Distribution of rare *LOXL1* missense alleles, haplotypes and diplotypes suggests association with reduced risk of glaucoma-related exfoliation syndrome.

Rob P. Igo, Jr.[1,2], Tyler G. Kinzy[1,2], Jessica N. Cooke Bailey[1,2], Chiea Chuen Khor[3,4], Tin Aung[4], Robert Ritch[5], Arthur J. Sit[6], Richard K. Lee[7], Margaret Pericak-Vance[8], Jae H. Kang[9], Jonathan L. Haines[1,2], Louis R. Pasquale[5], Janey L. Wiggs[10]

Affiliations

1. Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH

2. Cleveland Institute for Computational Biology, Case Western Reserve University, Cleveland, OH

3. Genome Institute of Singapore, Singapore

4. Singapore Eye Research Institute, Singapore

5.Department of Ophthalmology, Icahn School of Medicine at Mount Sinai, New York, New York.

6. Department of Ophthalmology, Mayo Clinic, Rochester, MN

7. Department of Ophthalmology, Bascom Palmer Eye Institute, University of Miami, Miami FL

8. Hussman Institute of Human Genetics, University of Miami, Miami, FL

9. Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA.

10. Department of Ophthalmology, Harvard Medical School, Massachusetts Eye and Ear, Boston, MA

ABSTRACT

Purpose: Common *LOXL1* protein-altering variants are significant genetic risk factors for exfoliation syndrome (XFS) and the related secondary glaucoma (XFG). A rare *LOXL1* missense allele has been associated with protective effects in a Japanese cohort, suggesting that other rare alleles may also exhibit protective effects. The goal of this study was to assess the contributions of rare *LOXL1* variants to XFS/XFG risk in cases and controls from the United States.

Methods: *LOXL1* rare (minor allele frequency less than 1%) variants were identified from Humanexome BeadArray (Illumina) data for 1118 XFS/XFG cases and 3661 controls. Distribution of rare variants, haplotypes (defined using IMPUTE2) and diplotypes were examined using the Fisher's exact test. Rare variant allele distribution was confirmed in an independent set of primary open angle glaucoma (POAG)

controls and multi-ethnic datasets. Correlation of *LOXL1* common allele homozygosity with disease risk used data from gnomAD (gnomad.broadinstitute.org/) and an existing multi-ethnic meta-analysis. **Results:** Four rare *LOXL1* missense alleles were identified, and all were more common in controls (combined P= 7.6E-4), with two of these located in a LOXL1 intrinsic disordered region (IDR) known to be involved in LOXL1 aggregation. Haplotypes that included the rare or minor variants were more common in controls compared to cases (OR= 0.33, P=1.7E-8). Heterozygous diplotypes were significantly associated with reduced risk overall (OR= 0.45 P= 1.7 E-89) with the largest effects observed for diplotypes with more than one heterozygous genotype (OR= 0.05, P= 1.0E-39). A homozygous diplotype was associated with increased disease risk (OR= 6.8, P= 4.7E-157) and homozygosity was correlated with disease risk for common *LOXL1* variants across multi-ethnic populations (Pearson= 0.92, P<0.001). **Conclusions:** Using exome array data from XFS/XFG cases and controls from the United States, we identify 4 rare protective *LOXL1* missense variants and show that the distribution of the corresponding haplotypes and diplotypes are associated with reduced risk of XFS/XFG. The diplotype results also demonstrate that *LOXL1* allelic heterozygosity is protective while homozygosity is associated with increased disease risk. These results suggest that *LOXL1* minor allele frequency variation among populations, with corresponding variation in genotype heterozgyosity and homozygosity, determines the XFS/XFG association effects and that genotypic effects may also impact protein aggregation involving intrinsic disordered regions.

INTRODUCTION

Exfoliation syndrome (XFS) is a systemic disorder characterized by progressive accumulation of abnormal fibrillar protein aggregates that can obstruct drainage of fluid from the eye, raising intraocular pressure (IOP) and causing exfoliation-related secondary glaucoma (XFG). XFG is the most common secondary glaucoma worldwide and is a major cause of irreversible blindness (Nazarali, 2018). XFS is also

associated with pre-mature cataract formation and complications during cataract surgery. Systemic conditions have also been associated with XFS, including cardiovascular disease (Chung, 2018), sleep apnea (Shumway, 2021), and obstructive pulmonary disease (Taylor, 2019).

LOXL1 (lysyl oxidase-like 1) is a major genetic risk factor for XFS/XFG in all populations examined (Aung, 2017), with *LOXL1* risk variants occurring in up to 98% of patients (Fan, 2011). Aggregated LOXL1 protein has been shown to be one component of the exfoliation fibrillar material (Zenkel, 2014; Sharma, 2009) that has been identified in extracellular spaces throughout the body but in particular in the ocular anterior segment (Schlötzer-Schrehardt, 2009). The LOXL1 protein includes several intrinsic disordered regions (IDRs) that could be sites of initiation of protein aggregates, and recent studies suggest that deletion of the most disordered regions reduces protein aggregation (Bernstein, 2019).

Two of the major *LOXL1* risk variants, (G153D and R141L) are common missence alleles located within the IDR with high likelihood of aggregation (Bernstein, 2019). In particular, the G153D variant falls within the largest protein segment with the highest disorder probability (**Figure 1**). Of interest, in most populations studied, the common allele 'G' at 153 and the common 'R' allele at 141 are associated with increased risk; however, in East Asian populations the 'L' allele at 141 is associated with increased risk (Hayashi 2008), and in South Africans, the 'D' allele at 153 is associated with increased risk (Williams 2010). Interestingly, both R141 'L' and G153 'D' are the 'common' alleles in East Asian and South African populations, respectively. A number of studies have investigated the role of *LOXL1* genetic variants including effects on protein function (Sharma 2016), gene expression (Pasutto, 2017) and dysregulation related to a long noncoding RNA also located within the LOXL1 genomic region (Hauser 2015). However, none of these studies have observed a consistent effect for the associated risk allele in all populations, which limits the determination of the underlying disease causing mechanism(s).

Recently, a rare *LOXL1* protein altering variant was associated with significantly reduced risk in a population of Japanese cases and controls (Aung, 2017). This finding, as well as the observation that the less frequent allele at G153D and R141L is protective in all populations studied so far, suggests that other rare or low frequency alleles may also be protective. To investigate this hypothesis, we examined the effects on risk of rare *LOXL1* coding variants in a cohort of XFS cases and controls with European descent from the United States.

METHODS

Study participants

This study adhered to the tenets of the Declaration of Helsinki and has been reviewed and approved by the Institutional Review Boards of the Massachusetts Eye and Ear Infirmary, Harvard School of Public Health, and the Brigham and Women's Hospital. Informed consent was obtained from the participants after explanation of the nature and possible consequences of the study.

Exfoliation cases were recruited from the Glaucoma service of the Massachusetts Eye and Ear Infirmary (MEEI, Harvard), Duke University School of Medicine Ophthalmology Department, the Glaucoma service of the Mayo Clinic, the Ophthalmology department from the University of Iowa School of Medicine, the Bascom Palmer Eye Institute (University of Miami), and the Nurses' Health Study (NHS) and Health Professionals Follow-up Study (HPFS). The controls were previously genotyped samples from the NHS and HPFS and were over the age of 40 and self-reported Caucasians with European ancestry and are representative of the United States population (Kang 2016). Controls have no evidence of XFS by clinical exam or medical record or repeatedly denied a self-report of glaucoma on biennial questionnaires administered over two decades. We also accessed genotype data from the MEEI and NEIGHBOR

controls. Detailed information on this dataset has been previously described (Wiggs, 2012; Margeta 2020).

All XFS cases are self-reported European ancestry Caucasians and are over the age of 40 and have documentation of characteristic ocular exfoliation material at the pupil margin or surface of the ocular lens, either though clinical exam or medical records. Clinical examination included measurement of visual acuity and intraocular pressure, slit lamp biomicroscopy, and fundoscopy. Cases also had visual field assessment primarily using the Humphrey automated visual fields. Individuals were excluded if other types of glaucoma (pigment dispersion, steroid-induced, uveitis) were evident on exam. All cases provided blood samples for DNA extraction.

Genotyping and quality control

Cases were genotyped at the Center for Inherited Disease Research (CIDR; http://www.cidr.jhmi.edu/) using the Illumina OmniExpress+Exome platform that includes 700,000 common SNPs (MAF >0.3) and 250,000 rare or low frequency functional exonic SNPs collected from 12,000 exomes (http://genome.sph.umich.edu/wiki/Exome_Chip). The XFS/XFG controls were previously genotyped using the Omni Express platform. The NEIGHBOR control data set was also genotyped using the Illumina HumanExome BeadArray (Illumina, Inc., San Diego, CA) at CIDR.

Quality control for human exome array genotype data (Igo 2016) for all cases and controls was carried out as follows. The Illumina Genome Studio (Illumina) and PLINK (Purcell 2007) were used for all quality controls (QC) steps except where noted. Basic QC for samples included screens for call rate (≥98.5%) and high (≥95%) concordance with a previous Illumina 660K panel run on the same sample where available (about 80% of samples). We verified recorded sex in the clinical records with genotyped sex by two criteria: mean fluorescence intensity on the X and Y chromosomes, plus genotype heterozygosity on the X chromosome and call rate on the Y, allowing male and female samples to have heterozygous X-linked

and successful Y-linked genotypes, respectively. We tested samples for pairwise relationships and unexpected duplication using KING (Manichaikul, 2010).

We verified European ancestry from the first two principal components derived from genotypes at 9000 ancestry-informative markers by means of the SNP weights program (Chen, 2013), including representative HapMap CEU, YRI, CHB, and JPT samples as reference populations. Moreover, we conducted a principal components analysis over 52,040 independent (pairwise $r^2 < 0.1$), common (MAF ≥0.005) SNPs using the smartpca program in EIGENSOFT (Price, 2006) to detect finer population structure.

Initial QC screens for markers included call rate (≥98%) and consistency with Hardy-Weinberg proportions ($P > 10^{-6}$ by Fisher's exact test). We screened markers for differences in allele frequency between whole-genome-amplified DNA samples and all other samples by the Fisher's exact test and removed from analysis all markers with $P < 0.0001$. All pseudoautosomal, Y-linked, and mitochondrial SNPs were subject to review in Illumina Genome Studio (Illumina, Inc.), and if necessary, were manually re-clustered. We also confirmed genotype clustering for rare (MAF <0.02) variants from fluorescence intensity data by means of zCall, run with a stringent z score threshold of $z = 21$ for calling heterozygous genotypes. Every rare SNP with two or more additional heterozygous calls by zCall than by GenCall was reviewed in Genome Studio, and if necessary, cluster locations were adjusted manually. Data for four rare coding variants (G120S, S159A, A160P, S427F) included on the genotyping platform were extracted for analyses.

Analyses

Distribution of the four rare *LOXL1* variants in cases compared to controls was assessed using the Fisher's exact test. Haplotypes were constructed using IMPUTE2 (Howie 2009) and haplotype distribution between cases and controls was also assessed using the Fisher's exact test. Diplotypes were identified using the haplotype data for each study participant. Pearson's correlation was used to assess percent homozygosity with disease risk for common *LOXL1* risk alleles.

RESULTS

Cases and controls. After sample and genotyping quality control, data was available for 1,118 cases and 3,661 XFS/XFG controls. Additionally, genotype data was available for 2,606 primary open angle glaucoma (POAG) subjects without XFS/G from the NEIGHBOR study (Wiggs 2013). Cases and controls have similar proportions of females (**Table 1**). The cases are on average older than both control cohorts; however, the overall difference was not significant (P= 0.26).

Distribution of rare missense alleles in XFS cases and controls suggest protective effects. We identified four *LOXL1* missense alleles with minor allele frequencies (MAFs) less than 1%: G120S, S159A, A160P and S427F (**Table 2**). Three of these have Combined Annotation Dependent Depletion (CADD) (Kircher 2014) scores >20, indicating potential deleterious effects on protein structure/function. Two missense alleles (S159A and A160P) are located within the *LOXL1* intrinsic disordered region (IDR) with high propensity to protein aggregation (**Figure 1**). All four missense variants were observed more often in controls compared to cases (**Table 2**) for both the XFS/XFG (combined P= 7.6E-4) and NEIGHBOR control datasets (combined P= 2.2E-3). Using data from an international XFS meta-analysis (Aung 2017), we assessed the allelic distributions for these variants in other populations. While these rare variants are not present in many populations, we were able to observe consistent protective effects in an Icelandic case control study for A160P and S427F and in a South African cohort for S159A (**Table 3**). G120S was

not associated (P> 0.05) with XFS in any other population (1 of 458 cases from Italy and 1 of 2827 cases from Japan and not in any controls from either population (267 and 3013 respectively).

LOXL1 rare haplotypes are more common in controls than in cases. To further explore the protective effects of the rare missense alleles, we investigated the distribution of haplotypes defined by the 4 missense alleles as well as the two common *LOXL1* variants, G153D and R141L. All haplotypes that include a rare (minor) missense allele are found more often in controls compared to cases (OR = 0.33, P=1.7E-173) (**Table 4**). In addition, only the haplotype that includes the common allele at all 6 missense variants was adversely associated with disease risk (OR = 4.76, P= 1.7E-173).

LOXL1 heterozygous diplotypes are more common in controls than in cases. As rare alleles are more likely to be present as heterozygous genotypes, we examined the distribution of the 6-variant diplotypes (derived from the haplotypes in Table 4) among XFS/XFG cases and controls. Similar to the haplotype distribution, only the homozygous diplotype comprised of the common alleles (Diplotype A, **Table 5**) is associated with increased disease risk (OR= 6.78, P=4.72E-157). All diplotypes that include heterozygous genotypes are protective with the exception of diplotype F, which was associated with an OR of 1.64 (**Table 5**) but was also very rare and the distribution between cases and controls was not statistically significant (P= 0.55). Overall heterozygous diplotypes are significantly protective (OR= 0.45, P=4.2E-89) (**Table 5**).

Protective effects of diplotypes are correlated with the number of heterozygous variants. Among the diplotypes identified, the number of heterozygous variants varies from 0 to 3. All diplotypes with heterozygous variants exhibit protective effects, however the observed ORs are lowest for diplotypes with more than one heterozygous variant (OR= 0.05, P= 1.0E-39) (**Table 5, Figure 2**). In fact, diplotypes

with 2 or 3 heterozygous variants are only observed in controls while diplotypes with 1 heterozygous variant are found in both cases and controls, but more commonly in controls (**Table 5**). Among the single heterozygous diplotypes, the G153D exhibits the most protective effects (Diplotype D, OR= 0.06) followed by S427F (Diplotype E, OR=0.47) and R141L (Diplotype C, OR= 0.70). The distribution of diplotypes among cases and controls does not vary substantially with age in this population (P>0.05, **Figure 3**).

Homozygosity for common G153D and R141L alleles is correlated with disease risk. Considering the consistent protective effects for heterozygous variants, we sought to examine the effects of homozygosity on disease risk. As homozygous genotypes are not found for the rare variants this analysis was limited to the alternate alleles for the common variants: allele 'D' for G153D and allele 'L' for R141L. We used population specific data for the alternate alleles from gnomAD (gnomad.broadinstitute.org/) and association results for similar populations from the XFS international meta-analysis (Aung 2017). **Figure 4** shows that the percent homozygosity for the alternate allele for G153D and also for R141L is strongly correlated with the risk effects for the same allele (Pearson = 0.92, P<0.001). The correlation is evident for populations where the alternate allele is the minor allele, and is also evident for populations where the 'alternate' allele is actually the common allele (i.e., the L allele at R141L is the 'common' allele in East Asians). Indeed this result suggests that the population minor allele frequency is a key factor in determining the observed associations due to the percent homozygosity with the most common alleles having highest percent homozgyosity and highest disease risk.

DISCUSSION

Common *LOXL1* coding variants are major XFS/XFG genetic risk factors, but despite the strong association with disease risk, the underlying pathogenic mechanism(s) are not known. The discovery of

a protective rare *LOXL1* variant in a Japanese case control study provided the first insight that *LOXL1* protein-altering variants could reduce disease risk. In this study, we have expanded this initial observation to include 4 other rare *LOXL1* missense alleles that are all more commonly found in unaffected controls compared to cases in a U.S. population with European ancestry. Of these variants, some are also found in other populations where they also exhibit protective effects. We also show that haplotypes and diplotypes that include these rare coding variants are more common in controls and that overall heterozygosity is protective and homozygosity is associated with increasing disease risk.

Only the diplotype that includes common homozygous variants at all 6 variants (diplotype A) is associated with increased disease risk. Interestingly, only one other homozygous diplotype was observed (diplotype B), and this diplotype, while rare, is found more often in controls. Diplotype B includes the homozygous 'alternate' R141 allele L that is frequently found as a heterozygous variant in the protective diplotypes that also include heterozygous rare variants. Homozygosity of R141L alone would be expected to be associated with increased risk, and it is possible that other, as yet undetected rare variants, are present on this diplotype that are underlying the protective effects. One diplotype (diplotype F) that includes a heterozygous rare variant is found more commonly in cases, however this association was not statistically significant. In this population, diplotype C that includes the single heterozygous genotype for R141L is less protective than diplotype D that includes the single heterozygous genotype for G153D. Accordingly, the cumulative protective effects of rare variants is more evident in diplotypes that include the 'L' allele for R141 and do not include the 'D' allele for G153.

Two of the rare variants as well as the common variant G153D are located within a LOXL1 intrinsic disordered region that has been shown to be correlated with protein aggregation (Bernstein 2019). This observation, coupled with the protective effects of heterozygosity, including the profound protective

effect of heterozygous G153D, suggests that protein variation within this region could reduce protein aggregation. IDRs can contribute to pathogenic aggregation for several neurodegenerative diseases including Alzheimer's, Parkinson's, Huntington's, and prion-related diseases, among others (Eftekharzadeh 2016). Understanding the nature of aggregation induced by IDRs is an area of intense investigation. Protein variants may increase or decrease aggregation (Candelise 2021). A missense allele protective for Alzheimer disease has been shown to reduce abeta aggregation (Das, 2017), supporting a hypothesis that heterozygous LOXL1 variants could potentially reduce protein aggregation through a similar mechanism.

In some populations, the risk allele at the common *LOXL1* variants (G153D and R141L) are reversed making it difficult to evoke a specific biological mechanism that explains disease risk. For example, a molecular interaction with the allele associated with risk or allele specific gene expression influencing risk would have to be population dependent. While possible, it seems unlikely that a gene variant could underly a drastically different function in one population compared to another. The results from this study, showing a strong correlation between homozygosity and disease risk combined with overall protective effects observed for heterozygous diplotypes supports a novel hypothesis that the observed risk allele reversal in some populations is simply the result of variation in minor allele frequency leading to variation in relative homozygosity and heterozygosity of risk variants. For example, R141 allele 'L' is the minor allele in most populations and is also associated with protective effects in most populations. However, in East Asians, where 'L' is the risk allele (Hayashi 2008), the 'L' allele is the more frequent allele and therefore is more likely to be homozygous and associated with disease risk. A similar argument can be made for the reversed risk observed for the G153D variant in South Africans (Williams 2010). This hypothesis also provides an explanation for the protective effects of rare *LOXL1* protein coding variants as these are likely to be heterozygous due to their low minor allele frequencies. Further

work will be required to confirm this hypothesis including testing heterozygous versus homozygous variants for protective functional effects.

There are several limitations of our study. First, our analyses are based on exome array data which does not comprehensively sample all *LOXL1* variants. Only the variants present on the array could be examined in this study. In future work, comprehensive sequence data will be necessary to more completely examine the effects of *LOXL1* protein altering variants on disease risk. Second, the differences in average ages for cases and controls may influence these results; however, the overall difference in age was not significant, and the distribution of diplotypes among case and control age groups was not significantly different. Finally, because only one haplotype including a rare variant did not also include the protective allele for the common variants, we could not examine the effects of rare variants independent of the common risk alleles. However, the observation that the increased number of risk variants correlates with increased protective effects supports an effect of rare variants independent of the common risk alleles.

In summary, we identify 4 rare protective *LOXL1* protein coding variants and show that the distribution of the haplotypes and diplotypes that include these variants suggest an association with reduced risk of XFS/XFG. We also show that only a homozygous diplotype comprised of all common alleles is significantly associated with disease risk and that homozygosity of the common risk variants G153D and R141L is correlated with disease risk across multi-ethnic populations. These results support a novel hypothesis that the reversal of *LOXL1* risk alleles observed in some populations is dependent on variation of minor allele frequencies among populations and that *LOXL1* allelic homozygosity is associated with increased disease risk potentially due to effects on protein aggregation involving

intrinsic disordered regions. Further work identifying additional protective variants and testing effects on protein aggregation will be necessary to confirm this hypothesis.

ACKNOWLEDGEMENTS

Supported by NIH/NEI R01 EY020928, EY015473, and an unrestricted challenge grant from Research to Prevent Blindness (NYC to Icahn School of Medicine at Mount Sinai)

REFERENCES

Aung T, et al., Genetic association study of exfoliation syndrome identifies a protective rare variant at LOXL1 and five new susceptibility loci Nat Genet. 2017 Jul;49(7):993-1004.

Bernstein AM, Ritch R, Wolosin JM. LOXL1 folding in exfoliation glaucoma.

Adv Protein Chem Struct Biol. 2019;118:273-288.

Candelise N, Scaricamazza S, Salvatori I, Ferri F, Valle C, Manganelli V, Garofalo T, Sorice M, Misasi R. Protein Aggregation Landscape in Neurodegenerative Diseases: Clinical Relevance and Future Applications Int J Mol Sci. 2021 Jun 2;22(11):6016.

Chen CY, Pollack S, Hunter DJ, Hirschhorn JN, Kraft P, Price AL. Improved ancestry inference using weights from external reference panels. *Bioinformatics*. 2013; 29: 1399–140.

Chung H, Arora S, Damji KF, Weis E. Association of pseudoexfoliation syndrome with cardiovascular and cerebrovascular disease: a systematic review and meta-analysis. Can J Ophthalmol. 2018 Aug;53(4):365-372.

Das P, Chacko AR, Belfort G. Alzheimer's Protective Cross-Interaction between Wild-Type and A2T Variants Alters Aβ 42 Dimer Structure ACS Chem Neurosci. 2017 Mar 15;8(3):606-618.

Eftekharzadeh B, Hyman BT, Wegmann S. Structural studies on the mechanism of protein aggregation in age related neurodegenerative diseases. Mech Ageing Dev. 2016 Jun; 156():1-13.

Fan BJ, Pasquale LR, Rhee D, Li T, Haines JL, Wiggs JL LOXL1 promoter haplotypes are associated with exfoliation syndrome in a U.S. Caucasian population. . Invest Ophthalmol Vis Sci. 2011 Apr 12;52(5):2372-8.

Hauser MA, Aboobakar IF, Liu Y, Miura S, Whigham BT, Challa P, Wheeler J, Williams A, Santiago-Turla C, Qin X, Rautenbach RM, Ziskind A, Ramsay M, Uebe S, Song L, Safi A, Vithana EN, Mizoguchi T, Nakano S, Kubota T, Hayashi K, Manabe S, Kazama S, Mori Y, Miyata K, Yoshimura N, Reis A, Crawford GE, Pasutto F, Carmichael TR, Williams SE, Ozaki M, Aung T, Khor CC, Stamer WD, Ashley-Koch AE, Allingham RR. Genetic variants and cellular stressors associated with exfoliation syndrome modulate promoter activity of a lncRNA within the LOXL1 locus. Hum Mol Genet. 2015 Nov 15;24(22):6552-63.

Hayashi H, Gotoh N, Ueda Y, Nakanishi H, Yoshimura N. Lysyl oxidase-like 1 polymorphisms and exfoliation syndrome in the Japanese population. Am J Ophthalmol. 2008 Mar;145(3):582-585.

Howie BN, Donnelly P and Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genetics* 5(6): e1000529.

Igo RP Jr, Cooke Bailey JN, Romm J, Haines JL, Wiggs JL. Quality Control for the Illumina HumanExome BeadChip. Curr Protoc Hum Genet. 2016 Jul 1;90:2.14.1-2.14.16.

Kang JH, Wu J, Cho E, Ogata S, Jacques P, Taylor A, Chiu CJ, Wiggs JL, Seddon JM, Hankinson SE, Schaumberg DA, Pasquale LR. Contribution of the Nurses' Health Study to the Epidemiology of Cataract, Age-Related Macular Degeneration, and Glaucoma. Am J Public Health. 2016 Sep;106(9):1684-9.
Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014 Mar;46(3):310-5.

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics*. 2010; 26: 2867–2873.

Margeta MA, Letcher SM, Igo RP Jr, Cooke Bailey JN, Pasquale LR, Haines JL, Butovsky O, Wiggs JL; NEIGHBORHOOD consortium. Association of *APOE* With Primary Open-Angle Glaucoma Suggests a Protective Effect for *APOE* epsilon4. Invest Ophthalmol Vis Sci. 2020 Jul 1;61(8):3.

Nazarali S, Damji F, Damji KF. What have we learned about exfoliation syndrome since its discovery by John Lindberg 100 years ago? Br J Ophthalmol. 2018 Oct;102(10):1342-1350.

Pasutto F, Zenkel M, Hoja U, Berner D, Uebe S, Ferrazzi F, Schödel J, Liravi P, Ozaki M, Paoli D, Frezzotti P, Mizoguchi T, Nakano S, Kubota T, Manabe S, Salvi E, Manunta P, Cusi D, Gieger C, Wichmann HE, Aung T, Khor CC, Kruse FE, Reis A, Schlötzer-Schrehardt U. Pseudoexfoliation syndrome-associated genetic variants affect transcription factor binding and alternative splicing of LOXL1. Nat Commun. 2017 May 23;8:15466.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006; 38: 904–909.

Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81: 559–575.

Schlötzer-Schrehardt U. Molecular pathology of pseudoexfoliation syndrome/glaucoma--new insights from LOXL1 gene associations. Exp Eye Res. 2009 Apr; 88(4):776-85.

Sharma S, Chataway T, Burdon KP, Jonavicius L, Klebe S, Hewitt AW, Mills RA, Craig JE. Identification of LOXL1 protein and Apolipoprotein E as components of surgically isolated pseudoexfoliation material by direct mass spectrometry. Exp Eye Res. 2009 Oct;89(4):479-85.

Sharma S, Martin S, Sykes MJ, Dave A, Hewitt AW, Burdon KP, Ronci M, Voelcker NH, Craig JE. Biological effect of LOXL1 coding variants associated with pseudoexfoliation syndrome. Exp Eye Res. 2016 May;146:212-223.

Shumway C, Curtin K, Taylor S, Sundar KM, Wirostko BM, Ritch R. Association between Obstructive Sleep Apnea and Exfoliation Syndrome: The Utah Project on Exfoliation Syndrome. Ophthalmol Glaucoma. 2021 May-Jun;4(3):260-267.

Taylor SC, Bernhisel AA, Curtin K, Allingham RR, Ritch R, Wirostko BM. Association between Chronic Obstructive Pulmonary Disease and Exfoliation Syndrome: The Utah Project on Exfoliation Syndrome. Ophthalmol Glaucoma. 2019 Jan-Feb;2(1):3-10.

Wiggs JL, Hauser MA, Abdrabou W, Allingham RR, Budenz DL, Delbono E, Friedman DS, Kang JH, Gaasterland D, Gaasterland T, Lee RK, Lichter PR, Loomis S, Liu Y, McCarty C, Medeiros FA, Moroi SE, Olson LM, Realini A, Richards JE, Rozsa FW, Schuman JS, Singh K, Stein JD, Vollrath D, Weinreb RN, Wollstein G, Yaspan BL, Yoneyama S, Zack D, Zhang K, Pericak-Vance M, Pasquale LR, Haines JL. The NEIGHBOR consortium primary open-angle glaucoma genome-wide association study: rationale, study design, and clinical variables. J Glaucoma. 2013 Sep;22(7):517-25.

Wiggs JL, Yaspan BL, Hauser MA, Kang JH, Allingham RR, Olson LM, Abdrabou W, Fan BJ, Wang DY, Brodeur W, Budenz DL, Caprioli J, Crenshaw A, Crooks K, Delbono E, Doheny KF, Friedman DS, Gaasterland D, Gaasterland T, Laurie C, Lee RK, Lichter PR, Loomis S, Liu Y, Medeiros FA, McCarty C, Mirel D, Moroi SE, Musch DC, Realini A, Rozsa FW, Schuman JS, Scott K, Singh K, Stein JD, Trager EH, Vanveldhuisen P, Vollrath D, Wollstein G, Yoneyama S, Zhang K, Weinreb RN, Ernst J, Kellis M, Masuda T, Zack D, Richards JE, Pericak-Vance M, Pasquale LR, Haines JL. Common variants at 9p21 and 8q22 are associated with increased susceptibility to optic nerve degeneration in glaucoma. PLoS Genet. 2012;8(4):e1002654.

Williams SE, Whigham BT, Liu Y, Carmichael TR, Qin X, Schmidt S, Ramsay M, Hauser MA, Allingham RR. Major LOXL1 risk allele is reversed in exfoliation glaucoma in a black South African population. Mol Vis. 2010 Apr 21;16:705-12.

Zenkel M, Schlötzer-Schrehardt U. The composition of exfoliation material and the cells involved in its production. J Glaucoma. 2014 Oct-Nov;23(8 Suppl 1):S12-4.

Tables and Figures

Table 1. Case and control demographics

	Exfoliation cases	Exfoliation controls	NEIGHBOR controls
N	1118	3661	2606
% Female	69	65	64
Age (average)	72.9 ± 9.6	54.0 ±7.8	65.5 ±13.5

Table 2. Distribution of rare *LOXL1* coding variants in XFS/XFG cases and controls

SNP [Minor allele]	**CADD Score**	**Case (N=2236) MAC (MAF)**	**Ctrl (N=7322) MAC (MAF)**	**OR (95%CI) P**	**NBR Ctrl (N=5212) MAC (MAF)**	**Combined Ctrls (N=12,534) MAC (MAF)**	**OR (95%CI) P**
rs201884374 G120S [A]	25.8	0 (0)	2 (2.7E-4)	NA	3 (5.8E-4)	5	NA
rs78803776 S159A [G]	20.6	0 (0)	2 (2.7E-4)	NA	2 (3.8E-4)	4	NA
rs138183635 A160P [C]	26.6	1 (4.5E-4)	19 (2.6E-3)	0.16 (0.02, 1.19) 0.06	8 (1.5E-3)	27	0.19 (0.03, 1.38) 0.11
rs189632283 S427F [T]	9.2	1 (4.5E-4)	21 (2.7E-3)	0.15 (0.02, 1.28) 0.04	8 (1.5E-3)	29	0.17 (0.02, 1.25) 0.08
Total		2	44	0.15 (0.04, 0.62) 7.6E-4	21	65	0.17 (0.04, 0.70) 2.2E-3

Rare variant distribution in 1118 XFS/XFG cases (N=2236 alleles) and 3661 XFS/XFG controls (N=7322 alleles) using Fisher's exact test. Minor allele frequencies were also compared to 2606 primary open angle glaucoma (POAG) controls (NBR, NEIGHBOR) (N=5212 alleles) and combined controls (total 12,534 alleles). Abbreviations: SNP, single nucleotide polymorphism; CADD, combined annotation dependent depletion; Ctrl, control; OR, odds ratio; NBR, NEIGHBOR; MAC, minor allele count; MAF, minor allele frequency.

Table 3. Rare *LOXL1* coding variant association in replication cohorts

	SNP [Minor allele]			
	rs201884374 G120S [A]	**rs78803776 S159A [G]**	**rs138183635 A160P [C]**	**rs189632283 S427F [T]**
Population	OR (P value)	OR (P value)	OR (P value)	OR (P value)
South Africa 90 cases 250 controls	NP	0.21 (0.0004)	NP	NP
Iceland 195 cases 14,474 controls	NP	NP	0.17 (0.0014)	0.16 (0.026)

rs78803776[G] was only present in the South African sample, while rs138183635[C] and rs189632283[T] were only present in the Icelandic sample. Abbreviations: OR, Odds ratio; NP, not present.

Table 4. *LOXL1* haplotypes with common and rare coding variants and XFS/XFG association.

Hap #	G120S	R141L	G153D	S159A	A160P	S427F	Cases (N=2236)	Controls (N=7322)	P	OR (95%CI)
							N	N		
1	**A**	**T**	**G**	**T**	**G**	**C**	0	2	NA	NA
2	**G**	**G**	**A**	**T**	**C**	**C**	1	19	0.06	0.16 (0.02, 1.19)
3	**G**	**G**	**A**	**T**	**G**	**C**	34	1219	7.90E-106	0.08 (0.06, 0.11)
4	**G**	**G**	**G**	**G**	**G**	**C**	0	2	NA	NA
5	**G**	**G**	**G**	**T**	**G**	**C**	1869	3785	1.7E-173	4.76 (4.20, 5.36)
6	**G**	**T**	**G**	**T**	**G**	**C**	331	2274	3.74E-56	0.39 (0.34, 0.44)
7	**G**	**T**	**G**	**T**	**G**	**T**	1	21	0.04	0.15 (0.02, 1.28)
All rare							367	3537	1.7E-173	0.21 (0.19, 0.24)

The minor (rare) allele at each variant is highlighted in yellow. Only haplotype #5 (highlighted in orange) that includes all the major (common) alleles is associated with disease risk. Abbreviations: NA, not applicable; OR, odds ratio.

Table 5. *LOXL1* diplotypes and XFS/XFG association

Diplotype	G120S	R141L	G153D	S159A	A160P	S427F	Cases	Freq.	Controls	Freq.	P	OR (95%CI)
A (H5/H5)	G	G	G	T	G	C	786	0.703	947	0.259	4.72E-157	6.78 (5.87,7.77)
	G	G	G	T	G	C						
B (H6/H6)	G	T	G	T	G	C	18	0.016	448	0.122	6.43E-34	0.13 (0.08,0.21)
	G	T	G	T	G	C						
Sum Homozygous							**804**	**0.719**	**1395**	**0.38**	**4.17E-89**	**4.16 (4.76,3.60)**
C (H5/H6)	G	**G**	G	T	G	C	290	0.259	1220	0.333	2.97E-06	0.7 (0.80,0.61)
	G	**T**	G	T	G	C						
D (H5/H3)	G	G	**G**	T	G	C	15	0.013	659	0.18	1.79E-68	0.06 (0.10,0.04)
	G	G	**A**	T	G	C						
E (H7/H6)	G	T	G	T	G	**T**	1	8.9E-4	7	1.9E-3	0.69	0.47 (3.78,0.06)
	G	T	G	T	G	**C**						
F (H2/H3)	G	G	A	T	**C**	C	1	8.9E-4	2	5.5E-4	0.55	1.64 (0.15,17.9)
	G	G	A	T	**G**	C						
Sum one Heterozygous							**307**	**0.27**	**1888**	**0.516**	**5.70E-47**	**0.36 (0.41,0.31)**
G (H3/H6)	G	**G**	**A**	T	G	C	7	6.3E-3	343	0.093	4.73E-32	0.06 (0.13,0.03)
	G	**T**	**G**	T	G	C						
H (H5/H2)	G	G	**G**	T	**G**	C	0	0	6	0.002	0.34	NA
	G	G	**A**	T	**C**	C						
I (H7/H5)	G	**T**	G	T	G	**T**	0	0	13	0.004	0.048	NA
	G	**G**	G	T	G	**C**						
J (H6/H4)	G	**T**	G	**T**	G	C	0	0	1	2.7E-4	NA	NA
	G	**G**	G	**G**	G	C						
K (H3/H4)	G	G	**A**	**T**	G	C	0	0	1	2.7E-4	NA	NA
	G	G	**G**	**G**	G	C						
L (H1/H5)	**A**	**T**	G	T	G	C	0	0	1	2.7E-4	NA	NA
	G	**G**	G	T	G	C						
Sum 2 heterozygous							**7**	**0.0063**	**365**	**0.0997**	**1.21E-34**	**0.06 (0.13,0.03)**
M (H2/H6)	G	**G**	**A**	T	**C**	C	0	0	11	0.003	0.078	NA
	G	**T**	**G**	T	**G**	C						
N (H3/H7)	G	**G**	**A**	T	G	**C**	0	0	2	5.5E-4	NA	NA
	G	**T**	**G**	T	G	**T**						
Sum 3 heterozygous							**0**	**0**	**13**	**0.0036**	**0.048**	**NA**
Sum 2 + 3 heterozygous							**7**	**6.3E-3**	**378**	**0.10**	**1.03E-39**	**0.05 (0.10,0.02)**
Sum all heterozygous							**314**	**0.28**	**2266**	**0.62**	**4.17E-89**	**0.45 (0.52,0.39)**

Heterozygous genotypes are shown in bold type. Only the homozygous diplotype A (H5/H5) is significantly associated with risk. Overall 72% of cases have a homozygous diplotype compared to 38%

of controls. The diplotypes that include two heterozygous genotypes are found in 7 cases and 342 controls while the diplotypes with three heterozygous genotypes are only found in controls.

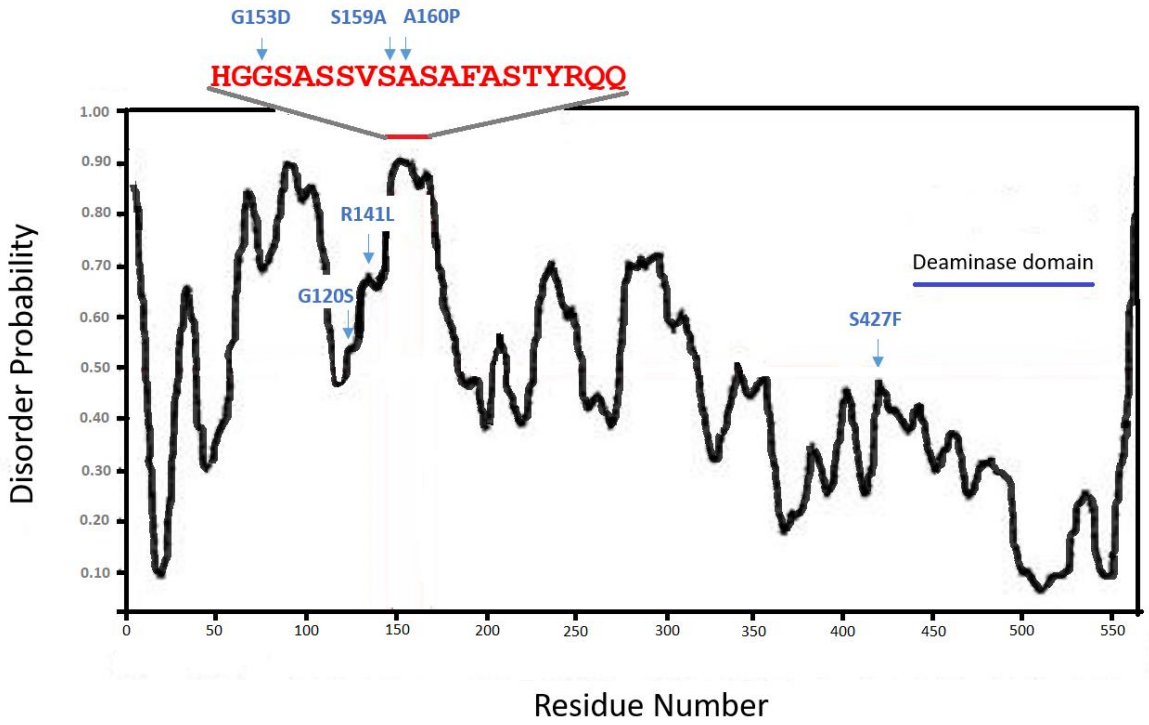

Figure 1. LOXL1 intrinsic disordered domains and location of protective variants. *LOXL1* variants are shown in blue font. The deaminase domain is indicated by the blue line. The region that includes G153D, S159A and A160P is the largest protein region with highest disordered probability indicating greatest likelihood of aggregation. Disordered probably greater than 50% is considered to be significant. The regions resulting in reduced aggregation when deleted were residues 28-95 and residues 125-185. This figure was modified from Bernstein et al. (Adv Protein Chem Struct Biol. 2019;118:273-288).

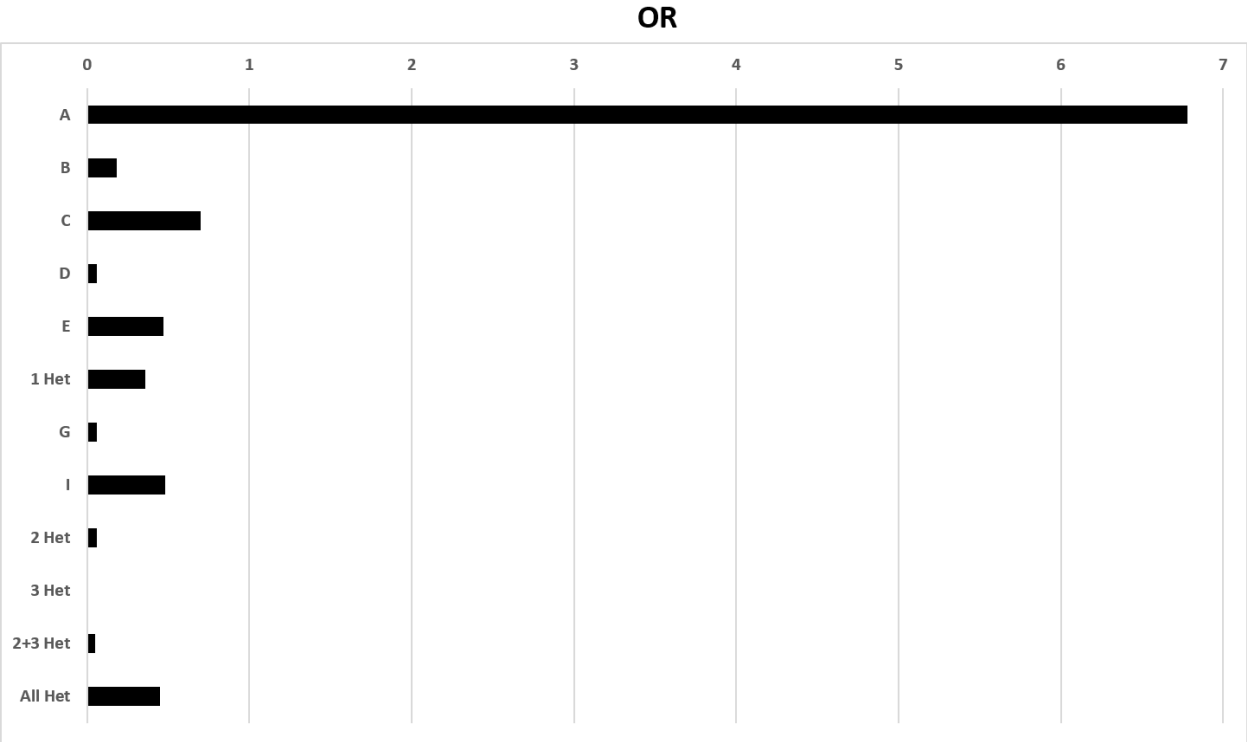

Figure 2. Odds ratios (OR) for *LOXL1* diplotypes. ORs for diplotypes with at least nominal evidence of association (P<0.05) are displayed on the x-axis for each diplotype labeled on the y-axis. The diplotype letter corresponds to the diplotype from Table 5. Only diplotype A that includes all homozygous genotypes is associated with increased XFS/XFG risk. All others have ORs less than 1 indicating protective effects.

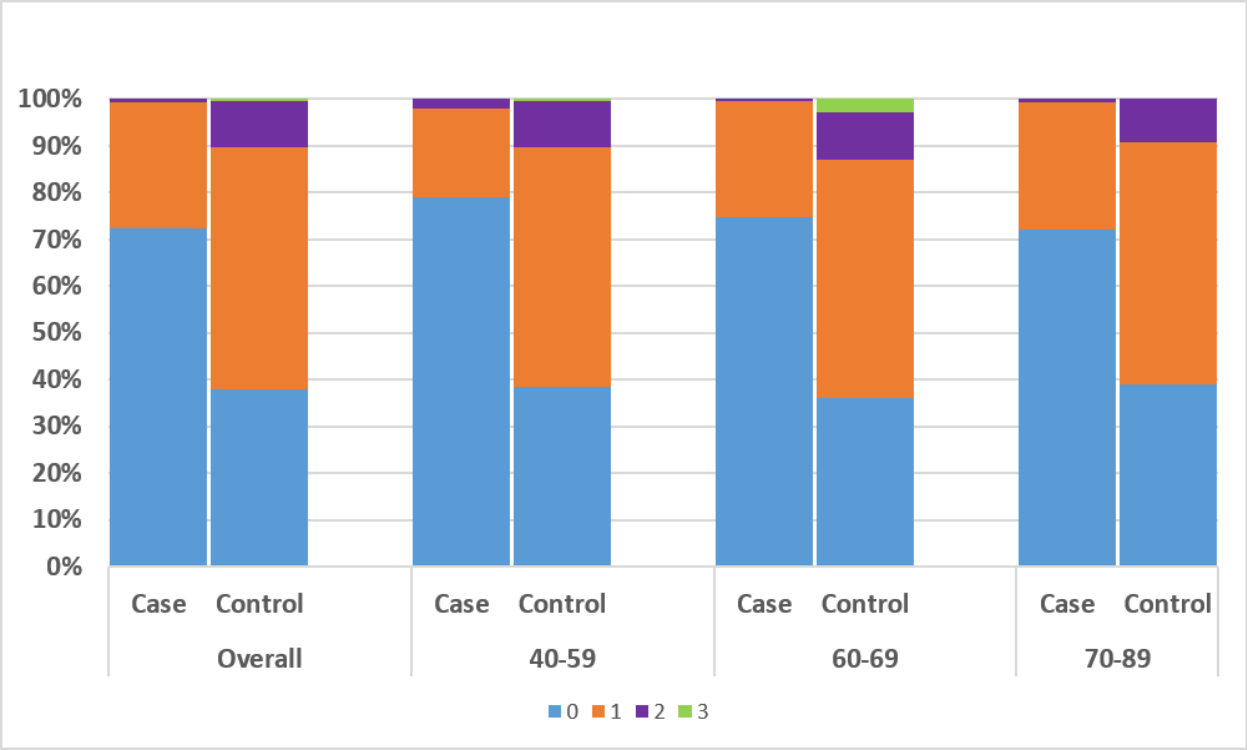

Figure 3. *LOXL1* diplotype distribution among cases and controls according to age. Total diplotypes with 0 heterozygous genotypes (ie homozygous) (blue), 1 heterozygous genotype (orange), 2 heterozygous genotypes (purple) and 3 heterozygous genotypes (green) were summed for cases and controls according to age groups: 40-59, 60-69, 70-89 and compared with the distribution among cases and controls overall. Similar distribution of diplotypes was observed for all age groups (P>0.05) with the exception of the rare diplotypes with 3 heterozygous variants which were found only in controls primarily in the 60-69 age range.

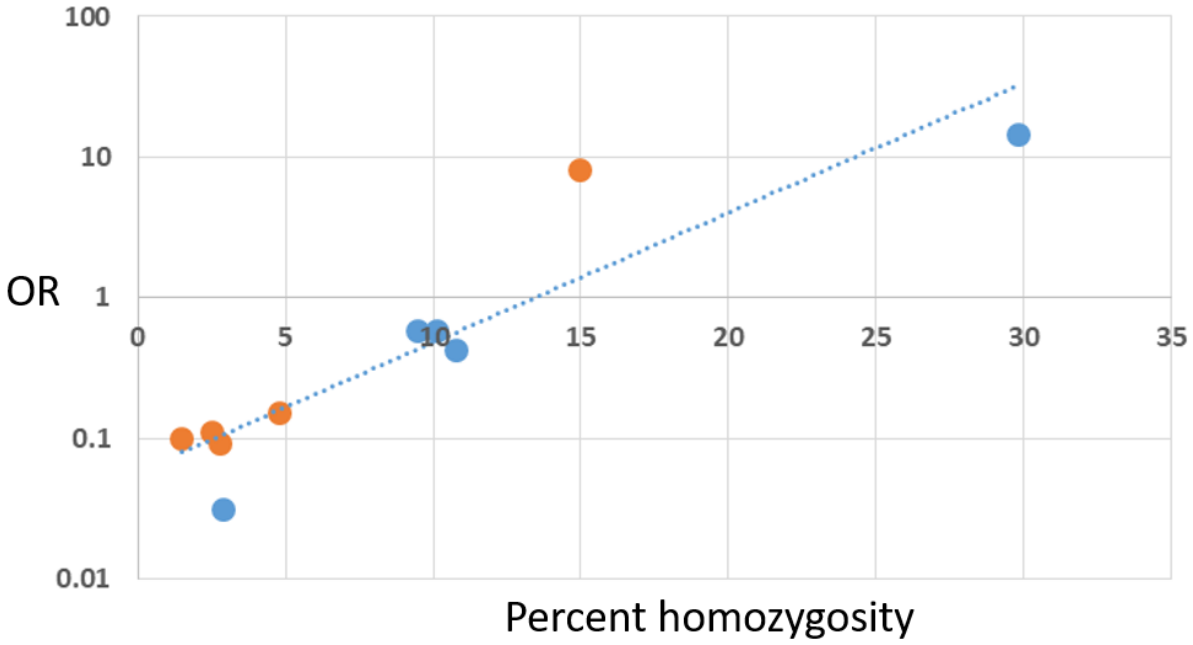

Figure 4. Correlation of *LOXL1* allele homozygosity for G153D and R141L with XFS/XFG risk (OR). The percent homozygosity for 5 populations for the 'alternate allele' for common missense variants G153D (allele D) (orange dots) and R141L (allele L) (blue dots) are displayed on the x-axis (linear scale) and the corresponding OR for XFS/XFG for that allele in the same population is shown on the y-axis (logarithmic scale). The regression line indicates excellent correlation (Pearson coefficient= 0.9233, P<0.001). The values for each population are shown in Table 6. Percent homozygosity was obtained from gnomAD (gnomad.broadinstitute.org/) and the ORs for XFS/XFG risk were obtained from Khor et al. (Nat Genet. 2017 Jul;49(7):993-1004.).

Table 6. Percent homozygosity and ORs for XFS/XFG for the 'alternate' allele for G153D and R141L in 5 populations.

SNP [alternate allele]	Population	Allele frequency	Percent homozygosity	XFS/XFG OR
G153D [D]	ENF	0.16	2.5	0.11
	EA	0.12	1.5	0.10
	AA	0.38	15	8.1
	LA	0.16	2.8	0.091
	SA	0.22	4.8	0.15
R141L [L]	ENF	0.32	10.8	0.42
	EA	0.54	29.8	14.6
	AA	0.17	2.9	0.031
	LA	0.31	9.5	0.58
	SA	0.32	10.1	0.59

Frequency= frequency of the alternate allele in the indicated population. Percent homozygosity is the fraction of the indicated population that is homozygous for the alternate allele and the OR is XFS/XFG risk estimate for the alternate allele in the indicated population. Note that for some populations the 'alternate allele' is the common allele. The percent homozygosity was obtained for each population from gnomAD (gnomad.braodinstititue.org/) and the ORs for XFS/XFG risk were obtained from Khor etal. (Nat Genet. 2017 Jul;49(7):993-1004.). ENF, European nonFinnish; EA, East Asian; AA, African, African-american; LA, Latino, Admixed American; SA, South Asian