants provided informed consent, and the study was approved by institutional review boards as previously described.

References

1. Friedman DS, O'Colmain BJ, Muñoz B. *et al.* Prevalence of age-related macular degeneration in the United States. *Arch Ophthalmol* 2004;**122**:564–572.
2. Resnikoff S, Pascolini D, Etya'ale D. *et al.* Global data on visual impairment in the year 2002. *Bull World Heal Organ* 2004;**82**:844–851.
3. Avisar R, Friling R, Snir M. *et al.* Estimation of prevalence and incidence rates and causes of blindness in Israel, 1998–2003. *Isr Med Assoc J* 2006;**8**:880–881.
4. Rein DB, Wittenborn JS, Zhang X. *et al.* Forecasting age-related macular degeneration through the year 2050: the potential impact of new treatments. *Arch Ophthalmol* 2009;**127**:533–540.
5. Marques AP, Ramke J, Cairns J. *et al.* The economics of vision impairment and its leading causes: a systematic review. *EClinicalMedicine* 2022;**46**:101354.
6. Gale R, Korobelnik JF, Yang Y. *et al.* Characteristics and predictors of early and delayed responders to ranibizumab treatment in neovascular age-related macular degeneration: a retrospective analysis from the ANCHOR, MARINA, HARBOR, and CATT trials. *Ophthalmologica* 2016;**236**:193–200.
7. Martin DF, Maguire MG, Ying GS. *et al.* Ranibizumab and bevacizumab for neovascular age-related macular degeneration. *N Engl J Med* 2011;**364**:1897–1908.
8. Schmidt-Erfurth U, Eldem B, Guymer R. *et al.* Efficacy and safety of monthly versus quarterly ranibizumab treatment in neovascular age-related macular degeneration: the EXCITE study. *Ophthalmology* 2011;**118**:831–839.
9. Klein ML, Mauldin WM, Stoumbos VD. Heredity and age-related macular degeneration. Observations in monozygotic twins. *Arch Ophthalmol* 1994;**112**:932–937.

10. Deangelis MM, Silveira AC, Carr EA. *et al.* Genetics of age-related macular degeneration: current concepts, future directions. *Semin Ophthalmol* 2011;**26**:77–93.
11. Yonekawa Y, Andreoli C, Miller JB. *et al.* Conversion to aflibercept for chronic refractory or recurrent neovascular age-related macular degeneration. *Am J Ophthalmol* 2013;**156**:29–35.e2.
12. Edwards AO, Ritter R 3rd, Abel KJ. *et al.* Complement factor H polymorphism and age-related macular degeneration. *Science* 2005;**308**:421–424.
13. Ferris FL, Fine SL, Hyman L. Age-related macular degeneration and blindness due to neovascular maculopathy. *Arch Ophthalmol* 1984;**102**:1640–1642.
14. Wang JJ, Ross RJ, Tuo J. *et al.* The LOC387715 polymorphism, inflammatory markers, smoking, and age-related macular degeneration. A population-based case-control study. *Ophthalmology* 2008;**115**:693–699.
15. Haines JL, Hauser MA, Schmidt S. *et al.* Complement factor H variant increases the risk of age-related macular degeneration. *Science* 2005;**308**:419–421.
16. Francis PJ, Hamon SC, Ott J. *et al.* Polymorphisms in C2, CFB and C3 are associated with progression to advanced age related macular degeneration associated with visual loss. *J Med Genet* 2009;**46**:300–307.
17. Gold B, Merriam JE, Zernant J. *et al.* Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration. *Nat Genet* 2006;**38**:458–462.
18. Yates JR, Sepp T, Matharu BK. *et al.* Complement C3 variant and the risk of age-related macular degeneration. *N Engl J Med* 2007;**357**:553–561.
19. Seddon JM, Yu Y, Miller EC. *et al.* Rare variants in CFI, C3 and C9 are associated with high risk of advanced age-related macular degeneration. *Nat Genet* 2013;**45**:1366–1370.
20. Fritsche LG, Igl W, Bailey JNC. *et al.* A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat Genet* 2016;**48**:134–143.
21. Chang D, Gao F, Slavney A. *et al.* Accounting for eXentricities: analysis of the X chromosome in GWAS reveals X-linked genes implicated in autoimmune diseases. *PLoS One* 2014;**9**:e113684.
22. König IR, Loley C, Erdmann J. *et al.* How to include chromosome X in your genome-wide association study. *Genet Epidemiol* 2014;**38**:97–103.
23. Wise AL, Gyi L, Manolio TA. EXclusion: toward integrating the X chromosome in genome-wide association analyses. *Am J Hum Genet* 2013;**92**:643–647.
24. Mailman MD, Feolo M, Jin Y. *et al.* The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet* 2007;**39**:1181–1186.
25. Hammer MF, Mendez FL, Cox MP. *et al.* Sex-biased evolutionary forces shape genomic patterns of human diversity. *PLoS Genet* 2008;**4**:e1000202.
26. Keinan A, Reich D. Can a sex-biased human demography account for the reduced effective population size of chromosome X in non-Africans? *Mol Biol Evol* 2010;**27**:2312–2321.
27. Gao F, Chang D, Biddanda A. *et al.* XWAS: a software toolset for genetic data analysis and association studies of the X chromosome. *J Hered* 2015;**106**:666–671.
28. Bonàs-Guarch S, Guindo-Martínez M, Miguel-Escalada I. *et al.* Re-analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes. *Nat Commun* 2018;**9**:321.
29. Ma L, Hoffman G, Keinan A. X-inactivation informs variance-based testing for X-linked association of a quantitative trait. *BMC Genomics* 2015;**16**:241.
30. Arbiza L, Gottipati S, Siepel A. *et al.* Contrasting X-linked and autosomal diversity across 14 human populations. *Am J Hum Genet* 2014;**94**:827–844.
31. Voskuhl R, Itoh Y. The X factor in neurodegeneration. *J Exp Med* 2022;**219**:e20211488.
32. Lerner DJ, Kannel WB. Patterns of coronary heart disease morbidity and mortality in the sexes: a 26-year follow-up of the Framingham population. *Am Heart J* 1986;**111**:383–390.
33. Muscat JE, Richie JPJ, Thompson S. *et al.* Gender differences in smoking and risk for oral cancer. *Cancer Res* 1996;**56**:5192–5197.
34. Ji Y, Hocker JD, Gattinoni L. Enhancing adoptive T cell immunotherapy with microRNA therapeutics. *Semin Immunol* 2016;**28**:45–53.
35. Hewagama A, Gorelik G, Patel D. *et al.* Overexpression of X-linked genes in T cells from women with lupus. *J Autoimmun* 2013;**41**:60–71.
36. Pinheiro I, Dejager L, Libert C. X-chromosome-located microRNAs in immunity: might they explain male/female differences? *BioEssays* **33**:791–802.
37. Khalifa O, Pers YM, Ferreira R. *et al.* X-linked miRNAs associated with gender differences in rheumatoid arthritis. *Int J Mol Sci* 2016;**17**:1852.
38. Kukurba KR, Parsana P, Balliu B. *et al.* Impact of the X chromosome and sex on regulatory variation. *Genome Res* 2016;**26**:768–777.
39. Grassmann F, Schoenberger PGA, Brandl C. *et al.* A circulating MicroRNA profile is associated with late-stage Neovascular age-related macular degeneration. *PLoS One* 2014;**9**:e107461.
40. Ren C, Liu Q, Wei Q. *et al.* Circulating miRNAs as potential biomarkers of age-related macular degeneration. *Cell Physiol Biochem* 2017;**41**:1413–1423.
41. ElShelmani H, Brennan I, Kelly DJ. *et al.* Differential circulating MicroRNA expression in age-related macular degeneration. *Int J Mol Sci* 2021;**22**:12321.
42. Jiang R, Dong J, Joo J. *et al.* Simple strategies for haplotype analysis of the X chromosome with application to age-related macular degeneration. *Eur J Hum Genet* 2011;**19**:801–806.
43. Vladan B, Biljana SP, Mandusic V. *et al.* Instability in X chromosome inactivation patterns in AMD: a new risk factor? *Med Hypothesis Discov Innov Ophthalmol* 2013;**2**:74–82.
44. Magini P, Poscente M, Ferrari S. *et al.* Cytogenetic and molecular characterization of a recombinant X chromosome in a family with a severe neurologic phenotype and macular degeneration. *Mol Cytogenet* 2015;**8**:58.
45. Winkler TW, Brandl C, Grassmann F. *et al.* Investigating the modulation of genetic effects on late AMD by age and sex: lessons learned and two additional loci. *PLoS One* 2018;**13**:e0194321.
46. Pennington KL, DeAngelis MM. Epidemiology of age-related macular degeneration (AMD): associations with cardiovascular disease phenotypes and lipid factors. *Eye Vis (Lon)* 2016;**3**:34.
47. Rein DB, Wittenborn JS, Burke-Conte Z. *et al.* Prevalence of age-related macular degeneration in the US in 2019. *JAMA Ophthalmol* 2022;**140**:1202–1208.
48. Voigt AP, Whitmore SS, Lessing ND. *et al.* Spectacle: an interactive resource for ocular single-cell RNA sequencing data analysis. *Exp Eye Res* 2020;**200**:108204.
49. Zhang J, Zhou Y. Identification of key genes and pathways associated with age-related macular degeneration. *J Ophthalmol* 2020;**2020**:2714746.
50. Grunin M, Beykin G, Rahmani E. *et al.* Association of a variant in VWA3A with response to anti-vascular endothelial growth factor treatment in neovascular AMD. *Investig Ophthalmol Vis Sci* 2020;**61**:68.

51. Ahmad F, Sun Q, Patel D. *et al.* Cholesterol metabolism: a potential therapeutic target in glioblastoma. *Cancers (Basel)* 2019;**11**:146.
52. Wanjari UR, Mukherjee AG, Gopalakrishnan AV. *et al.* Role of metabolism and metabolic pathways in prostate cancer. *Meta* 2023;**13**:183.
53. Ueki M, Cordell HJ. Improved statistics for genome-wide interaction analysis. *PLoS Genet* 2012;**8**:e1002625.
54. Wan X, Yang C, Yang Q. *et al.* BOOST: a fast approach to detecting gene-gene interactions in genome-wide case-control studies. *Am J Hum Genet* 2010;**87**:325–340.
55. Farkas MH, Grant GR, White JA. *et al.* Transcriptome analyses of the human retina identify unprecedented transcript diversity and 3.5 Mb of novel transcribed sequence via significant alternative splicing and novel genes. *BMC Genomics* 2013;**14**:486.
56. Ratnapriya R, Sosina OA, Starostik MR. *et al.* Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. *Nat Genet* 2019;**51**:606–610.
57. Yan W, Peng Y-R, van Zyl T. *et al.* Cell atlas of the human fovea and peripheral retina. *Sci Rep* 2020;**10**:9802.
58. Lu Y, Shiau F, Yi W. *et al.* Single-cell analysis of human retina identifies evolutionarily conserved and species-specific mechanisms controlling development. *Dev Cell* 2020;**53**:473–491.e9.
59. Orozco LD, Chen H-H, Cox C. *et al.* Integration of eQTL and a single-cell atlas in the human eye identifies causal genes for age-related macular degeneration. *Cell Rep* 2020;**30**:1246–1259.e6.
60. Xiang X, Mei H, Qu H. *et al.* miRNA-584-5p exerts tumor suppressive functions in human neuroblastoma through repressing transcription of matrix metalloproteinase 14. *Biochim Biophys Acta* 2015;**1852**:1743–1754.
61. Zheng L, Chen Y, Ye L. *et al.* miRNA-584-3p inhibits gastric cancer progression by repressing yin Yang 1- facilitated MMP-14 expression. *Sci Rep* 2017;**7**:8967.
62. Wang C-C, Han L, Hou Y-H. *et al.* MiRNA-584 suppresses the progression of NK/T-cell lymphoma by targeting FOXO1. *Eur Rev Med Pharmacol Sci* 2020;**24**:4404–4411.
63. Grassmann F, Weber BHF, Veitia RA. Insights into the loss of the Y chromosome with age in control individuals and in patients with age-related macular degeneration using genotyping microarray data. *Hum Genet* 2020;**139**:401–407.
64. Waksmunski AR, Grunin M, Kinzy TG. *et al.* Pathway analysis integrating genome-wide and functional data identifies PLCG2 as a candidate gene for age-related macular degeneration. *Investig Ophthalmol Vis Sci* 2019;**60**:4041–4051.
65. Ricotti V, Jägle H, Theodorou M. *et al.* Ocular and neurodevelopmental features of Duchenne muscular dystrophy: a signature of dystrophin function in the central nervous system. *Eur J Hum Genet* 2016;**24**:562–568.
66. Schmitz F, Drenckhahn D. Dystrophin in the retina. *Prog Neurobiol* 1997;**53**:547–560.
67. Sigesmund DA, Weleber RG, Pillers DA. *et al.* Characterization of the ocular phenotype of Duchenne and Becker muscular dystrophy. *Ophthalmology* 1994;**101**:856–865.
68. Zhao Y, Jaber V, Alexandrov PN. *et al.* microRNA-based biomarkers in Alzheimer's disease (AD). *Front Neurosci* 2020;**14**:585432.
69. Lukiw WJ, Vergallo A, Lista S. *et al.* Biomarkers for Alzheimer's disease (AD) and the application of precision medicine. *J Pers Med* 2020;**10**:138.
70. Ratnayaka JA, Serpell LC, Lotery AJ. Dementia of the eye: the role of amyloid beta in retinal degeneration. *Eye (Lond)* 2015;**29**:1013–1026.
71. Striebel JF, Race B, Leung JM. *et al.* Prion-induced photoreceptor degeneration begins with misfolded prion protein accumulation in cones at two distinct sites: cilia and ribbon synapses. *Acta Neuropathol Commun* 2021;**9**:17.
72. Kuhn DE, Martin MM, Feldman DS. *et al.* Experimental validation of miRNA targets. *Methods* 2008;**44**:47–54.
73. Gustafson D, Tyryshkin K, Renwick N. microRNA-guided diagnostics in clinical samples. *Best Pract Res Clin Endocrinol Metab* 2016;**30**:563–575.
74. Moffitt JR, Hao J, Wang G. *et al.* High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc Natl Acad Sci* 2016;**113**:39.
75. Wang G, Moffitt JR, Zhuang X. Multiplexed imaging of high-density libraries of RNAs with MERFISH and expansion microscopy. *Sci Rep* 2018;**8**:4847.
76. Waksmunski AR, Grunin M, Kinzy TG. *et al.* Statistical driver genes as a means to uncover missing heritability for age-related macular degeneration. *BMC Med Genet* 2020;**13**:95.
77. Holden M, Deng S, Wojnowski L. *et al.* GSEA-SNP: applying gene set enrichment analysis to SNP data from genome-wide association studies. *Bioinformatics* 2008;**24**:2784–2785.
78. Zhang K, Chang S, Cui S. *et al.* ICSNPathway: identify candidate causal SNPs and pathways from genome-wide association study by one analytical framework. *Nucleic Acids Res* 2011;**39**:W437–W443.
79. Huang DW, Sherman BT, Tan Q. *et al.* DAVID bioinformatics resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res* 2007;**35**:e101.
80. Bailey TL, Boden M, Buske FA. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 2009;**37**:W202–W208.