Ancestry of the Timorese: age-related macular degeneration associated genotype and allele sharing among human populations from throughout the world

Margaux A. Morrison[1], Tiago R. Magalhaes[2,3], Jacqueline Ramke[4], Silvia E. Smith[1], Sean Ennis[3,5], Claire L. Simpson[6], Laura Portas[6,7], Federico Murgia[6,7], Jeeyun Ahn[8,9], Caitlin Dardenne[1], Katie Mayne[1], Rosann Robinson[1], Denise J. Morgan[1], Garry Brian[4], Lucy Lee[4,10], Se J. Woo[8,11], Fani Zacharaki[12], Evangelia E. Tsironi[12], Joan W. Miller[13], Ivana K. Kim[13], Kyu H. Park[8,11], Joan E. Bailey-Wilson[6], Lindsay A. Farrer[14], Dwight Stambolian[15] and Margaret M. DeAngelis[1]*

[1] Ophthalmology and Visual Sciences, John A. Moran Eye Center, University of Utah, Salt Lake City, UT, USA, [2] National Children's Research Centre, Our Lady's Children's Hospital, Dublin, Ireland, [3] Academic Centre on Rare Diseases, School of Medicine and Medical Science, University College Dublin, Dublin, Ireland, [4] The Fred Hollows Foundation New Zealand, Auckland, New Zealand, [5] National Centre for Medical Genetics, Our Lady's Children's Hospital, Dublin, Ireland, [6] Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, MD, USA, [7] Institute of Population Genetics, The National Research Council, Sassari, Italy, [8] Department of Ophthalmology, Seoul National University College of Medicine, Seoul, South Korea, [9] Department of Ophthalmology, Seoul Metropolitan Government Seoul National University Boramae Medical Center, Seoul, South Korea, [10] London School of Hygiene and Tropical Medicine, University of London, London, UK, [11] Department of Ophthalmology, Seoul National University Bundang Hospital, Seoungnam, South Korea, [12] Department of Ophthalmology, University of Thessaly School of Medicine, Larissa, Greece, [13] Retina Service and Ophthalmology, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA, USA, [14] Departments of Medicine, Ophthalmology, Neurology, Epidemiology, and Biostatistics, Boston University Schools of Medicine and Public Health, Boston, MA, USA, [15] Department of Ophthalmology, University of Pennsylvania, Philadelphia, PA, USA

We observed that the third leading cause of blindness in the world, age-related macular degeneration (AMD), occurs at a very low documented frequency in a population-based cohort from Timor-Leste. Thus, we determined a complete catalog of the ancestry of the Timorese by analysis of whole exome chip data and haplogroup analysis of SNP genotypes determined by sequencing the Hypervariable I and II regions of the mitochondrial genome and 17 genotyped YSTR markers obtained from 535 individuals. We genotyped 20 previously reported AMD-associated SNPs in the Timorese to examine their allele frequencies compared to and between previously documented AMD cohorts of varying ethnicities. For those without AMD (average age > 55 years), genotype and allele frequencies were similar for most SNPs with a few exceptions. The major risk allele of *HTRA1* rs11200638 (10q26) was at a significantly higher frequency in the Timorese, as well as 3 of the 5 protective *CFH* (1q32) SNPs (rs800292, rs2284664, and rs12066959). Additionally, the most commonly associated AMD-risk SNP, CFH rs1061170 (Y402H), was also seen at a much lower frequency in the Korean and Timorese populations than in the assessed Caucasian populations (C ∼7 vs. ∼40%, respectively). The difference in allele frequencies between the Timorese population and

the other genotyped populations, along with the haplogroup analysis, also highlight the genetic diversity of the Timorese. Specifically, the most common ancestry groupings were Oceanic (Melanesian and Papuan) and Eastern Asian (specifically Han Chinese). The low prevalence of AMD in the Timorese population (2 of 535 randomly selected participants) may be due to the enrichment of protective alleles in this population at the 1q32 locus.

Introduction

The population of Timor-Leste is afflicted by a very high prevalence of vision problems and ocular morbidity. In 2010, prevalence of blindness in adults aged ≥40 years was 3.6%, and cataract was the most common cause of blindness (Ramke et al., 2012). Although age-related macular degeneration (AMD) is the leading cause of blindness in developed countries, and the third leading cause world-wide, there is a low prevalence in Timor-Leste (Ramke et al., 2012). We therefore hypothesized that the genetic makeup of the Timorese may be enriched for protective variants.

The Democratic Republic of Timor-Leste (also known as East Timor), is a Southeast Asian country, inhabited by an ethnically diverse population which has not been extensively genetically characterized. Previously, two studies showed ∼6-8% allele frequency differences for Y-chromosome short tandem repeat (Y-STR) markers between and among East Timorese and other neighboring populations (Kayser et al., 2001; Souto et al., 2006). Previous ancestry studies of Southeast Asian populations considered only mitochondrial DNA (mtDNA) markers (e.g., Melton et al., 1998; Hill et al., 2007; Tabbada et al., 2010) Y-STRs (e.g., Kayser et al., 2001; Karafet et al., 2010; Zhong et al., 2011; Van Oven et al., 2012), or both (e.g., Kayser et al., 2003; Hurles et al., 2005; Mona et al., 2009; Stoneking and Delfin, 2010). These studies consistently showed a preponderance of Y haplogroups O and M, and mtDNA haplogroups M, B, and F, although many other haplogroups are represented in Southeast Asia at lower frequencies.

We set out to characterize a sample population from Timor-Leste using a more comprehensive panel of both Y-STRs and mitochondrial DNA markers than had been previously reported, and data from a whole exome chip. Additionally, we used these data along with genotypes from 20 disease AMD associated single nucleotide polymorphisms (SNPs) to offer an explanation for the low frequency of AMD among Timorese.

Materials and Methods

Timorese Cohort: Recruitment and Assessment

The study protocol was reviewed and approved by the Institutional Review Board at the University of Utah and by the Timor Ministry of Health and conformed to the tenets of the Declaration of Helsinki. Participants were enrolled in this study after giving informed consent verbally and/or in writing before all data collection and examinations. Communications occurred in Tetum, the local language, or another language, depending on the participant's preference as previously described (Ramke

et al., 2012). Six hundred and three subjects were recruited from Timor-Leste via a population-based cross-sectional survey using multistage cluster random sampling that has been described previously (Ramke et al., 2012). Briefly, individuals were phenotyped by external and intraocular examinations by a trained specialist (GB and JR). This included magnified assessment of the anterior and indirect ophthalmoscopy of the dilated posterior segment as previously described (Ramke et al., 2012). Both blood samples and epidemiological information were available for 535 subjects. Leukocyte DNA was purified by using standard phenol chloroform extraction methods, the DNAzol (Invitrogen) extraction protocol, or the QIAamp DNA Blood Maxi Kit (Qiagen). HbA1C levels were measured using a Roche/Hitachi COBAS C-311 instrument with reagent Tina-quant Hemoglobin A1c Gen.2. for each of the participants to rule out any diabetic disease.

Ancestry of the Timorese

We sequenced the Hypervariable I and II region of the mitochondrial genome of each individual. Reactions were prepared in $25\,\mu l$ volume as follows: $2.5\,\mu l$ of 10X PCR Rxn Buffer (Invitrogen), $0.5\,\mu l$ of 10 mM dNTP Mix (Invitrogen), $0.75\,\mu l$ of 50 mM $MgCl_2$(Invitrogen), $1\,\mu l$ of each $10\,\mu M$ PCR primer (mt-F15971 and mt-R484) (Integrated DNA Technologies Incorporated), $1.25\,\mu l$ of dimethyl sulfoxide (Sigma), $0.2\,\mu l$ of 5 units/μl Taq DNA polymerase (Invitrogen), and $10\,\mu l$ of 2 ng/μl template DNA. Amplification reactions were carried out in a BIORAD MyCycler thermal cycler with the following protocol: initial denaturation 94°C for 2 min, 38 cycles at 94°C for 15 s, 54°C for 30 s, and 72°C for 1 min and 15 s, and a final extension at 72°C for 5 min. We digested $5\,\mu l$ of the post-PCR product with $2\,\mu l$ of ExoSAP-IT (USB), according to the manufacturer's protocol. We added $1\,\mu l$ of each forward $10\,\mu M$ sequencing primer (mt-F15 and mt-F15971) to $7\,\mu l$ of digested post-PCR product (Integrated DNA Technologies Incorporated), and $1\,\mu l$ of $10\,\mu M$ reverse sequencing primer (mt-R141) to $7\,\mu l$ of digested post-PCR product (Integrated DNA Technologies Incorporated). Sequences were analyzed using Sequencher 5.0 (Gene Codes Corporation) independently by two researchers to ensure that mutations were true events rather than mere sequencing errors.

We assessed the degree of Y-chromosome variation among the Timorese males represented in our population sample using 17 short tandem repeats (STRs) loci (DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, and Y-GATA H4) using the AmpF*l*STR Yfiler Amplification Kit (Applied Biosystems) using the manufacturer's

recommendations. DNA was amplified in a GeneAMp PCR System 9700 equipped with silver blocks and DNA fragment separation and detection was carried out on an ABI Prism 3100 Genetic Analyzer (Applied Biosystems). Genescan500 LIZ (ABI) was utilized as an internal size standard. Data were analyzed with an allelic ladder as well as positive and negative controls using GeneMapper® 3.2 (ABI) and were double-checked manually for accuracy.

Haplogroup assignation for the Y-STR markers was carried out with World Haplogroup & Haplo-I Subclade Predictor (http://members.bex.net/jtcullen515/haplotest.htm) and confirmed using Althey's Haplogroup Predictor (http://www. hprg.com/hapest5/). The two identifying SNPs for haplogroups O3 (M122 and P198) and O2 (P31 and M268) were confirmed using a Taqman assay and direct sequencing, respectively [following Van Oven (Van Oven et al., 2012) for primer design]. Haplogroup assignment for the Hypervariable I and II region of the mitochondrial genome was determined using MitoTool (www.mitotool.org) and confirmed using HaploGrep (haplogrep.uibk.ac.at).

Genotyping using the Illumina HumanExome array version 1.1 chip was also performed at the Center for Inherited Disease Research (CIDR) as part of a refractive error consortium study. This chip not only includes >240,000 SNPs including 30,000 custom probes but most importantly 3468 Ancestry informative markers. CIDR laboratory standard quality control procedures were applied to the entire dataset prior to release of the genotype data from the laboratory. Once sample collection was finalized, samples were plated controlled-randomized to avoid spurious associations due to plate effects. This QC included the following: Blind duplicates and HapMap controls were distributed across plates for concordance checking. Samples with suspected mixtures or unusual X and Y patterns or gender mismatch were identified and dropped before release of genotype data from the CIDR lab. SNP clustering was performed on all SNPs in the project and SNP genotypes with genotype quality (GC) score less than 0.15 were recoded as missing genotypes. Autosomal SNPs with less than 85% call rate, cluster separation of less than 0.3 and heterozygote rate greater than 80% were dropped prior to release of genotype data from the laboratory. A subset of SNPs was also manually reviewed that included all Y, XY pseudoautosomal and mitochondrial SNPs and other various circumstances.

After receiving data from CIDR, additional quality control measures were applied. Sex discrepancies were calculated in PLINK (Purcell et al., 2007) (http://pngu.mgh.harvard. edu/~purcell/plink/) and samples which did not appear sufficiently matched to their recorded sex were dropped. Any unexpected duplicate samples were identified using PREST-PLUS (http://utstat.toronto.edu/sun/Software/Prest/) and one of the duplicate pair dropped. SNPs with >1 errors in blind duplicates or HapMap controls were dropped and SNPs with >1 Mendelian error after correction of pedigree relationships were also removed. Testing for batch effects was performed using a homogeneity test of minor allele frequency for each SNP on each plate compared to all other plates (Pluzhnikov et al., 2008, 2010). These statistics were then averaged over all SNPs to determine

how the plates deviate from each other in PLINK. Heterozygosity rates across samples were checked and outlier samples excluded. Examination of samples for chromosomal abnormalities was performed and problematic samples were identified. Autosomal SNPs with sex difference in allelic frequency >0.2, sex difference in heterozygosity > 0.3 were also excluded. Monomorphic variants were also excluded.

Resulting genotype data were then merged with the samples from the Human Genome Diversity Project (HGDP, http://www. hagsc.org/hgdp/index.html) using—merge in PLINK, ensuring that all overlapping SNPs were on the same strand, and removing all AT and CG SNPs. Assessment of population sub-structure in this resulting merged HGDP-Timor dataset was performed using principal components analysis with Eigensoft (Patterson et al., 2006; Price et al., 2006) (http://www.hsph.harvard. edu/alkes-price/software/), Admixture version 1.2 (Alexander et al., 2009) (http://www.genetics.ucla.edu/software/admixture/ download.html), and Ancestry Mapper (Magalhaes et al., 2012) (http://cran.r-project.org/web/packages/AncestryMapper/ index.html). Ancestry Mapper is a program that uses SNPs to create a unique individual genetic identifier and inform ancestry by comparing that identifier to those of individuals with known genetic ancestry. For this study, the Timorese individuals were compared to the 51 populations included in the HGDP. This includes individuals a world wide range of populations covering all continents including Africa, Europe, the Middle East, South and Central Asia, East Asia, Oceania and the Americas (Cann et al., 2002; Rosenberg et al., 2002, 2005) and is a widely used dataset to study genetics of populations (Cavalli-Sforza et al., 1991; Tishkoff et al., 2009; International HapMap 3; Li et al., 2008; López Herráez et al., 2009; International HapMap 3 Consortium et al., 2010). Using all three methods, which rely on different algorithms, assures the consistency and cross-validation of the results.

AMD-Associated SNPs

To investigate the genetic basis for the low prevalence of AMD in this cohort, we genotyped 20 previously reported AMD-associated SNPs to examine their allele frequencies in the Timorese and other populations of different ethnicities. This included six SNPs from the *CFH* (1q32) region (*CFH* SNPs rs800292, rs16840422, rs1061170, rs12144939, and rs2284664, and *CFHR2* rs3790414) (Klein et al., 2005; Li et al., 2006; Zhang et al., 2008; Sivakumaran et al., 2011), along with the *CFHR1–3Δ* change (Hughes et al., 2006), 3 *ROBO1* (3p12) SNPs (rs1387665, rs4513416, and rs9309833) (Jun et al., 2011), 2 SNPs from the 6p21 region where *C2* resides (*C2* rs9332739 and rs547154) (Gold et al., 2006), 6 SNPs from *ARMS2/HTRA1* (10q26) region (*ARMS2* rs10490924 and rs10664316, and *HTRA1* rs11200638, rs2672598, rs1049331, and rs2293870) (Dewan et al., 2006; Yang et al., 2006; Deangelis et al., 2008), and 3 *RORA* (15q22) SNPs (rs12900948, rs8034864, and rs730754) (Schaumberg et al., 2010; Silveira et al., 2010; Jun et al., 2011). These SNPs were also genotyped in three diverse and carefully AMD-phenotyped populations including a family-based cohort from New England (*n* = 657) (DeAngelis et al., 2004; Silveira et al., 2010), a case-control cohort from Central Greece (*n* = 436) (Silveira

et al., 2010), and a case-control cohort recruited from Seoul National University Bundang Hospital ($n = 818$) (Fritsche et al., 2013).

All *CFH, CFHR2, C2, ROBO1, ARMS2, HTRA1* (rs11200638 and rs2672598), and *RORA* SNPs were genotyped using a combination of pre-designed and Custom Taqman SNP Genotyping Assays (Applied Biosystems). Each assay was run in a 15 ul reaction containing $2\times$ Taqman GTXpress master mix, $40\times$ or $80\times$ probe, and 10 ng of DNA. Thermal cycling was performed according to the manufacturer's protocol. The ABI 7500 Real-Time PCR System, with the accompanying software, was used to analyze the genotypes. Direct sequencing was used to genotype *HTRA1* SNPs (rs1049331 and rs2293870) using previously reported oligonucleotide primers and methods (Deangelis et al., 2008). For sequencing reactions, PCR products were digested according to the manufacturer's protocol using ExoSAP-IT (USB). DNA sequencing was performed at the University of Utah DNA Sequencing Core Laboratory. Electropherograms were read independently by two evaluators without knowledge of the subject's disease status. *CFHR1–3Δ* status was determined directly in the Timorese cohort using a pre-designed Taqman copy number assay (Applied Biosystems). Each sample was run in triplicate in 20 ul reactions containing $2\times$ Taqman GTXpress master mix, $20\times$ probe, $20\times$ RNaseP, and 20 ng of genomic DNA. Data was analyzed using CopyCaller v2.0 software. *CFHR1–3Δ* status was compared to the T allele of rs12144939 to determine its tagging validity.

Allele frequencies in each cohort were calculated for normal subjects, or those without any evidence of documented AMD. Linkage disequilibrium among the genotyped SNPs was calculated using only those without any signs of AMD in Haploview (http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview).

Results

Subjects

Six hundred and 3 subjects from Timor-Leste were examined as part of this study. Of these 603 subjects, both blood samples and epidemiological information were available for 535 subjects. Of these 535 subjects, 267 were males and 268 were females. Average HbA1c was 5.57 ± 0.71 for this cohort. A diagnosis of early AMD was given to two subjects from the Timor population: a 62 year old female and a 75 year old male. The average age of the study participants without AMD was 55.12 years (range 40–94). Subject characteristics of all populations examined in this study are shown in **Table 1**.

Ancestry Analysis

The Y-STR analysis yielded genotypes for 265 of the 267 Timorese males. Using the publicly available databases World Haplogroup & Haplo-I Subclade Predictor (http://members.bex.net/jtcullen515/haplotest.htm) and Althey's Haplogroup Predictor (http://www.hprg.com/hapest5/), 11 haplogroups were assigned to 248 of these subjects (**Figure 1A**). The origin of the
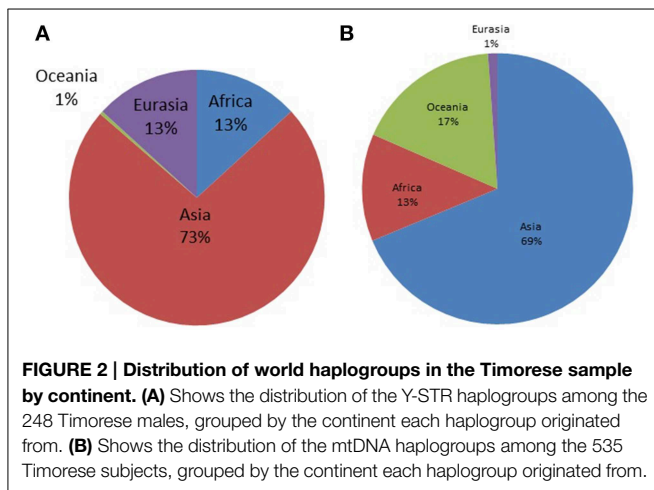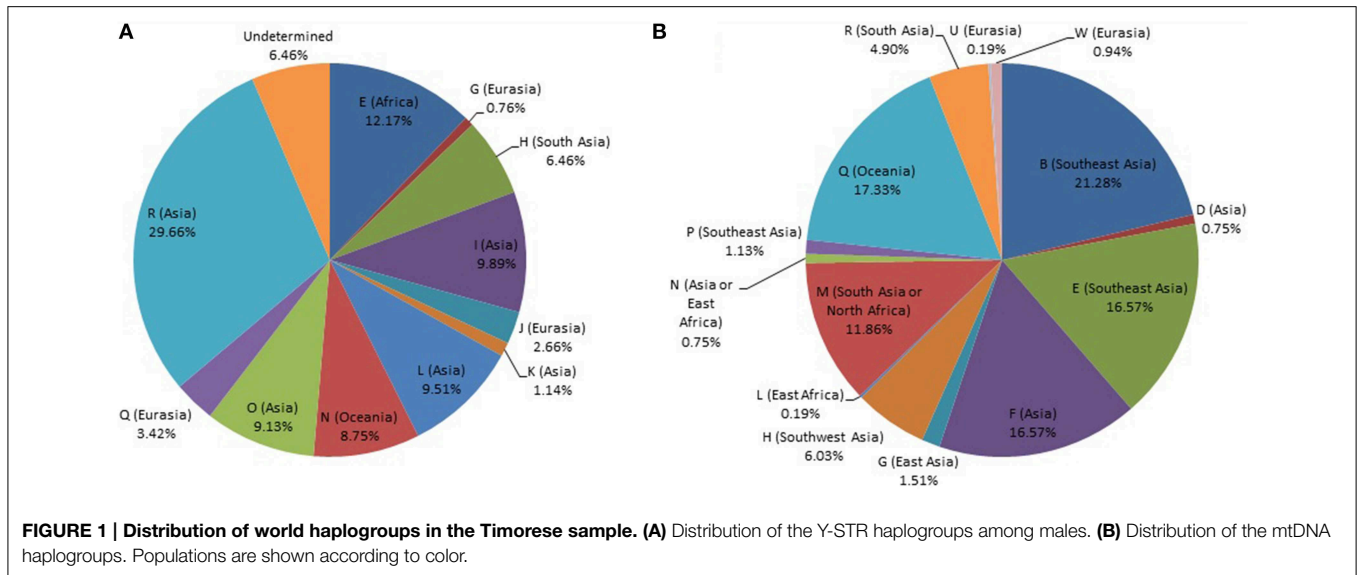
TABLE 1 | Characteristics of subjects without AMD.

Characteristic	NESC	Greek	Korean	Timor
n	198	213	384	533
Males (% total)	87 (43.9%)	100 (46.9%)	194 (50.5%)	268 (50.1%)
Age (range)	75.45 (50–94)	73.78 (48–95)	68.46 (50–87)	55.12 (40–94)

haplogroups is mostly Asia (73%), followed by Africa and Eurasia (13% each), and a small percent Oceana (1%) (**Figure 2A**).

MtDNA genotypes were determined for all 535 Timorese subjects using the publicly available reference data and are distributed among 14 haplogroups (**Figure 1B**). The haplogroups distribution shows that the origin of the Timorese males is mostly Asian (69%), followed by Oceanic (17%), African (13%), and a small percent Eurasian (1%), (**Figure 2B**). Given the within-group diversity present, it is likely that either multiple migrations or a few migrations of distantly related male individuals from Asia and other parts of the Old World have populated the islands, which is consistent with pre-historical and historical events.

After QC and data cleaning, genotypes from 253,405 SNPs were available for 489 Timorese subjects with a call rate of >98% and were included in the subsequent analyses. There were 11,064 common SNPs between the HGDP dataset and the Timorese dataset used for subsequent ancestry analyses. Principal components analysis showed that the first principal component, PC1, was most similar between the Timorese and Papuans, followed by the She, Tujia, Dai, and Han Chinese (mean PC1 ~ -0.02; **Figure 3**). Using admixture, we have used several number of groupings and have found that $K = 5$ organizes the populations into the known ancestral populations, i.e., African, Indo-European, Oceania, Amerindian and Asian. Those global ancestry components allow us to separate HGDP, a dataset with worldwide coverage into their well-established continental differences since the known populations are correctly assigned. Where $k = 5$, the Timorese samples were showing the greatest proportion of Admix group 2, followed by a smaller proportion of Admix group 4 (**Figure 4A**). Individuals with the highest values for Admix group 2 are mostly of Oceanic populations, such as the Melanesian and Papuan samples, which are geographically close to Timor-Leste. Individuals with high values for Admix group 4 are from Eastern Asia, including Japanese and several Chinese populations. By increasing the number of ancestral assignments (k) to 10, there is a clearer separation of the Timorese samples, in that they show a closer relationship with the Papuan samples, and not as much the Melanesian (**Figure 4B**). Using Ancestry Mapper, the largest percentage of the Timorese was assigned to the Han Chinese reference population (~85%, **Figure 5**).

Only two individuals in our cohort were affected by AMD. The first was a 62 year old female, whose mitochondrial haplogroup is F3b1 (thus implying maternal Southeast Asian ancestry) and whose principal component 1 was very close to the average of the group and the Papuan reference (PC1 = -0.0250). Using admixture, this subject was shown to be Oceanic ($k = 5$) and when $k = 10$, this individual showed the most similarity to Han Chinese (50.5%) and also Papuan (43.3%). Finally, Ancestry Mapper showed that aside from the Timorese reference group,

FIGURE 1 | Distribution of world haplogroups in the Timorese sample. (A) Distribution of the Y-STR haplogroups among males. **(B)** Distribution of the mtDNA haplogroups. Populations are shown according to color.

FIGURE 2 | Distribution of world haplogroups in the Timorese sample by continent. (A) Shows the distribution of the Y-STR haplogroups among the 248 Timorese males, grouped by the continent each haplogroup originated from. **(B)** Shows the distribution of the mtDNA haplogroups among the 535 Timorese subjects, grouped by the continent each haplogroup originated from.

this individual was given the highest score to the Han Chinese reference group (79.0%), followed by the Cambodian reference group (75.9%). The second AMD individual was a 75 year old male, who belongs to Y-haplogroup H (thus implying a likely South Asian paternal origin) and mitochondrial haplogroup B4b1 (thus implying maternal Southeast Asian ancestry). This subject was shown by principal components to be closest to the Papuan reference (PC1 = −0.0217) and by Admixture to be Oceanic when $k = 5$ but a split of Han Chinese and Papuan when $k = 10$ (48.3 and 48.0%, respectively). Finally, using Ancestry Mapper, after the Timorese reference group, this individual was given the highest score to the Han Chinese reference group (81.1%), followed by the Hazara reference (77.3%).

AMD-Associated SNPs

The frequency of alleles of the 20 genotyped SNPs was statistically significantly different between the Timorese and other genotyped cohorts for the majority of the SNPs investigated. Specifically, statistically significantly differences were observed in 11 SNPs

between the Timorese and the NESC cohort, 16 SNPs between the Timor and Greek cohorts, and 18 SNPs between the Timor and Korean cohorts ($p < 0.05$, **Table 2**). For all cohorts examined, the most statistically significant differences in allele frequencies were observed in the *CFH* (1q32) region, followed by the *ARMS2/HTRA1* (10q26) region. The two most consistently AMD-associated SNPs, *CFH* rs1061170 (Y402H) and *ARMS2* rs10490924 (A69S) were not statistically different between the Timorese and all cohorts. The alleles of Y402H were significantly differently distributed between the Timor and Caucasian cohorts (NESC and Greek) but not the Koreans. Specifically, the risk allele, C, was significantly less frequent in the Timorese than in the Caucasians. The alleles of A69S were statistically different between the Timorese and the Greeks and Koreans but not the Caucasians. Specifically, the risk allele, T, was significantly higher in the Timorese than the Greeks but lower than the Koreans.

Notably, the neither of the two subjects with early AMD in this Timorese cohort had risk alleles at either of the two major risk loci, *CFH* Y402H (rs1061170) or *ARMS2* A69S (rs10490924). Interestingly, the *CFH* SNP rs12144939, which has been shown to tag for the protective CFHR1-3Δ, was observed at a relatively similar frequency in each of the cohorts examined (MAF ∼ 5–20%, with the Koreans having the lowest frequency), although it's tagging ability was only validated in the NESC cohort.

The linkage disequilibrium (LD) patterns among the genotyped SNPs showed differences between the four populations (Supplementary Table 1). On chromosome 1 where *CFH* resides, no haplotype blocks were observed for the NESC, Greek or Korean cohorts, while in the Timorese, a haplotype block encompassing 5 SNPs was defined by the Gabriel rule (Gabriel et al., 2002) (*CFH* rs16840422, rs1061170, rs2284664, and rs12144939). On chromosome 10, the LD patterns were similar between the Caucasian cohorts and between the Korean and Timorese cohorts. Specifically, in the Caucasian cohorts there was high LD ($r^2 > 0.8$) among 3 pairs of
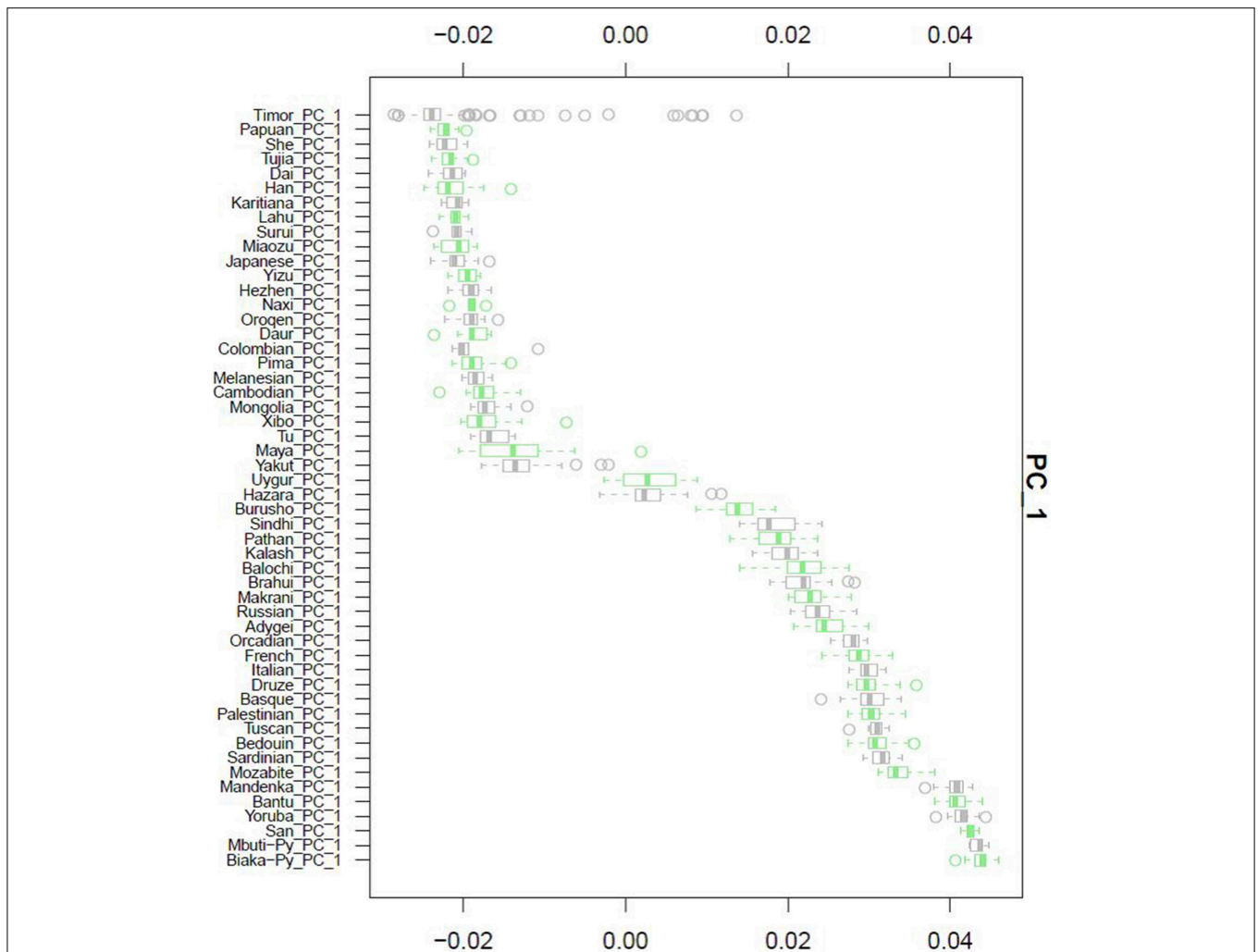
FIGURE 3 | Principal Component Analysis. Principal component 1, PC1, as calculated for each of the 51 HGDP references and compared to the Timorese individuals investigated in this study. The Y-axis shows each of the reference populations. The X-axis shows the calculated PC1 value for each population.
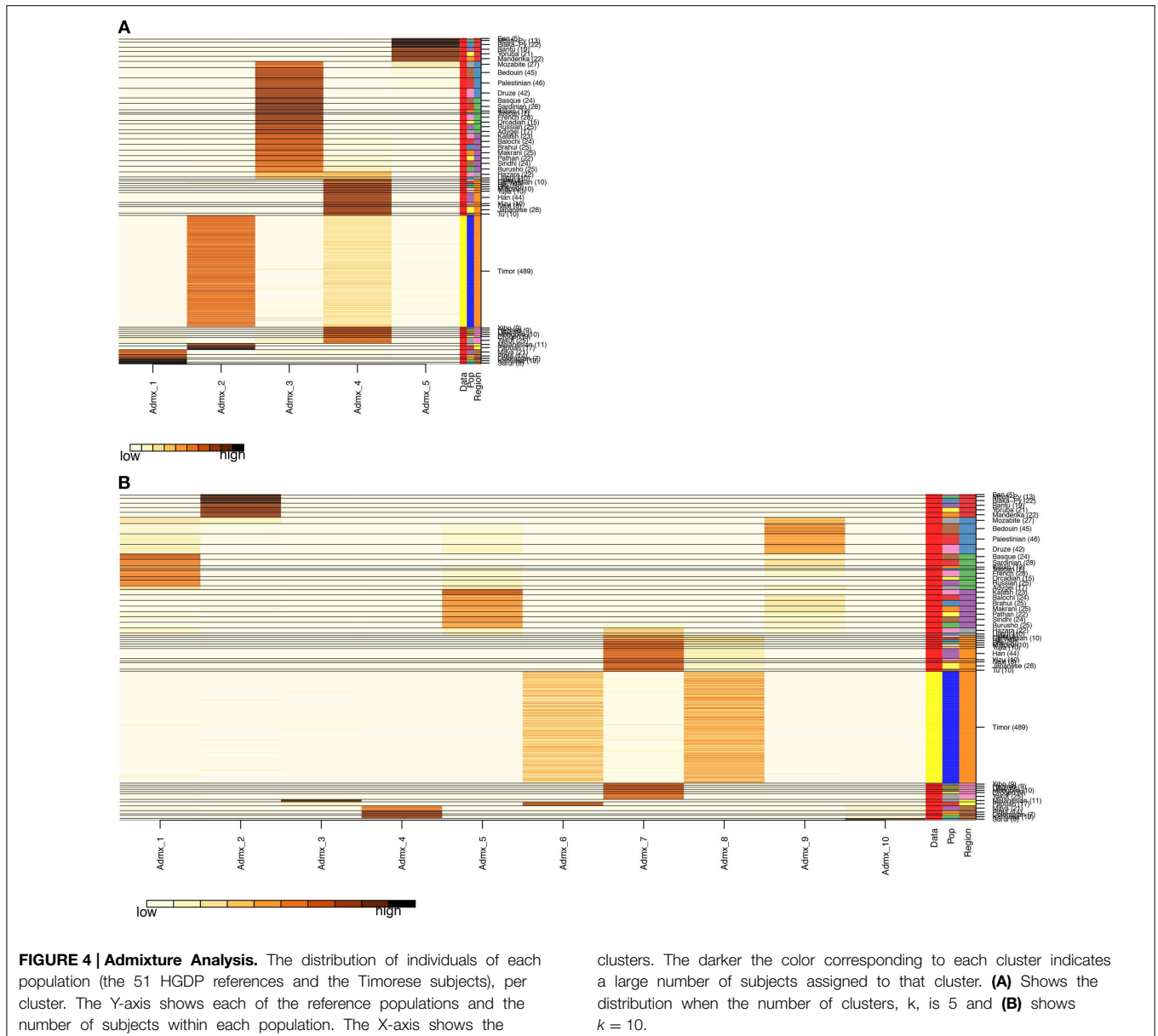
SNPs, while in the Korean and Timorese cohorts there was high LD among 6 sets of SNPs. On 15q22, there was high LD among two *RORA* SNPs in all cohorts, while the Korean and Timorese cohorts also showed high LD among the third. No significant differences in LD were observed on chromosome 3 (*ROBO1*) or chromosome 6 (*C2*).

Discussion

Utilization of the Timorese cohort provides a unique opportunity to study protection from a complex disease, AMD, in a geographically isolated population with low disease prevalence. AMD is the only complex disease for which 2 loci, *CFH* and *ARMS2/HTRA1*, explain a large proportion of risk (Edwards et al., 2005; Hageman et al., 2005; Haines et al., 2005; Klein et al., 2005; Dewan et al., 2006; Yang et al., 2006, p. 200; Swaroop et al., 2007; Deangelis et al., 2008; Fritsche et al., 2013). In addition, AMD risk is explained by many other loci which implicate several mechanisms and pathways underlying AMD pathophysiology

(for review please see, Swaroop et al., 2009; Deangelis et al., 2011; Fritsche et al., 2013; Miller, 2013).

Other studies have shown that population structure and admixture can have strong confounding effects on the determination of genetic association analyses (e.g., Shtir et al., 2009). This is an important factor for which to account when studying the genetics of AMD, as there are inconsistencies in the distribution of AMD-related alleles and AMD prevalence among different world populations (e.g., Grassi et al., 2006; Mori et al., 2007; Nonyane et al., 2010). Although this is not meant to be an exhaustive literature review, studies have shown that the risk allele at Y402H (rs1061170) located in *CFH*, which is one of the most strongly associated markers to any complex disease [odds rations (ORs) >5 for the homozygous risk allele genotype, Klein et al., 2005], has a high frequency in Caucasians (Edwards et al., 2005; Hageman et al., 2005; Haines et al., 2005; Klein et al., 2005). These observations have been documented independently by different researchers who studied different populations, including the Japanese (Grassi et al., 2006; Mori et al., 2007),

FIGURE 4 | Admixture Analysis. The distribution of individuals of each population (the 51 HGDP references and the Timorese subjects), per cluster. The Y-axis shows each of the reference populations and the number of subjects within each population. The X-axis shows the clusters. The darker the color corresponding to each cluster indicates a large number of subjects assigned to that cluster. **(A)** Shows the distribution when the number of clusters, k, is 5 and **(B)** shows $k = 10$.
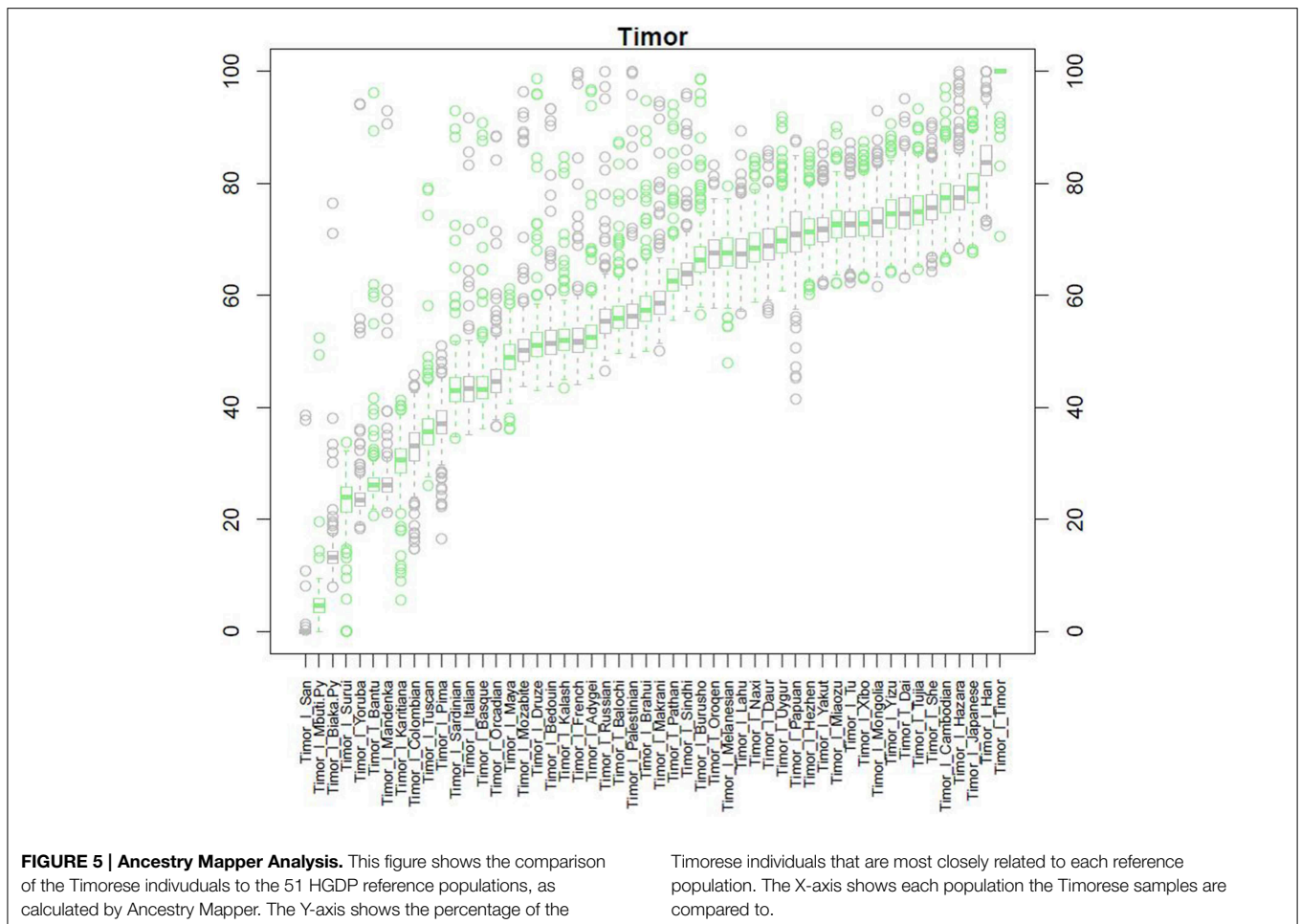
Chinese (Ng et al., 2008; Xu et al., 2008), and Koreans (Kim et al., 2008). On the other hand, Africans were shown to have comparatively high frequencies of the deleterious alleles, but low AMD prevalence (though not adjusted for life expectancy) (Friedman et al., 1999; Klein et al., 1999; Muñoz et al., 2000; Ziskind et al., 2008).

Our work has shown that the Timorese population represented by 535 phenotyped individuals exhibits an increased frequency of some protective genotypes (*CFH* rs800292, rs1684022, and rs2284664) and a decreased frequency of a major risk genotype (at *CFH* rs1061170), thus further confirming the implication of the two main loci on chromosome 1 and 10. However, the Timorese cohort exhibits a lower frequency of some protective alleles as compared to the other population(s) reported here (*CFH* rs12144939 and rs12066959; *HTRA1* rs2672598; *ARMS2* rs10664316), thus emphasizing the difficulty

of identifying causative alleles implicated in the pathogenesis of complex diseases.

Studying the genetic variation in different populations where AMD prevalence and disease expression (i.e., phenocopy) varies is of paramount importance if we wish to untangle the relationship between genotype and phenotype, and genotype and environment. Additionally, comparing genotypes among populations can facilitate the identification of causal variants, a process that can be masked by the effect of linkage disequilibrium (LD). In a case where causal and non-causal variants are inherited together as part of one LD block, the identification of causal variants is difficult, as both causal and non-causal variants may have similar frequencies. Populations from different lineages have an increased chance of having lost some of the variants due to recombination, and may differ from other populations with diverse evolutionary histories. Finally, this study shows that

FIGURE 5 | Ancestry Mapper Analysis. This figure shows the comparison of the Timorese indivuduals to the 51 HGDP reference populations, as calculated by Ancestry Mapper. The Y-axis shows the percentage of the Timorese individuals that are most closely related to each reference population. The X-axis shows each population the Timorese samples are compared to.

ancestry plays a remarkable role in the risk of developing a disease, given differential exposure to selective stressors during the natural history of each population.

This ancestry analysis in the Timorese builds on the work that was performed by Souto et al. (2006), as the world haplogroup databases have grown in the past several years. We confirm the genetic diversity of this population found by Souto but add to the diversity with the addition of Oceanic groupings. The mtDNA analyses conducted in this study show that the Timorese populations represented in our sample cohort have a complex natural history. Since mtDNA is maternally inherited, it does not recombine, and has a high (but fairly constant) mutation rate, it allows us to study lineage-specific variation in maternal lineages. Additionally, we typed the mitochondrial control region (which includes HVR I and II) as an ancestry marker, because it is subjected to low selective pressures because it is non-coding, with its main function being mtDNA replication. These analyses indicate that a number of lineages are represented in this population, with haplogroups B (found mainly in East Asia), E (common among Malaysian people), F (being a widespread Asian haplogroup found in high frequencies in Asia and Japan), and Q (having its highest frequency in Oceania), being the most common. The pattern of mtDNA variation is somewhat different from that of the Y-chromosome. The most common

Y-chromosome haplogroup was R, which is most common in Western Asia. Similar findings were seen in the analysis of the whole exome data, in that the most common groupings were Oceanic (Melanesian and Papuan) and Eastern Asian (specifically Han Chinese).

More recently, a similar study was performed on the population of West Timor, the population inhabiting the other side of the island, by Tumonggor et al. (2014). Like their study, we found slightly different haplogroups, and thus origins, of paternal and maternal lineages. Both studies observed the highest frequencies of B mtDNA haplogroups (24.4% in the West Timorese and 21% in the East Timorese), followed by F, Q, and E, while the Y-STR haplogroups differed between studies. Unlike their study, the most common Y-STR haplogroup in the current study was R (30%), while in the West Timorese it was C (26.6%). Using autosomal markers, although through different methodology (whole exome sequencing vs. 37 ancestry informative markers), our analysis showed similar results to what was found in the West Timorese: there are strong Han Chinese and Papuan lineages.

Our study has several limitations. Although our analysis of Y-chromosome and mtDNA markers allowed robust interpretation of the ancestral origins of the East Timorese, genotype information for a much larger number of autosomal markers

TABLE 2 | AMD-Associated SNP Allele Frequencies.

SNP	AMD Effect*	Chr.	Position	Alleles	Timor	NESC	p-value	Greek	p-value	Korean	p-value
CFH rs800292	A is Protective	1	194908856	G/A	A = 0.688	A = 0.205	<0.0001	A = 0.220	< 0.0001	A = 0.397	< 0.0001
CFH rs16840422	T is Protective	1	194919457	C/T	T = 0.155	T = 0.131	0.2550	T = 0.110	0.0237	T = 0.083	< 0.0001
CFH rs1061170	C is Risk	1	194925860	T/C	C = 0.067	C = 0.429	<0.0001	C = 0.373	< 0.0001	C = 0.069	0.8601
CFH rs12144939	T is Protective, tags for CFHR1-3Δ	1	194965568	G/T	T = 0.146	T = 0.170	0.2654	T = 0.213	0.0018	T = 0.053	< 0.0001
CFH rs2284664	T is Protective	1	194969148	C/T	T = 0.613	T = 0.197	<0.0001	T = 0.207	< 0.0001	T = 0.364	< 0.0001
CFHR2 rs3790414	A is Protective	1	195186922	T/A	A = 0.764	A = 0.240	<0.0001	A = 0.250	< 0.0001	A = 0.683	0.0002
ROBO1 rs1387665	T is Risk	3	79512501	C/T	T = 0.553	T = 0.497	0.0706	T = 0.493	0.0411	T = 0.380	< 0.0001
ROBO1 rs4513416	A is Protective	3	79573493	G/A	A = 0.346	A = 0.413	0.0191	A = 0.407	0.0305	A = 0.468	< 0.0001
ROBO1 rs9309833	C is Risk	3	79894409	T/C	C = 0.178	C = 0.124	0.0159	C = 0.172	0.7890	C = 0.308	< 0.0001
C2 rs9332739	C is Protective	6	32011783	G/C	C = 0.006	C = 0.035	<0.0001	C = 0.050	< 0.0001	C = 0.024	0.0011
C2 rs547154	T is Protective	6	32018917	G/T	T = 0.166	T = 0.073	<0.0001	T = 0.070	< 0.0001	T = 0.101	< 0.0001
ARMS2 rs10490924	T is Risk	10	124204438	G/T	T = 0.336	T = 0.311	0.3680	T = 0.237	0.0002	T = 0.398	0.0064
ARMS2 rs10664316	- is Protective	10	124206375	delAT	− = 0.331	− = 0.396	0.0223	− = 0.415	0.0023	− = 0.358	0.2431
HTRA1 rs11200638	A is Risk	10	124210534	G/A	A = 0.635	A = 0.303	<0.0001	A = 0.226	< 0.0001	A = 0.405	< 0.0001
HTRA1 rs2672598	C is Risk	10	124210672	T/C	C = 0.523	C = 0.492	0.3028	C = 0.534	0.7109	C = 0.711	< 0.0001
HTRA1 rs1049331	T is Risk	10	124211260	C/T	T = 0.358	T = 0.306	0.0643	T = 0.231	< 0.0001	T = 0.407	0.0320
HTRA1 rs2293870	T and C are Risk	10	124211266	G/T/C	T = 0.355, C = 0.001	T = 0.308, C = 0.089	0.1489	T = 0.226, C = 0.164	0.2251	T = 0.407, C = 0.000	0.0288
RORA rs12900948	G is Protective	15	59227963	A/G	G = 0.413	G = 0.410	0.9246	G = 0.502	0.0020	G = 0.580	< 0.0001
RORA rs8034864	A is Risk	15	59233645	C/A	A = 0.417	A = 0.209	<0.0001	A = 0.290	< 0.0001	A = 0.587	< 0.0001
RORA rs730754	G is Protective	15	59238628	A/G	G = 0.420	G = 0.397	0.4331	G = 0.468	0.0941	G = 0.618	< 0.0001

*p-values shown are those comparing allele frequencies between each cohort and Timorese. *AMD References: CFH region (1q32) (Hageman et al., 2005; Haines et al., 2005; Klein et al., 2005; Hughes et al., 2006; Li et al., 2006; Zhang et al., 2008; Sivakumaran et al., 2011); ROBO1 (3p12) (Jun et al., 2011); C2 (6p21 region) (Gold et al., 2006); ARMS2/HTRA1 (10q26) (Dewan et al., 2006; Yang et al., 2006; Deangelis et al., 2008); and RORA (15q22) (Schaumberg et al., 2010; Silveira et al., 2010; Jun et al., 2011).*

is needed to do a thorough population admixture analysis. In addition, our sample included too few AMD subjects to make definitive conclusions about the frequencies of established AMD risk variants in AMD subjects in East Timor. Nonetheless, the low prevalence of AMD in this community is explained in part by the very low frequency of several AMD risk variants that are prevalent in other populations throughout the world.

The difference in allele frequencies between the Timorese population and the other genotyped populations, along with the ancestry analysis, show the genetic diversity of this population. Their reduced frequency of deleterious alleles and increased frequency of protective alleles for AMD-associated variants gives reason for their low prevalence of disease. Their ancestry analysis shows they are closely related to Eastern Asians, who also have low prevalence of AMD.

Author Contributions

MM, JR, DS, and MD were involved in the conception and design of the work. JR, JA, GB, LL, SW, FW, ET, JM, IK, KP, DS, and MD contributed to the acquisition of data for analysis. MM, TM, JR, SS, SE, CS, LP, FM, CD, KM, RR, DM, GB, FZ, JW, IK, KP, LF, J B-W, DS, MD were responsible for analysis and/or interpretation of data. MM, TM, SS, SE, CS, LP, FM, JA, CD, KM, RR, DM, SW, EE, KP, and MD were involved in the drafting the manuscript. MM, TM, JR, SS, GB, LL, FZ, JM, IK, JB, LF, DS, and MD were responsible for revising the manuscript critically for

important intellectual content. MM, TM, JR, SS, SE, CS, LP, FM, JA, CD, KM, RR, DM, GB, LL, SW, FZ, ET, JW, IK, KP, JB, LF, DS, and MD gave final approval of the version to be published; and agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fgene.2015.00238

References

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109

Cann, H. M., de Toma, C., Cazes, L., Legrand, M.-F., Morel, V., Piouffre, L., et al. (2002). A human genome diversity cell line panel. *Science* 296, 261–262. doi: 10.1126/science.296.5566.261b

Cavalli-Sforza, L. L., Wilson, A. C., Cantor, C. R., Cook-Deegan, R. M., and King, M. C. (1991). Call for a worldwide survey of human genetic diversity: a vanishing opportunity for the Human Genome Project. *Genomics* 11, 490–491.

Deangelis, M. M., Ji, F., Adams, S., Morrison, M. A., Harring, A. J., Sweeney, M. O., et al. (2008). Alleles in the HtrA serine peptidase 1 gene alter the risk of neovascular age-related macular degeneration. *Ophthalmology* 115, 1209–1215.e7. doi: 10.1016/j.ophtha.2007.10.032

DeAngelis, M. M., Lane, A. M., Shah, C. P., Ott, J., Dryja, T. P., and Miller, J. W. (2004). Extremely discordant sib-pair study design to determine risk factors for neovascular age-related macular degeneration. *Arch. Ophthalmol.* 122, 575–580. doi: 10.1001/archopht.122.4.575

Deangelis, M. M., Silveira, A. C., Carr, E. A., and Kim, I. K. (2011). Genetics of age-related macular degeneration: current concepts, future directions. *Semin. Ophthalmol.* 26, 77–93. doi: 10.3109/08820538.2011.577129

Dewan, A., Liu, M., Hartman, S., Zhang, S. S.-M., Liu, D. T. L., Zhao, C., et al. (2006). HTRA1 promoter polymorphism in wet age-related macular degeneration. *Science* 314, 989–992. doi: 10.1126/science.1133807

Edwards, A. O., Ritter, R. 3rd, Abel, K. J., Manning, A., Panhuysen, C., and Farrer, L. A. (2005). Complement factor H polymorphism and age-related macular degeneration. *Science* 308, 421–424. doi: 10.1126/science.1110189

Friedman, D. S., Katz, J., Bressler, N. M., Rahmani, B., and Tielsch, J. M. (1999). Racial differences in the prevalence of age-related macular degeneration: the Baltimore Eye Survey. *Ophthalmology* 106, 1049–1055. doi: 10.1016/S0161-6420(99)90267-1

Fritsche, L. G., Chen, W., Schu, M., Yaspan, B. L., Yu, Y., Thorleifsson, G., et al. (2013). Seven new loci associated with age-related macular degeneration. *Nat. Genet.* 45, 433–439, 439e1–439e2. doi: 10.1038/ng.2578

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., et al. (2002). The structure of haplotype blocks in the human genome. *Science* 296, 2225–2229. doi: 10.1126/science.1069424

Gold, B., Merriam, J. E., Zernant, J., Hancox, L. S., Taiber, A. J., Gehrs, K., et al. (2006). Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration. *Nat. Genet.* 38, 458–462. doi: 10.1038/ng1750

Grassi, M. A., Fingert, J. H., Scheetz, T. E., Roos, B. R., Ritch, R., West, S. K., et al. (2006). Ethnic variation in AMD-associated complement factor H polymorphism p.Tyr402His. *Hum. Mutat.* 27, 921–925. doi: 10.1002/humu.20359

Hageman, G. S., Anderson, D. H., Johnson, L. V., Hancox, L. S., Taiber, A. J., Hardisty, L. I., et al. (2005). A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc. Natl. Acad. Sci. U.S.A.* 102, 7227–7232. doi: 10.1073/pnas.0501536102

Haines, J. L., Hauser, M. A., Schmidt, S., Scott, W. K., Olson, L. M., Gallins, P., (2005). Complement factor H variant increases the risk of age-related macular degeneration. *Science* 308, 419–421. doi: 10.1126/science.1110359

Hill, C., Soares, P., Mormina, M., Macaulay, V., Clarke, D., Blumbach, P. B., et al. (2007). A mitochondrial stratigraphy for island southeast Asia. *Am. J. Hum. Genet.* 80, 29–43. doi: 10.1086/510412

Hughes, A. E., Orr, N., Esfandiary, H., Diaz-Torres, M., Goodship, T., and Chakravarthy, U. (2006). A common CFH haplotype, with deletion of CFHR1 and CFHR3, is associated with lower risk of age-related macular degeneration. *Nat. Genet.* 38, 1173–1177. doi: 10.1038/ng1890

Hurles, M. E., Sykes, B. C., Jobling, M. A., and Forster, P. (2005). The dual origin of the Malagasy in Island Southeast Asia and East Africa: evidence from maternal and paternal lineages. *Am. J. Hum. Genet.* 76, 894–901. doi: 10.1086/430051

International HapMap 3 Consortium, Altshuler, D. M., Gibbs, R. A., Peltonen, L., Altshuler, D. M., Gibbs, R. A., (2010). Integrating common and rare genetic variation in diverse human populations. *Nature* 467, 52–58. doi: 10.1038/nature09298

Jun, G., Nicolaou, M., Morrison, M. A., Buros, J., Morgan, D. J., Radeke, M. J., et al. (2011). Influence of ROBO1 and RORA on risk of age-related macular degeneration reveals genetically distinct phenotypes in disease pathophysiology. *PLoS ONE* 6:e25775. doi: 10.1371/journal.pone.0025775

Karafet, T. M., Hallmark, B., Cox, M. P., Sudoyo, H., Downey, S., Lansing, J. S., et al. (2010). Major east-west division underlies Y chromosome stratification across Indonesia. *Mol. Biol. Evol.* 27, 1833–1844. doi: 10.1093/molbev/msq063

Kayser, M., Brauer, S., Weiss, G., Schiefenhövel, W., Underhill, P. A., and Stoneking, M. (2001). Independent histories of human Y chromosomes from Melanesia and Australia. *Am. J. Hum. Genet.* 68, 173–190. doi: 10.1086/316949

Kayser, M., Brauer, S., Weiss, G., Schiefenhövel, W., Underhill, P., Shen, P., et al. (2003). Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am. J. Hum. Genet.* 72, 281–302. doi: 10.1086/346065

Kim, N. R., Kang, J. H., Kwon, O. W., Lee, S. J., Oh, J. H., and Chin, H. S. (2008). Association between complement factor H gene polymorphisms and neovascular age-related macular degeneration in Koreans. *Invest. Ophthalmol. Vis. Sci.* 49, 2071–2076. doi: 10.1167/iovs.07-1195

Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J.-Y., Sackler, R. S., Haynes, C., et al. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science* 308, 385–389. doi: 10.1126/science.1109557

Klein, R., Klein, B. E., and Cruickshanks, K. J. (1999). The prevalence of age-related maculopathy by geographic region and ethnicity. *Prog. Retin. Eye Res.* 18, 371–389.

Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., et al. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319, 1100–1104. doi: 10.1126/science.1153717

Li, Y., Xu, L., Jonas, J. B., Yang, H., Ma, Y., and Li, J. (2006). Prevalence of age-related maculopathy in the adult population in China: the Beijing eye study. *Am. J. Ophthalmol.* 142, 788–793. doi: 10.1016/j.ajo.2006.06.001

López Herráez, D., Bauchet, M., Tang, K., Theunert, C., Pugach, I., Li, J., et al. (2009). Genetic variation and recent positive selection in worldwide human populations: evidence from nearly 1 million SNPs. *PLoS ONE* 4:e7888. doi: 10.1371/journal.pone.0007888

Magalhaes, T. R., Casey, J. P., Conroy, J., Regan, R., Fitzpatrick, D. J., Shah, N., et al. (2012). HGDP and HapMap analysis by ancestry mapper reveals local and global population relationships. *PLoS ONE* 7:e49438. doi: 10.1371/journal.pone.0049438

Melton, T., Clifford, S., Martinson, J., Batzer, M., and Stoneking, M. (1998). Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. *Am. J. Hum. Genet.* 63, 1807–1823. doi: 10.1086/302131

Miller, J. W. (2013). Age-related macular degeneration revisited–piecing the puzzle: the LXIX Edward Jackson memorial lecture. *Am. J. Ophthalmol.* 155, 1–35.e13. doi: 10.1016/j.ajo.2012.10.018

Mona, S., Grunz, K. E., Brauer, S., Pakendorf, B., Castrì, L., Sudoyo, H., et al. (2009). Genetic admixture history of Eastern Indonesia as revealed by Y-chromosome and mitochondrial DNA analysis. *Mol. Biol. Evol.* 26, 1865–1877. doi: 10.1093/molbev/msp097

Mori, K., Gehlbach, P. L., Kabasawa, S., Kawasaki, I., Oosaki, M., Iizuka, H., et al. (2007). Coding and noncoding variants in the CFH gene and cigarette smoking influence the risk of age-related macular degeneration in a Japanese population. *Invest. Ophthalmol. Vis. Sci.* 48, 5315–5319. doi: 10.1167/iovs.07-0426

Muñoz, B., West, S. K., Rubin, G. S., Schein, O. D., Quigley, H. A., Bressler, S. B., et al. (2000). Causes of blindness and visual impairment in a population of older Americans: The Salisbury Eye Evaluation Study. *Arch. Ophthalmol.* 118, 819–825. doi: 10.1001/archopht.118.6.819

Ng, T. K., Chen, L. J., Liu, D. T. L., Tam, P. O. S., Chan, W. M., Liu, K., et al. (2008). Multiple gene polymorphisms in the complement factor h gene are associated with exudative age-related macular degeneration in chinese. *Invest. Ophthalmol. Vis. Sci.* 49, 3312–3317. doi: 10.1167/iovs.07-1517

Nonyane, B. A. S., Nitsch, D., Whittaker, J. C., Sofat, R., Smeeth, L., Chakravarthy, U., et al. (2010). An ecological correlation study of late age-related macular degeneration and the complement factor H Y402H polymorphism. *Invest. Ophthalmol. Vis. Sci.* 51, 2393–2402. doi: 10.1167/iovs.09-4228

Patterson, N., Price, A. L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet* 2:e190. doi: 10.1371/journal.pgen.0020190

Pluzhnikov, A., Below, J. E., Konkashbaev, A., Tikhomirov, A., Kistner-Griffin, E., Roe, C. A., et al. (2010). Spoiling the whole bunch: quality control aimed at preserving the integrity of high-throughput genotyping. *Am. J. Hum. Genet.* 87, 123–128. doi: 10.1016/j.ajhg.2010.06.005

Pluzhnikov, A., Below, J. E., Tikhomirov, A., Konkashbaev, A., Roe, C., Nicolae, D., et al. (2008). "Differential bias in genotype calls between plates due to the effect of a small number of lower DNA quality and/or contaminated samples," in *Genetic Epidemiology*, Vol. 32 (Hoboken, NJ: WILEY-LISS), 676.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909. doi: 10.1038/ng1847

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Ramke, J., Brian, G., Naduvilath, T., Lee, L., and Qoqonokana, M. Q. (2012). Prevalence and causes of blindness and low vision revisited after 5 years of Eye Care in Timor-Leste. *Ophthal. Epidemiol.* 19, 52–57. doi: 10.3109/09286586.2011.645108

Rosenberg, N. A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J. K., and Feldman, M. W. (2005). Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genet.* 1:e70. doi: 10.1371/journal.pgen.0010070

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., et al. (2002). Genetic structure of human populations. *Science* 298, 2381–2385. doi: 10.1126/science.1078311

Schaumberg, D. A., Chasman, D., Morrison, M. A., Adams, S. M., Guo, Q., Hunter, D. J., et al. (2010). Prospective study of common variants in the retinoic acid receptor-related orphan receptor α gene and risk of neovascular age-related macular degeneration. *Arch. Ophthalmol.* 128, 1462–1471. doi: 10.1001/archophthalmol.2010.261

Shtir, C. J., Marjoram, P., Azen, S., Conti, D. V., Le Marchand, L., Haiman, C. A., et al. (2009). Variation in genetic admixture and population structure among Latinos: the Los Angeles Latino eye study (LALES). *BMC Genet.* 10:71. doi: 10.1186/1471-2156-10-71

Silveira, A. C., Morrison, M. A., Ji, F., Xu, H., Reinecke, J. B., Adams, S. M., et al. (2010). Convergence of linkage, gene expression and association data demonstrates the influence of the RAR-related orphan receptor alpha (RORA) gene on neovascular AMD: a systems biology based approach. *Vision Res.* 50, 698–715. doi: 10.1016/j.visres.2009.09.016

Sivakumaran, T. A., Igo, R. P. Jr., Kidd, J. M., Itsara, A., Kopplin, L. J., Chen, W., et al. (2011). A 32 kb critical region excluding Y402H in CFH mediates risk for age-related macular degeneration. *PLoS ONE* 6:e25598. doi: 10.1371/journal.pone.0025598

Souto, L., Gusmão, L., Ferreira, E., Amorim, A., Côrte-Real, F., and Vieira, D. N. (2006). Y-chromosome STR haplotypes in East Timor: forensic evaluation and population data. *Forensic Sci. Int.* 156, 261–265. doi: 10.1016/j.forsciint.2005.02.015

Stoneking, M., and Delfin, F. (2010). The human genetic history of East Asia: weaving a complex tapestry. *Curr. Biol.* 20, R188–R193. doi: 10.1016/j.cub.2009.11.052

Swaroop, A., Branham, K. E., Chen, W., and Abecasis, G. (2007). Genetic susceptibility to age-related macular degeneration: a paradigm for dissecting complex disease traits. *Hum. Mol. Genet.* 16, R174–R182. doi: 10.1093/hmg/ddm212

Swaroop, A., Chew, E. Y., Rickman, C. B., and Abecasis, G. R. (2009). Unraveling a multifactorial late-onset disease: from genetic susceptibility to disease mechanisms for age-related macular degeneration. *Annu. Rev. Genom. Hum. Genet.* 10, 19–43. doi: 10.1146/annurev.genom.9.081307.164350

Tabbada, K. A., Trejaut, J., Loo, J.-H., Chen, Y.-M., Lin, M., Mirazón-Lahr, M., et al. (2010). Philippine mitochondrial DNA diversity: a populated viaduct between Taiwan and Indonesia? *Mol. Biol. Evol.* 27, 21–31. doi: 10.1093/molbev/msp215

Tishkoff, S. A., Reed, F. A., Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., et al. (2009). The genetic structure and history of Africans and African Americans. *Science* 324, 1035–1044. doi: 10.1126/science.1172257

Tumonggor, M. K., Karafet, T. M., Downey, S., Lansing, J. S., Norquest, P., Sudoyo, H., et al. (2014). Isolation, contact and social behavior shaped genetic diversity in West Timor. *J. Hum. Genet.* 59, 494–503. doi: 10.1038/jhg.2014.62

Van Oven, M., van den Tempel, N., and Kayser, M. (2012). A multiplex SNP assay for the dissection of human Y-chromosome haplogroup O representing the major paternal lineage in East and Southeast Asia. *J. Hum. Genet.* 57, 65–69. doi: 10.1038/jhg.2011.120

Xu, Y., Guan, N., Xu, J., Yang, X., Ma, K., Zhou, H., et al. (2008). Association of CFH, LOC387715, and HTRA1 polymorphisms with exudative age-related macular degeneration in a northern Chinese population. *Mol. Vis.* 14, 1373–1381.

Yang, Z., Camp, N. J., Sun, H., Tong, Z., Gibbs, D., Cameron, D. J., et al. (2006). A variant of the HTRA1 gene increases susceptibility to age-related macular degeneration. *Science* 314, 992–993. doi: 10.1126/science.1133811

Zhang, H., Morrison, M. A., Dewan, A., Adams, S., Andreoli, M., Huynh, N., et al. (2008). The NEI/NCBI dbGAP database: genotypes and haplotypes that may specifically predispose to risk of neovascular age-related macular degeneration. *BMC Med. Genet.* 9:51. doi: 10.1186/1471-2350-9-51

Zhong, H., Shi, H., Qi, X.-B., Duan, Z.-Y., Tan, P.-P., Jin, L., et al. (2011). Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route. *Mol. Biol. Evol.* 28, 717–727. doi: 10.1093/molbev/msq247

Ziskind, A., Bardien, S., van der Merwe, L., and Webster, A. R. (2008). The frequency of the H402 allele of CFH and its involvement with age-related maculopathy in an aged Black African Xhosa population. *Ophthalmic Genet.* 29, 117–119. doi: 10.1080/13816810802216472