

Common tandem repeat variants associated with glaucoma risk in individuals of African ancestry

Kenneth Pham^{1,2,3,4,5}, Roy Lee^{1,2}, Isabel Di Rosa^{1,2}, Rebecca Salowe^{1,2}, Fangming Jin^{2,6}, Yuki Bradford^{3,7}, Gui-Shuang Ying^{2,6}, Jennifer E. Phillips-Cremins^{3,4,5}, Shefali Setia-Verma^{7,8}, Joan O'Brien^{1,2*}

¹ Penn Medicine Center for Genetics of Complex Disease, Department of Ophthalmology, University of Pennsylvania, Philadelphia, PA, USA

² Department of Ophthalmology, University of Pennsylvania, Philadelphia, PA, USA

³ Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

⁴ Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA, USA

⁵ Epigenetics Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, PA, USA

⁶ Center for Preventive Ophthalmology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

⁷ Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA.

⁸ Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

*Corresponding Author: Joan O'Brien (joan.obrien@penncmedicine.upenn.edu)

Abstract

The contribution of common tandem repeats (TR) variants to common, complex disease remains unknown, especially in populations historically underrepresented in genetic research. We identified common TR variants associated with risk of primary open-angle glaucoma (POAG) in individuals of African ancestry. The POAG-associated TR variants were predominantly found at Alu poly(A) tail elements, regions, retinal development enhancers, and harbor binding sites of a POAG-associated transcription factor, LMX1B, suggesting a convergent mechanism of how common TR variation arises and contributes to POAG pathophysiology.

Main

Glaucoma is the leading cause of irreversible blindness worldwide¹. Primary open-angle glaucoma (POAG), the most common form of the disease²⁻⁴, disproportionately affects populations of African ancestry³ with more severe disease outcomes⁵. Since early stages of POAG are asymptomatic⁶, with approximately half of patients unaware they have the disease⁷, tools that identify individuals at risk of developing POAG are needed to enable earlier intervention to prevent disease progression and preserve vision.

Several large genome-wide association studies (GWAS) in multi-ancestry populations⁸⁻¹⁰ and individuals of African ancestry¹¹ have identified POAG risk-associated loci. Primarily, studies followed up with these risk-associated SNPs by implicating nearby genes and the biological pathways in which these genes act^{8,10,11}. However, genes constitute a minority of the genome, and other genetic elements, such as tandem repeats (TRs), contribute to diversity across human genomes. A recent GWAS found an association between a variable tandem number repeat (VNTR) in *TMCO1* and POAG in individuals of European ancestry¹². Here, we investigated whether common TR variants associate with POAG risk in individuals of African ancestry.

First, we leveraged four large-scale GWAS conducted in African ancestry or multi-ancestry populations: (1) POAG mega-analysis in three cohorts of African ancestry individuals, including the Primary Open-Angle African American Glaucoma Genetics (POAAGG) study¹¹; (2) multi-ancestry POAG GWAS¹⁰; (3) multi-ancestry POAG GWAS across 15 biobanks globally⁸; and (4) a multi-ancestry POAG GWAS with the Department of Veterans Affairs Million Veteran Program (MVP) data⁹ (**Fig 1a**). With the compiled genome-wide significant SNPs, we identified SNPs in high linkage disequilibrium with the POAG-associated SNPs in individuals of African ancestry. We then leveraged a reference panel of phased TRs and SNPs¹³ built with whole genome sequencing (WGS) data from individuals of African ancestry (**Methods**). We filtered for TRs with unit repeat length greater than two, as reported disease-associated TRs have predominantly been repeat unit lengths of three or greater¹⁴, yielding 17 potential TRs (**Table 1**).

To test if any of the nominated TRs were associated with POAG risk, we conducted WGS on 99 individuals of African ancestry with POAG in the POAAGG study¹⁵ (**Fig 1b**). We found that within the POAAGG WGS dataset, 16 out of 17 of the nominated TRs were polymorphic (**Fig 1c, Suppl. Fig. 1**), with 11 being dimorphic, and 5 with greater than two alleles in our study population (three allele lengths: 3 TRs; four allele lengths: 1 TR; and five allele lengths: 1 TR). When compared to a reference dataset of TR from an ancestry-matched population¹³ (**Fig 1b, Methods**), we found that 3 of the 17 TRs (chr1:219735425-219735441 AAAC; chr10:58547670-58547681 ATGT; chr14:31626151-31626171 TTATT, hg38) had a differential proportion of allele lengths in the POAG cases relative to the ancestry-matched population comparison (FDR = 5%) (**Fig 1d, Suppl. Fig. 2a**).

Since an intronic VNTR in *TMCO1* has recently been associated with POAG risk in individuals of European ancestry¹², we compared the frequency of expanded intronic *TMCO1* TR in POAG cases to the frequency in individuals of African ancestry in the 1000 Genomes Project^{12,16}. We found that the expanded *TMCO1* TR had a frequency of 11.6% in the POAG cases compared to 8.3% in the population reference (**Suppl. Fig. 2b**). Together, our data contributes three new common TR variants associated with POAG risk to the previously identified *TMCO1* VNTR variant.

Of the four POAG-associated TR variants, three of the four TRs identified fall into the intron of protein coding genes: the ATGT TR in *BICC1*, the TTATT TR in *NUBPL*, the 28 bp VNTR in *TMCO1* (**Fig. 2a**). To assess whether the protein products of these genes are implicated in contributing to POAG risk, we accessed UK Biobank gene-trait association studies based on whole exome sequencing (WES) data¹⁷. We assessed *BICC1*, *NUBPL* and *TMCO1* for associations with POAG or elevated intraocular pressure (IOP) (**Fig. 2b**) and did not find evidence of association with either POAG or elevated IOP.

Rare, penetrant TR variants—especially intronic TRs with AT rich sequences—that result in human disorders often localize at the poly(A) tail of *Alu* elements^{18,19}. We assessed whether the common, POAG-risk associated TR variants share characteristics with rare, penetrant TR variants. Among the TRs associated with POAG, we found that all but one lie within or adjacent to an *Alu* element poly(A) tail (**Fig. 2c**). The AAAC TR and the *TMCO1* TR localize to the poly(A) tail of an *AluJb* element. We found the *NUBPL* TR within the poly(A) tail of an *AluSx1* element. The exception, the *BICC1* TR, lays within a LINE-1 repeat. Similar to rare, penetrant variants^{18,19}, POAG-risk associated TR variants predominantly localize to poly(A) tails of *Alu* elements.

Rare, penetrant TR variants can modify disease phenotypes through influence on gene regulation^{18,20-22}, including serving as sites of transcription factor (TF) binding^{18,23}. To investigate whether the TRs could be TF binding sites of relevance to POAG, we considered TFs with human genetic evidence of association with POAG. Manual curation of the primary literature identified ten genome-wide significant loci where the POAG-associated SNPs are either in a gene or the nearest gene encodes for a TF (**Suppl. Fig. 3**). We utilized the UK Biobank WES data¹⁷ to assess which of these TFs associated with POAG risk and found *LMX1B* and *SIX6* variants associated with greater risk of POAG (**Suppl. Fig. 3**). When we assessed the four POAG-associated TRs, we found predicted binding sites for *LMX1B* within the TTATT TR in *NUBPL* and in *TMCO1* VNTR, but not within the other two TRs (**Fig. 2d**).

Next, we considered to assess the regulatory potential of these TRs, we analyzed single cell epigenomic data from developing human retinal organoids²⁴, specifically H3K27ac, a histone modification that marks enhancers. We found that H3K27ac marks or borders all four TRs across the developmental stages from diencephalon to retinal ganglion cells, with the most pronounced co-localization in retinal progenitor cells (**Fig.2e**).

Together, these data show TRs with POAG-associated risk variants tend to be intronic TRs with binding sites for *LMX1B*, the POAG-associated TF, located at Alu element poly(A) tails with evidence of enhancer activity during retinal development (**Fig. 2f**).

In conclusion, we report TR variants associated with POAG in individuals of African ancestry. Using epigenomic and genomic analyses, we nominated the POAG-associated TRs as putative enhancers in the development of retina and binding sites of a POAG-associated TF, *LMX1B*. Additionally, POAG-associated TR expansions lie within gene introns, specifically adjacent to or within Alu element poly(A) tails. Together, these data inspire a model wherein Alu element poly(A) tails predispose for TR variation as a general hotspot of TR variants. Consequently, TR variants modulate regulatory activity via TF binding to contribute to the pathophysiology of POAG.

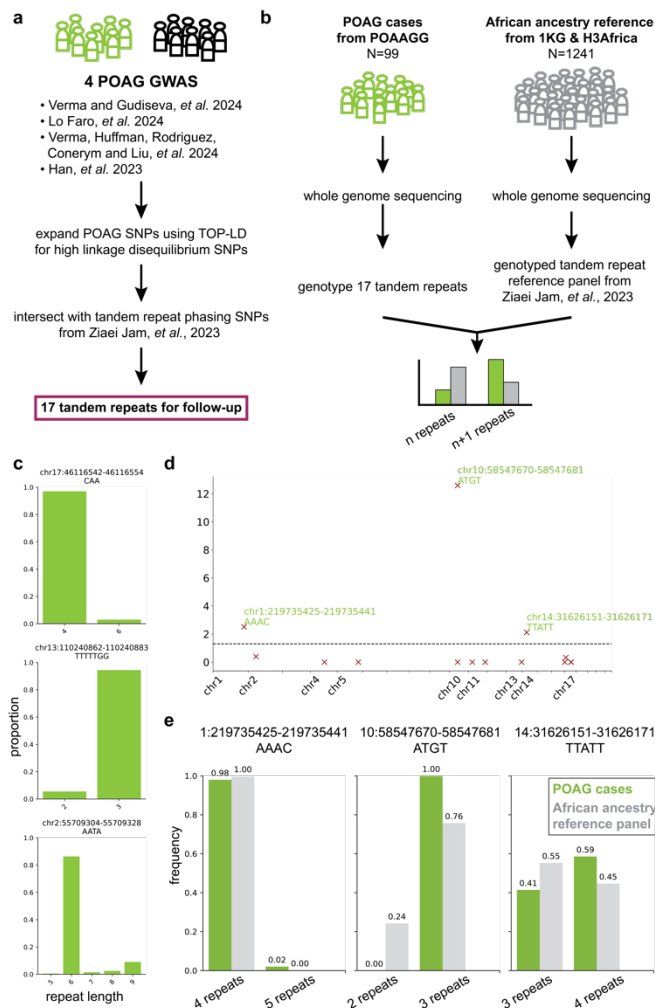


Fig 1. Tandem repeat variants associated with POAG risk in individuals of African ancestry

a, Schematic summarizing approach to nominate tandem repeats, starting with 4 large GWAS conducted in populations of multiple ancestries⁸⁻¹⁰ or individuals of African descent¹¹, identifying SNPs in high linkage disequilibrium, and intersecting with SNPs that phase tandem repeats in individuals of African ancestry¹³. **b**, Schematic of approach to identify tandem repeat variants associated with POAG risk in individuals of African ancestry. **c**, Three tandem repeat length distributions from WGS of POAG cases in the Primary Open-Angle African American Glaucoma Genetics (POAAGG) study. **d**, Manhattan plot of 17 nominated TRs from **a** and their association with POAG in individuals of African ancestry. **e**, Frequency of TR lengths in POAG cases from the POAAGG study and the African ancestry reference panel¹³ for the 3 significant TRs associated with POAG identified in **d**.

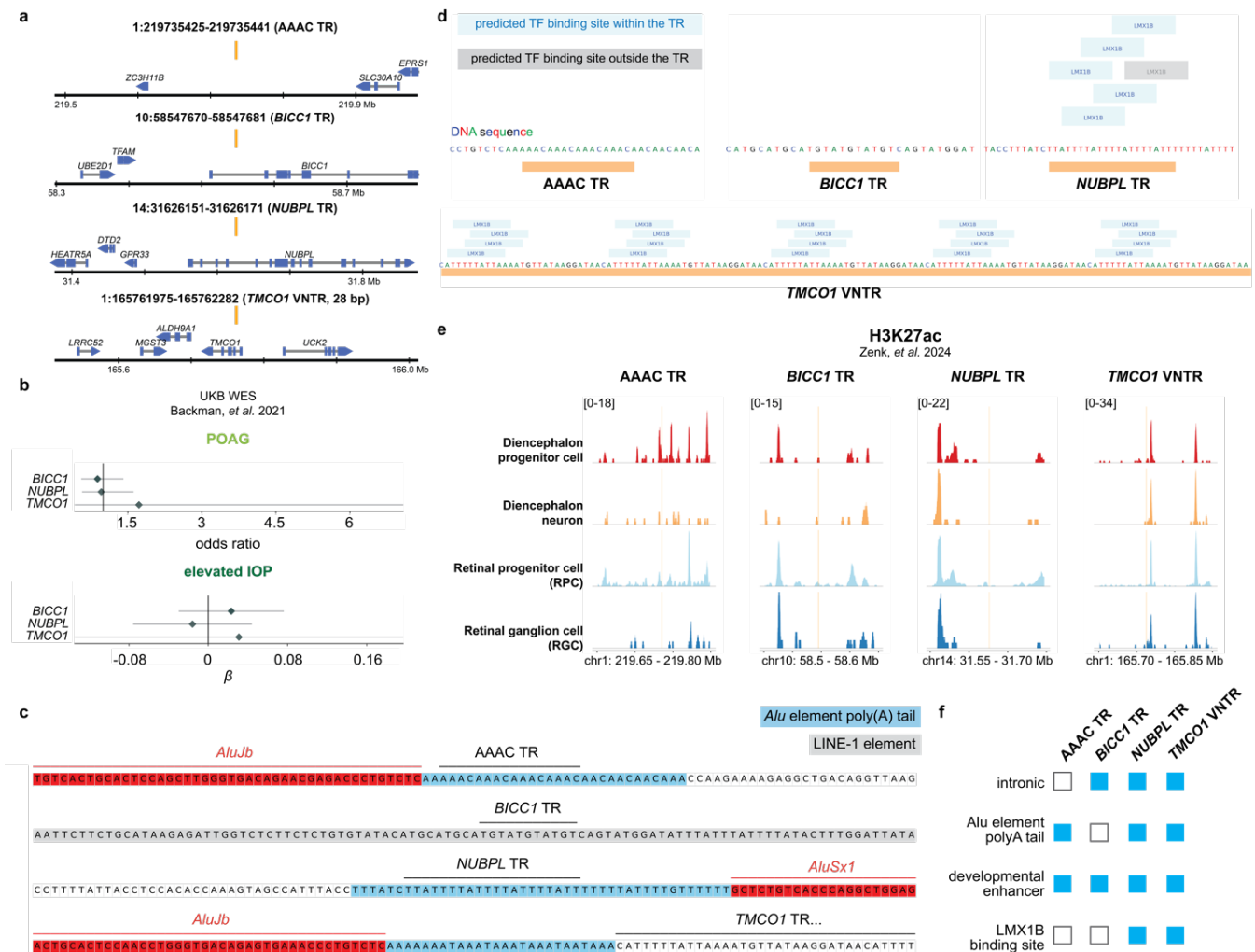
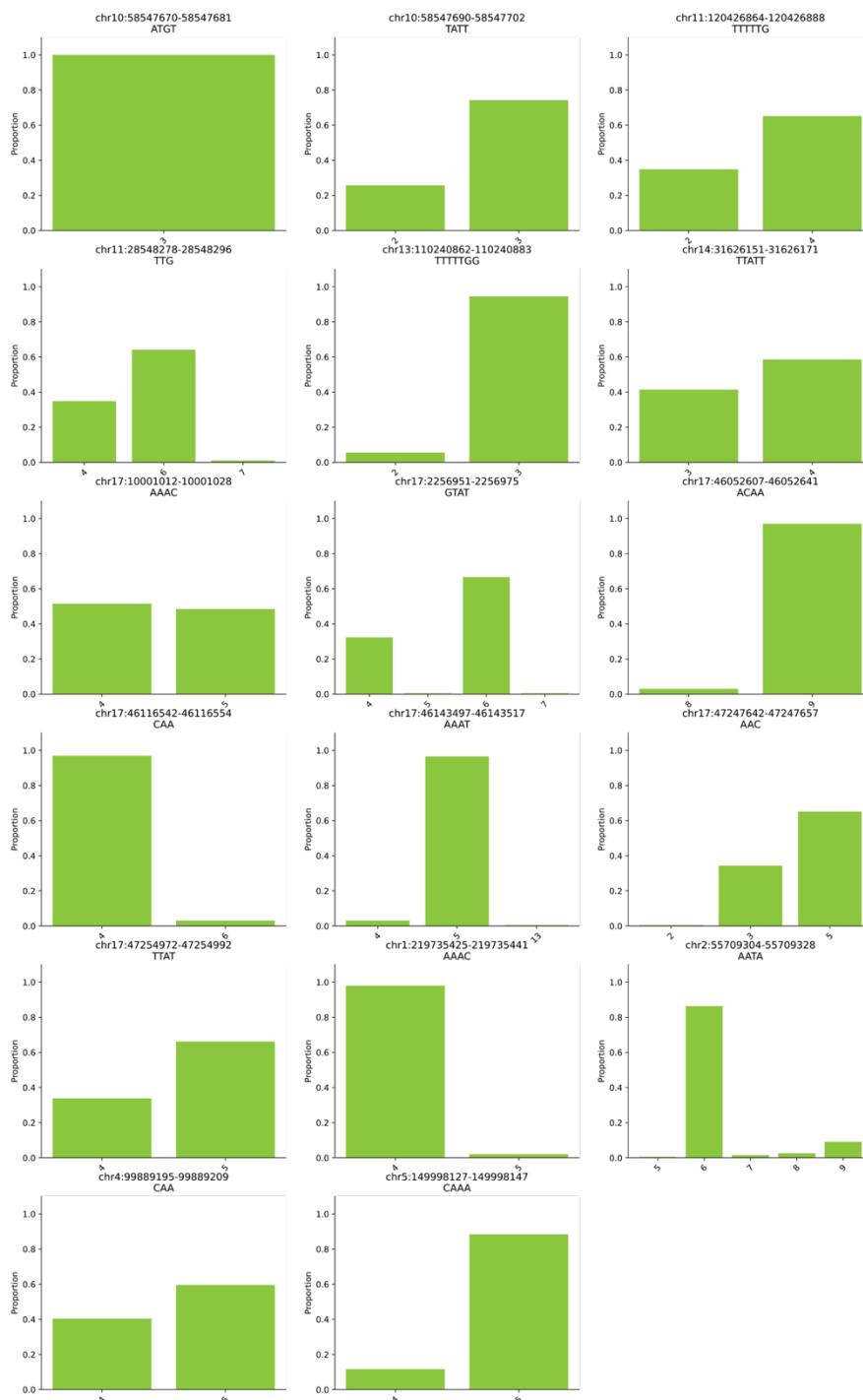
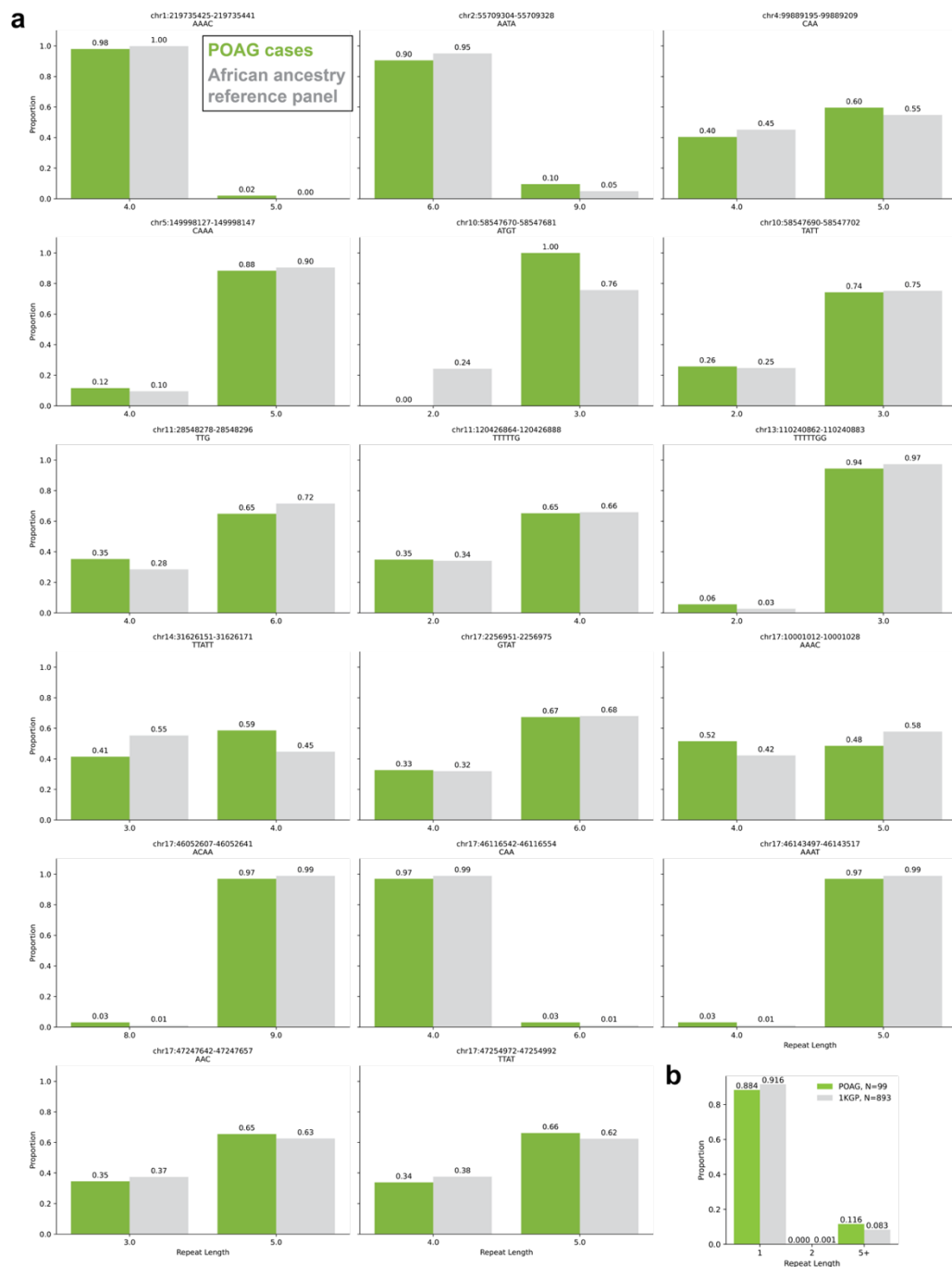


Fig 2. POAG risk-associated tandem repeats harbor binding sites of POAG-associated transcription factors and localize at developmental enhancers at poly(A) tails of *Alu* elements

a, Genomic location of the three TRs associated with POAG in individuals of African descent and the VNTR associated with POAG in individuals of European descent from ref.12¹². **b**, Association of aggregates pLOF and deleterious missense variants with POAG and elevated IOP from ref.17¹⁷. **c**, Genomic context of the four POAG-associated TRs with repetitive elements labeled. (*Alu* elements in red, *Alu* element poly(A) tails in blue, and LINE element in grey.) **d**, Predicted binding sites at the POAG-associated TRs of the POAG-associated TFs, *LMX1B* and *SIX6*. (Binding sites within the TR in blue, and binding sites not contained by the TR in grey.) **e**, H3K27ac single nucleus CUT&Tag signal across retinal development from human retinal organoids, pseudo-bulked by cell type²⁴. **f**, Summary of shared characteristics of the four POAG-associated TRs.

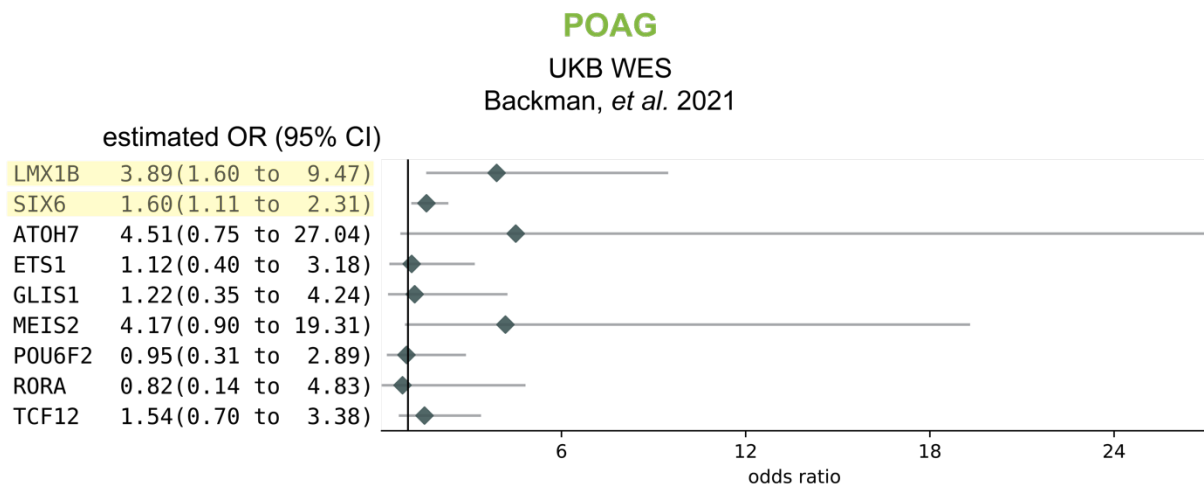


Supplementary Fig. 1 Distribution of tandem repeat lengths in POAG cases
Tandem repeat length distributions from WGS of POAG cases in the POAAGG study for the 17 nominated TRs in **Fig. 1a**



Supplementary Fig. 2 Comparison of tandem repeat lengths in POAG cases and African ancestry reference panel

a, Tandem repeat length distributions for the two most common alleles from WGS of POAG cases in the POAAGG study for the 17 nominated TRs compared to the distribution found in an African ancestry reference panel¹³. **b**, Length distributions for a VNTR associated with POAG in individuals of European descent identified in ref.12¹² in the POAG cases and 1KGP individuals of African ancestry.



Supplementary Fig. 3 Association of transcription factors with POAG

Association of aggregate pLOF and deleterious missense variants with POAG and elevated IOP from ref.17¹⁷ for transcription factors with evidence of association with POAG from human genetic studies (*LMX1B*^{25,26}, *SIX6*²⁷, *ATOH7*²⁸, *ETS1*²⁹, *GLIS1*²⁹, *MEIS2*²⁵, *POU6F2*²⁹, *RORA*³⁰ and *TCF12*²⁹). Estimated odds ratio and 95% confidence interval (CI) are included.

Methods

Identification of TRs associated with POAG

We compiled SNPs associated with POAG from four studies⁸⁻¹¹. Then, we used TOP-LD³¹ to identify SNPs in high linkage disequilibrium with the POAG-associated SNPs with the following settings: African ancestry; $R^2 > 0.9$; MAF > 0 . Next, we intersected the expanded list of SNPs a SNP-TR reference panel⁵¹ to nominate TRs for further study.

POAAGG study population

The POAAGG study population includes self-identified African ancestry subjects (Black, Afro-Caribbean, or African American), aged 35 years or older, recruited from the Philadelphia area. Glaucoma specialists or ophthalmologists classified subjects as cases, controls, or suspects based on previously published criteria¹⁵. All subjects signed an informed consent form under IRB-approved protocols. The University of Pennsylvania Institutional Review Board approved the study, which adheres to the tenets of the Declaration of Helsinki.

Whole genome sequencing

We selected 99 cases from the POAAGG study for WGS. Inclusion criteria for selection included active enrollment at a University of Pennsylvania site (i.e. at least 3 visits over a period of 2 years), without any confounding ocular disease. Of these cases, individuals with a visual acuity of 20/200 or worse, mean deviation on visual field testing of -14dB or worse, and/or average retinal nerve fiber layer thickness less than 60 μm were selected. A manual chart review was performed to ensure that POAG was the primary cause of vision loss. We collected whole blood and isolated DNA. PCR-free WGS was performed at the New York Genome Center to generate 30x coverage whole-genome sequences.

Risk-associated tandem repeat variant identification

To search for tandem repeat variants associated with POAG risk in African ancestry individuals, we used ExpansionHunter³² to genotype the 99 POAG cases with WGS because systematic assessment of tandem repeat genotyping tools demonstrated that ExpansionHunter had the highest accuracy³³. We genotyped the 17 nominated TRs (see **Identification of TRs associated with POAG**). We then compared the proportions of the two most common allele lengths in POAG cases to the proportions in individuals of African ancestry¹³ that used WGS data from over 1200 individuals of African ancestry¹³ from the 1000 Genomes Project¹⁶ and H3Africa³⁴ cohorts. We used a chi-square test with Bonferroni correction and an FDR of 5%. For the *TMCO1* VNTR, we genotyped the 99 POAG cases using code from ref. 12¹².

UKB WES for associations with POAG and elevated IOP

We accessed summary statistics for the whole-exome sequencing data from the UK Biobank. We tested associations by using the odds ratio of POAG and/or elevated IOP with burden tests that combined predicted loss-of-function and predicted deleterious missense mutation with minor allele frequency 1%.

Alu element poly(A) tail annotation

We annotated Alu elements and their poly(A) tails based on RepeatMasker³⁵ annotations accessed on the UCSC Genome Browser³⁶.

Predicted transcription factor binding sites

We mined JASPAR³⁷ and HOCOMOCO³⁸ for DNA binding motifs for the POAG-associated TFs, *LMX1B* and *SIX6*, and then used FIMO³⁹ to search for predicted binding sites at the two TRs with POAG-associated variants.

H3K27ac data from human developing retinal organoids

We accessed H3K27ac single-nucleus CUT&Tag data from ref.24²⁴ from 12 weeks old human retinal organoids for four cell types: diencephalon progenitor cells, diencephalon neurons, retinal progenitor cells, and retinal ganglion cells.

Data availability

The WGS of the POAG cases in the POAAGG study are available from the corresponding author on reasonable request and with permission of the University of Pennsylvania Institutional Review Board. All other data were publicly available.

Code availability

We used previously published tools, as indicated in the **Methods** to perform the analyses. No custom code was developed or used.

Acknowledgements

We thank members of the O'Brien lab for helpful discussions.

Funding

The Primary Open-Angle African American Glaucoma Genetics (POAAGG) study is supported by the National Eye Institute, Bethesda, Maryland (grant #1R01EY023557) and Vision Research Core Grant (P30 EY001583). Funds also come from the F.M. Kirby Foundation, Research to Prevent Blindness, The UPenn Hospital Board of Women Visitors, and The Paul and Evanina Bell Mackall Foundation Trust. K. Pham is a recipient of a F30 Ruth L. Kirschstein National Research Service Award Individual Predoctoral Fellowship (F30HD114405). K. Pham and J. Phillips-Cremens acknowledge funding from NIH NIMH (1R01MH120269; J.E.P.-C.) and CZI Neurodegenerative Disease Pairs Award (2020-221479-5022; DAF2022-250430; J.E.P.-C.). Additional support also came from the Ophthalmology Department at the Perelman School of Medicine and the VA Hospital in Philadelphia, PA. The funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

1. Flaxman, S.R. *et al.* Global causes of blindness and distance vision impairment 1990-2020: a systematic review and meta-analysis. *Lancet Glob Health* **5**, e1221-e1234 (2017).
2. Quigley, H.A. & Broman, A.T. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol* **90**, 262-7 (2006).
3. Bourne, R.R. *et al.* Number of People Blind or Visually Impaired by Glaucoma Worldwide and in World Regions 1990 - 2010: A Meta-Analysis. *PLoS One* **11**, e0162229 (2016).
4. Tham, Y.C. *et al.* Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology* **121**, 2081-90 (2014).
5. Munoz, B. *et al.* Causes of blindness and visual impairment in a population of older Americans: The Salisbury Eye Evaluation Study. *Arch Ophthalmol* **118**, 819-25 (2000).
6. Kastner, A. & King, A.J. Advanced glaucoma at diagnosis: current perspectives. *Eye (Lond)* **34**, 116-128 (2020).
7. Mitchell, P., Smith, W., Attebo, K. & Healey, P.R. Prevalence of open-angle glaucoma in Australia. The Blue Mountains Eye Study. *Ophthalmology* **103**, 1661-9 (1996).
8. Lo Faro, V. *et al.* Novel ancestry-specific primary open-angle glaucoma loci and shared biology with vascular mechanisms and cell proliferation. *Cell Rep Med* **5**, 101430 (2024).
9. Verma, A. *et al.* Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program. *Science* **385**, eadj1182 (2024).
10. Han, X. *et al.* Large-scale multitrait genome-wide association analyses identify hundreds of glaucoma risk loci. *Nat Genet* **55**, 1116-1125 (2023).
11. Verma, S.S. *et al.* A multi-cohort genome-wide association study in African ancestry individuals reveals risk loci for primary open-angle glaucoma. *Cell* **187**, 464-480 e10 (2024).
12. Mukamel, R.E. *et al.* Repeat polymorphisms underlie top genetic risk loci for glaucoma and colorectal cancer. *Cell* **186**, 3659-3673 e23 (2023).
13. Ziaei Jam, H. *et al.* A deep population reference panel of tandem repeat variation. *Nat Commun* **14**, 6711 (2023).
14. Gall-Duncan, T., Sato, N., Yuen, R.K.C. & Pearson, C.E. Advancing genomic technologies and clinical awareness accelerates discovery of disease-associated tandem repeat sequences. *Genome Res* **32**, 1-27 (2022).
15. Charlson, E.S. *et al.* The primary open-angle african american glaucoma genetics study: baseline demographics. *Ophthalmology* **122**, 711-20 (2015).
16. Byrska-Bishop, M. *et al.* High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell* **185**, 3426-3440 e19 (2022).
17. Backman, J.D. *et al.* Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628-634 (2021).

18. Grasberger, H. *et al.* STR mutations on chromosome 15q cause thyrotropin resistance by activating a primate-specific enhancer of MIR7-2/MIR1179. *Nat Genet* **56**, 877-888 (2024).
19. Ishiura, H. & Tsuji, S. Advances in repeat expansion diseases and a new concept of repeat motif-phenotype correlation. *Curr Opin Genet Dev* **65**, 176-185 (2020).
20. Narumi, S. *et al.* Functional variants in a TTTG microsatellite on 15q26.1 cause familial nonautoimmune thyroid abnormalities. *Nat Genet* **56**, 869-876 (2024).
21. Law, M.J. *et al.* ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell* **143**, 367-78 (2010).
22. Taylor, A.K. *et al.* Tissue heterogeneity of the FMR1 mutation in a high-functioning male with fragile X syndrome. *Am J Med Genet* **84**, 233-9 (1999).
23. Horton, C.A. *et al.* Short tandem repeats bind transcription factors to tune eukaryotic gene expression. *Science* **381**, eadd1250 (2023).
24. Zenk, F. *et al.* Single-cell epigenomic reconstruction of developmental trajectories from pluripotency in human neural organoid systems. *Nat Neurosci* **27**, 1376-1386 (2024).
25. Shiga, Y. *et al.* Genome-wide association study identifies seven novel susceptibility loci for primary open-angle glaucoma. *Hum Mol Genet* **27**, 1486-1496 (2018).
26. Park, S. *et al.* Genetic risk for primary open-angle glaucoma determined by LMX1B haplotypes. *Invest Ophthalmol Vis Sci* **50**, 1522-30 (2009).
27. Wiggs, J.L. *et al.* Common variants at 9p21 and 8q22 are associated with increased susceptibility to optic nerve degeneration in glaucoma. *PLoS Genet* **8**, e1002654 (2012).
28. Ramdas, W.D. *et al.* Common genetic variants associated with open-angle glaucoma. *Hum Mol Genet* **20**, 2464-71 (2011).
29. Gharahkhani, P. *et al.* Genome-wide meta-analysis identifies 127 open-angle glaucoma loci with consistent effect across ancestries. *Nat Commun* **12**, 1258 (2021).
30. Choquet, H. *et al.* A multiethnic genome-wide association study of primary open-angle glaucoma identifies novel risk loci. *Nat Commun* **9**, 2278 (2018).
31. Huang, L. *et al.* TOP-LD: A tool to explore linkage disequilibrium with TOPMed whole-genome sequence data. *Am J Hum Genet* **109**, 1175-1181 (2022).
32. Dolzhenko, E. *et al.* Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Res* **27**, 1895-1903 (2017).
33. Weisburd, B., Tiao, G. & Rehm, H.L. Insights from a genome-wide truth set of tandem repeat variation. *bioRxiv* (2023).
34. Choudhury, A. *et al.* High-depth African genomes inform human migration and health. *Nature* **586**, 741-748 (2020).
35. Smit, A., Hubley, R. & Green, P. RepeatMasker.
36. Nassar, L.R. *et al.* The UCSC Genome Browser database: 2023 update. *Nucleic Acids Res* **51**, D1188-D1195 (2023).

37. Rauluseviciute, I. *et al.* JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **52**, D174-D182 (2024).
38. Vorontsov, I.E. *et al.* HOCOMOCO in 2024: a rebuild of the curated collection of binding models for human and mouse transcription factors. *Nucleic Acids Res* **52**, D154-D163 (2024).
39. Grant, C.E., Bailey, T.L. & Noble, W.S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-8 (2011).