TITLE

The co-occurrence of genetic variants in the *TYR* and *OCA2* genes confers susceptibility to albinism

AUTHOR LIST & AFFILIATIONS

David J. Green[1], Vincent Michaud[2,3], Eulalie Lasseaux[2], Claudio Plaisant[2], UK Biobank Eye and Vision Consortium, Tomas Fitzgerald[4], Ewan Birney[4], Graeme C. Black[1,5,#], Benoît Arveiler[2,3,#], Panagiotis I. Sergouniotis[1,4,5,6]

[1] Division of Evolution, Infection and Genomics, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK.

[2] Department of Medical Genetics, University Hospital of Bordeaux, Bordeaux, France.

[3] INSERM U1211, Rare Diseases, Genetics and Metabolism, University of Bordeaux, Bordeaux, France.

[4] European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Cambridge, UK.

[5] Manchester Centre for Genomic Medicine, Saint Mary's Hospital, Manchester University NHS Foundation Trust, Manchester, UK.

[6] Manchester Royal Eye Hospital, Manchester University NHS Foundation Trust, Manchester, UK.

# These authors contributed equally: Graeme C. Black, Benoît Arveiler

Correspondence to

Panagiotis I. Sergouniotis (panagiotis.sergouniotis@manchester.ac.uk).

ABSTRACT

Although rare genetic conditions are mostly caused by DNA sequence alterations that functionally disrupt individual genes, large-scale studies using genome sequencing have started to unmask additional complexity. Understanding how combinations of variants in different genes shape human phenotypes is expected to provide important insights into the clinical and genetic heterogeneity of rare disorders. We used albinism, an archetypal rare condition associated with hypopigmentation, as an exemplar for the study of genetic interactions. We analysed data from the Genomics England 100,000 Genomes Project alongside a cohort of 1,121 individuals with albinism, and investigated the effect of dual heterozygosity for the combination of two established albinism-related variants: *TYR*:c.1205G>A (p.Arg402Gln) [rs1126809] and *OCA2*:c.1327G>A (p.Val443Ile) [rs74653330]. As each of these changes alone is insufficient to cause disease when present in the heterozygous state, we sought evidence of synergistic effects. We found that, when both variants are present, the probability of receiving a diagnosis of albinism is significantly increased (odds ratio 7.9; 95% confidence interval 4.4 – 13.2; p-value 4 x 10^{-9}). Further analyses in an independent cohort, the UK Biobank, supported this finding and highlighted that heterozygosity for the *TYR*:c.1205G>A and *OCA2*:c.1327G>A variant combination is associated with statistically significant alterations in visual acuity and central retinal thickness (traits that are considered albinism endophenotypes). The approach discussed in this report opens up new avenues for the investigation of oligogenic patterns in apparently Mendelian disorders.

KEYWORDS: epistasis, digenic, oligogenic, bi-locus, genetic interaction

MAIN TEXT

INTRODUCTION

An individual affected by a condition that appears to segregate as a Mendelian trait, would be expected to carry large-effect genetic variation in a single gene/locus. However, for a significant proportion of affected probands, single-gene-centric genomic analyses fail to identify a clear genetic diagnosis.[1-6] Although this diagnostic and knowledge gap is, to a degree, due to imperfect phenotyping, sequencing or annotation, other factors may be contributing. There is for example growing evidence supporting the hypothesis that many apparently Mendelian phenotypes are caused by the interaction of multiple variants in more than one locus.[7-10] Uncovering such genetic interactions has the potential to provide insights not only into disease mechanisms and diagnostics but also into phenomena like incomplete penetrance and variable expressivity.[11-13]

There are many patterns through which genetic changes in more than one gene can synergistically shape a phenotype. A simple scenario, would involve two heterozygous variants in two distinct loci cooperating to mediate disease (when each individual variant in isolation fails to explain the clinical presentation). Only a small number of cases with this specific form of digenic inheritance have been described[14-16], the majority detected through qualitative approaches involving variant segregation analyses in a small number of families. In these cases, interaction-based models have generally been used to provide a post hoc explanation of the observed phenomena. By contrast, we describe a quantitative, case-control approach that allows obtaining statistical evidence of digenic patterns. We focused on specific genotypes in the *TYR* and *OCA2* genes; biallelic variants in each of these two genes

are associated with albinism, a clinically and genetically heterogeneous group of conditions characterised by reduced levels of melanin pigment. Key features of albinism include visual abnormalities and under-development of the central retina (i.e. the fovea); skin and/or hair hypopigmentation may also be present.[17] Building on recent work[18-21], we aimed to advance our understanding of the genetic complexity of this rare disorder.

RESULTS & DISCUSSION

A cohort of 1,016 people with albinism underwent testing of ≤19 albinism-related genes; these individuals were not known to be related and had predominantly European ancestries (Supplementary Data 1). A further 105 probands with albinism were identified in the Genomics England 100,000 Genomes Project dataset.[22] A 'control' cohort of 29,451 unrelated individuals that had no recorded diagnosis or features of albinism was also identified in this resource (Fig.1, Supplementary Table 1, Methods).

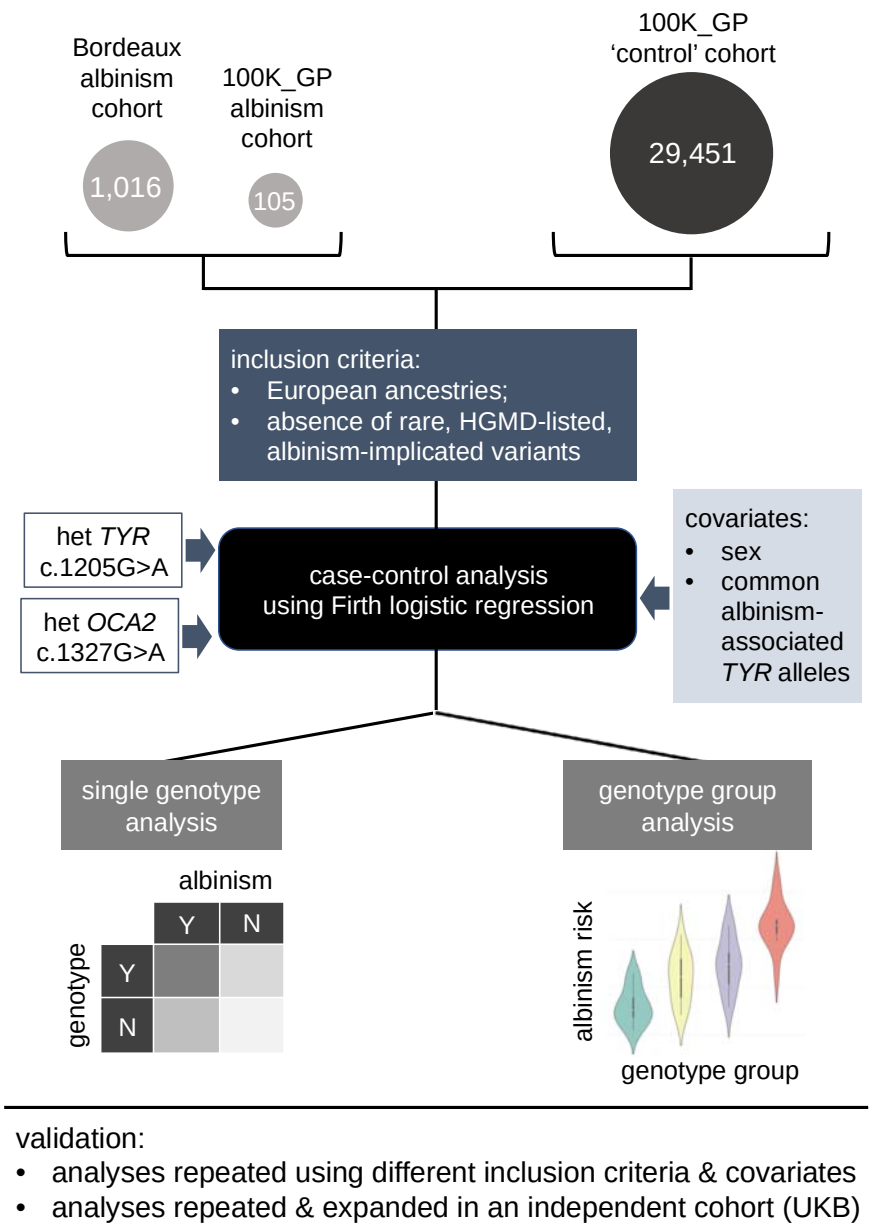


Figure 1. Outline of the case-control study design.

A case-control analysis was performed to gain insights into the contribution of genotypes involving heterozygosity for the *TYR*:c.1205G>A (p.Arg402Gln) [rs1126809] and *OCA2*:c.1327G>A (p.Val443Ile) [rs74653330] variant combination in albinism. These two missense changes were selected as they are the commonest albinism-related changes in European populations.

The majority of participants in the 'case' cohort (1,016/1,121) were identified through the database of the University Hospital of Bordeaux Molecular Genetics Laboratory, France. All these probands had at least one key ocular feature of albinism, *i.e.* nystagmus or prominent foveal hypoplasia. The remaining 105/1,121 cases were identified through the Genomics England 100,000 Genomes Project dataset and had a diagnosis of albinism or partial/ocular albinism. The 'control' cohort included 29,451 unrelated individuals from the Genomics England 100,000 Genomes Project

dataset, none of whom had a recorded diagnosis of albinism.

The primary analysis involved using logistic regression to study the subsets of the case and control cohorts that (i) had European ancestries and (ii) did not carry any other rare genetic changes that had been labelled as disease-causing (DM) in the Human Gene Mutation Database (HGMD) v2021.2. The following two covariates were used: sex; number of common albinism-associated alleles, *i.e.* relevant DNA sequence alterations in the genomic locations corresponding to *TYR* c.−301C>T [rs4547091] and *TYR* c.575C>A (p.Ser192Tyr) [rs1042602].

To assess the robustness of the findings we performed additional regression analyses in the same cohort using broader inclusion criteria (e.g not just individuals of European ancestries) and incorporating additional covariates. Validation studies in an independent cohort (UK Biobank) were also conducted.

100K_GP, Genomics England 100,000 Genomes Project; HGMD, Human Gene Mutation Database v2021.2; het, heterozygous; UKB, UK Biobank. *TYR* and *OCA2* variant numbering was based on the transcripts with the following identifiers: NM_000372.5/ENST00000263321.6. and NM_000275.3/ENST00000354638.8.

Searching for genetic elements underlying digenic traits may be carried out at the level of genes, genotypes or DNA variants. It is noted that a given pair of DNA variants comprises up to nine pairs of genotypes while a given gene may contain a large number of variants. Each of these search strategies has inherent advantages and pitfalls.[23] We opted for a genotype-based approach that offers a high level of precision and is more likely to provide insights that can be used for clinical genome interpretation.

Our genotype-based analysis focused on two prevalent albinism-related changes, *TYR*:c.1205G>A (p.Arg402Gln) [rs1126809] and *OCA2*:c.1327G>A (p.Val443Ile) [rs74653330]. We hypothesised that when these two missense variants are both present in the heterozygous state ("dual heterozygosity"), their interaction is driving the pathology observed in albinism. Notably, both these changes have been shown to reduce the activity of their respective protein products *in vitro*.[24-28] It is also highlighted that previous studies have pointed to a functional interaction between the TYR and OCA2 proteins: the latter has a role in the pH regulation of the

melanosome which, in turn, can affect the catalytic activity of the former (i.e. tyrosinase, the rate-limiting enzyme for the synthesis of melanin in the melanosomes of retinal pigment epithelia and melanocytes).[17,29] Both *TYR*:c.1205G>A and *OCA2*:c.1327G>A have been implicated in albinism when *in trans* with pathogenic changes in the same gene.[20,30] Also, multiple associations have been recorded for these two variants including with skin/hair pigmentation and skin cancer.[31] It is noted that these two changes were chosen as they are thought to be the two commonest albinism-related variants in populations of European ancestries, a feature that can help mitigate potential issues with statistical power. The minor allele frequencies of the *TYR*:c.1205G>A and *OCA2*:c.1327G>A changes in European populations are 27% and 0.3% respectively (Non-Finnish European subset of the Genome Aggregation Database [gnomAD] v2.1.1).[32]

We used Firth regression analysis[33,34] to study how heterozygosity for the *TYR*:c.1205G>A and *OCA2*:c.1327G>A variant combination affects the probability of receiving a diagnosis of albinism ("risk of albinism"). This logistic regression approach has been designed to handle small, imbalanced datasets (which are common in studies of rare conditions) and allows for adjustment of key covariates (which is not possible in contingency table methods) (Fig.1).

We performed two different regression analyses. The first one involved a direct case-control comparison between those individuals who were heterozygous for the *TYR*:c.1205G>A and *OCA2*:c.1327G>A variant combination and those who were not. The focus of this primary analysis was the subset of participants who had European ancestries and carried no other rare albinism-implicated variants (i.e., low-

frequency genetic changes that had been labelled as disease-causing (DM) in the Human Gene Mutation Database (HGMD) v2021.2)[35]. Sex and the number of common albinism-associated alleles were factored in as covariates. The point estimate for the relevant odds ratio (OR) was 7.9 (the 95% confidence interval was 4.4 – 13.2 and the p-value was 4×10^{-9}). It is noted that this effect size is similar to what has been calculated for established disease-associated genotypes including pathogenic and likely pathogenic variants in *BRCA1*/*BRCA2*[36,37] and a prevalent glaucoma-predisposing variant in *MYOC,* c.1102C>T, (p.Gln368Ter) [rs74315329][38,39].

Our second analysis involved case-control comparisons between different groups of genotypes linked to the *TYR*:c.1205G>A and *OCA2*:c.1327G>A changes. The studied groups were:

- *group A* (reference group): homozygous for *TYR*:c.1205= and *OCA2*:c.1327=
- *group B* (single heterozygote group): heterozygous for either the *TYR*:c.1205G>A and *OCA2*:c.1327= variant combination or for the *TYR*:c.1205= and *OCA2*:c.1327G>A variant combination.
- *group C* (dual heterozygote group): heterozygous for the *TYR*:c.1205G>A and *OCA2*:c.1327G>A variant combination.
- *group D* (homozygous combination group): either homozygous for *TYR*:c.1205G>A and heterozygous for *OCA2*:c.1327G>A; or heterozygous for *TYR*:c.1205G>A and homozygous for *OCA2*:c.1327G>A; or homozygous for both changes.

These groups correspond to different levels of dysfunction of the OCA2 and TYR molecules and the findings of this regression analysis revealed a synergistic potentiation of the changes at the two loci (see Fig.2 and Supplementary Table 2).

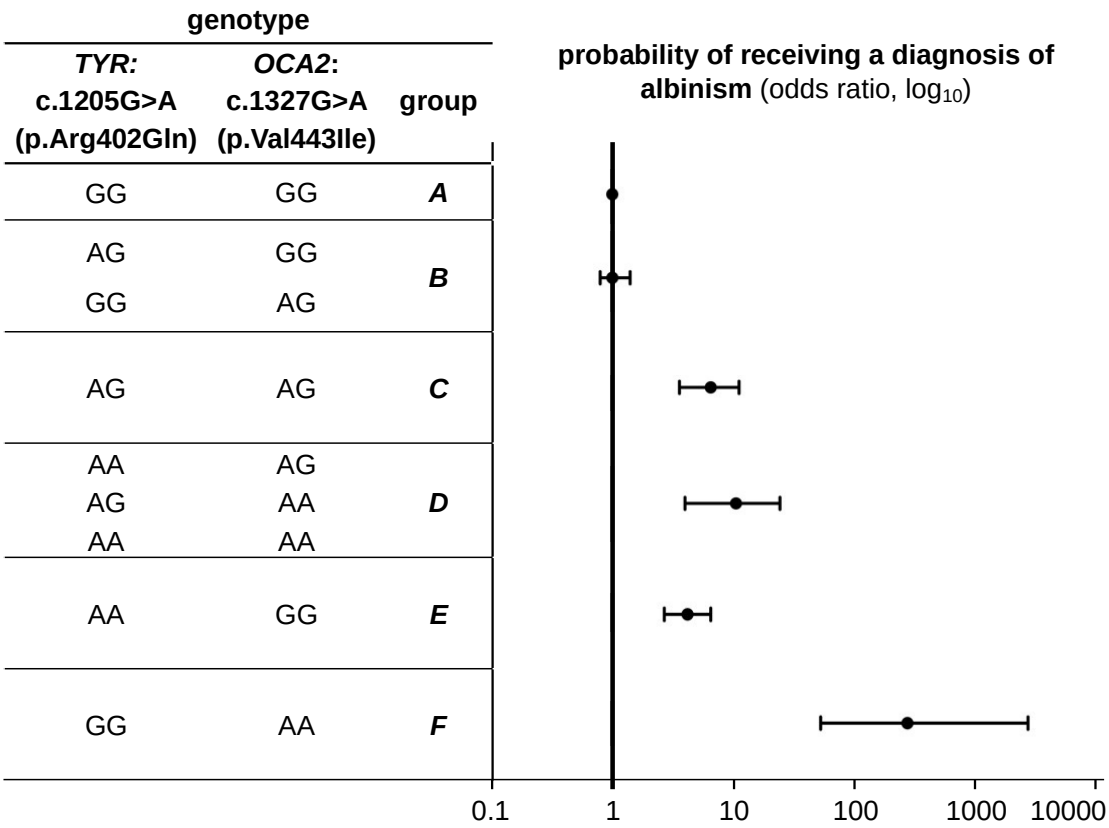


Figure 2. The combination of the *TYR*:c.1205G>A (p.Arg402Gln) and *OCA2*:c.1327G>A (p.Val443Ile) variants in a dual heterozygous state confers susceptibility to albinism.

The probability of receiving a diagnosis of albinism (i.e. the risk of albinism) was calculated for the following genotype groups:

- ***group A*** (reference group): homozygous for *TYR*:c.1205= and *OCA2*:c.1327=
- ***group B*** (single heterozygote group): heterozygous for either the *TYR*:c.1205G>A and *OCA2*:c.1327= variant combination or for the *TYR*:c.1205= and *OCA2*:c.1327G>A variant combination.
- ***group C*** (dual heterozygote group): heterozygous for the *TYR*:c.1205G>A and *OCA2*:c.1327G>A variant combination.
- ***group D*** (homozygous combination group): either homozygous for *TYR*:c.1205G>A and heterozygous for *OCA2*:c.1327G>A; or heterozygous for *TYR*:c.1205G>A and homozygous for *OCA2*:c.1327G>A; or homozygous for both changes.

Findings in individuals who are homozygous for *TYR*:c.1205G>A (***group E***) and individuals who are homozygous for *OCA2*:c.1327G>A (***group F***) are also shown. As the former is a common, “hypomorphic” change, it is expected to have a smaller effect size compared to the latter.

The circle in the middle of each horizontal line (95% confidence interval) represent the point estimate of each odds ratio. A $\log_{10}$ scale with 1 as the reference point is used.

This analysis focused on case and control individuals that (i) had European ancestries and (ii) did not carry any other rare genetic changes that had been labelled as disease-causing (DM) in the Human Gene Mutation Database (HGMD) v2021.2. Two covariates were factored in (see Fig.1). Group A was used as the reference group to which all other groups were compared using Firth regression analysis (i.e. the odds ratio for this group was fixed at 1). Further information including numerical data can be found in Supplementary Table 2.

To increase confidence in these observations, we performed additional analyses in an independent cohort, the UK Biobank.[40] We found that UK Biobank volunteers who were heterozygous for the *TYR*:c.1205G>A and *OCA2*:c.1327G>A variant combination had not only a higher chance of receiving a diagnosis of albinism (OR>3.8, p-value=0.0017) but also had, on average, slightly worse visual acuity and marginally thicker central retina (the respective Kruskal-Wallis p-values were 1×10^{-4} and 9.6×10^{-6}; see Fig.3 and Supplementary Tables 3&4). It is noted that visual acuity and central retinal thickness are quantitative endophenotypes of albinism and these findings suggest that the studied *TYR*/*OCA2* variant combination is both contributing to the genetic architecture of albinism and having a role in the development of the visual system in individuals from the general population.

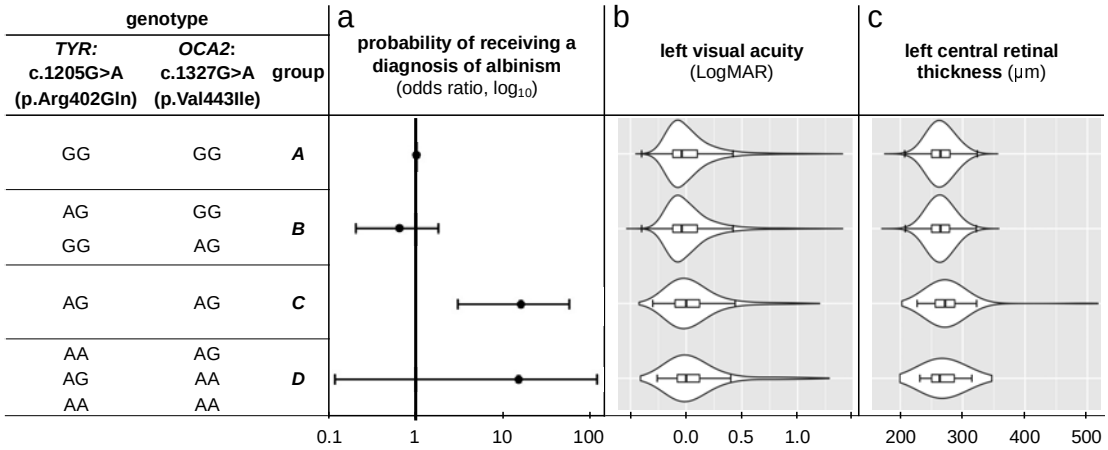


Figure 3. The combination of the *TYR*:c.1205G>A (p.Arg402Gln) and *OCA2*:c.1327G>A (p.Val443Ile) variants in a dual heterozygous state is associated with (a) a higher probability of receiving an albinism diagnosis (i.e. a higher risk of albinism), (b) lower visual acuity and (c) increased central retinal thickness in UK Biobank participants.

The following genotype groups were studied:

- ***group A*** (reference group): homozygous for *TYR*:c.1205= and *OCA2*:c.1327=
- ***group B*** (single heterozygote group): heterozygous for either the *TYR*:c.1205G>A and *OCA2*:c.1327= variant combination or for the *TYR*:c.1205= and *OCA2*:c.1327G>A variant combination.
- ***group C*** (dual heterozygote group): heterozygous for the *TYR*:c.1205G>A and *OCA2*:c.1327G>A variant combination.
- ***group D*** (homozygous combination group): either homozygous for *TYR*:c.1205G>A and heterozygous for *OCA2*:c.1327G>A; or heterozygous for *TYR*:c.1205G>A and homozygous for *OCA2*:c.1327G>A; or homozygous for both changes.

It is noted that individuals with albinism tend to have reduced visual acuity (*i.e.* higher LogMAR value than 0.2) and increased central retinal thickness (due to underdevelopment of the fovea). Further information including numerical data can be found in Supplementary Tables 3&4.

We provide evidence that the *TYR*:c.1205G>A and *OCA2*:c.1327G>A changes appear to have a joint effect when each is encountered in the heterozygous state. However, the interpretation of this variant combination in clinical diagnostic settings should be approached cautiously. Although the detected effect size (odds ratio) is significant, it would be more in keeping with a predisposing[41] rather than a pathogenic[42] genotype. Further, while it is tempting to view our observations as an example of a digenic pattern, this would have reductionist connotations; interactions

are ubiquitous in nature and we expect that increased sample sizes will elucidate longer “chains” of interconnected genetic variation. It is thus plausible that the effect size of the studied genotype will increase when further genetic changes that impact on tyrosinase function or melanosome pH are added to the model. Indeed, observations in the presented data already point to this possibility (Supplementary Data 2).

One potential limitation of this study is our inability to stringently match the albinism cases with the unaffected controls, especially in terms of recent ancestry.[43] Although ancestry was considered, our analysis was imperfect as it was not possible to reliably assign genetic ancestry to most albinism cases. We used a combination of orthogonal approaches to evaluate the robustness and generalisability of our findings. First, we used 35 presumed neutral single-nucleotide variants to calculate the genomic inflation factor lambda (λGC)[44,45]; the λ_{median} was found to be 1.04, in keeping with limited confounding by ancestry (Supplementary Table 5). Subsequently, we performed targeted sub-analyses of the available cohorts; the results of three focused case-control studies supported our key findings and increased confidence in the validity of the detected associations (Supplementary Fig.1 and Supplementary Tables 6-8).

In conclusion, we have shown that dual heterozygosity for a *TYR* and an *OCA2* variant confers susceptibility to albinism. Shedding light into this and other variant interactions is expected to increase our understanding of the pathways and cellular processes affected in albinism. Notably, the concepts discussed herein are likely to be relevant to other rare disorders, and systems-based approaches using biological

modules are expected to narrow the molecular diagnostic gap. Future work will focus on increasing sample sizes, embracing more diverse populations and searching for high-order interactions by considering more than two variants at a time.

METHODS

Cohort characteristics and genotyping

University Hospital of Bordeaux albinism cohort

Individuals with albinism were identified through the database of the University Hospital of Bordeaux Molecular Genetics Laboratory, France. This is a national reference laboratory that has been performing genetic testing for albinism since 2003 and has been receiving samples from individuals predominantly based in France (or French-administered overseas territories).

Information on the dermatological and ophthalmological phenotypes was available for all affected individuals, and each of these cases had at least one of the key ocular features of albinism, *i.e.* infantile nystagmus or absence of a foveal pit (prominent foveal hypoplasia). No pre-screening based on genotype was undertaken other than selecting individuals in whom position c.-301 of *TYR* was sequenced. Only individuals who were not knowingly related were included.

Genetic testing, bioinformatic analyses, and variant interpretation were performed as previously described.[20,21] Briefly, most participants had gene-panel testing of ≤19 genes associated with albinism (*TYR*, *OCA2*, *TYRP1*, *SLC45A2*, *SLC24A5*, *C10ORF11*, *GPR143*, *HPS 1* to 10, *LYST*, *SLC38A8*) using IonTorrent platforms. High-resolution array-CGH (comparative genomic hybridization) was used to detect copy number variants in these genes. All genetic changes of interest were confirmed with an alternative method (e.g. Sanger sequencing or quantitative PCR). Clinical interpretation of variants was performed using criteria consistent with the 2015 American College of Medical Genetics and Genomics (ACMG) best practice guidelines.[42] It is noted that genetic findings in this cohort have been partly reported in previous publications from our group.[18,21]

Due to the limited number of genes screened in this cohort, it was not possible to reliably assess genetic ancestry and to objectively assign individuals to ancestry groups. Attempting to mitigate this, we processed available data on self-identified ethnicity that were collected through questionnaires. Responses were inspected and stratification into five broad continental groups (European, African, Admixed American, East Asian, South Asian) was performed.

Genomics England 100,000 Genomes Project cohort

Clinical and genomic data from the Genomics England 100,000 Genomes Project were accessed through a secure Research Environment that is available to registered users. This dataset was collected as part of a national genome sequencing initiative.[46] Enrolment was coordinated by Genomics England Limited, and participants were recruited mainly at National Health Service (NHS) Hospitals in the UK.[22] Clinical information was recorded in Human Phenotype Ontology (HPO)[47] terms and International Classification of Diseases (ICD) codes. Genome sequencing was performed in DNA samples from 78,195 individuals using Illumina HiSeq X systems (150 base-pair paired-end format). Reads were aligned using the iSAAC Aligner v03.16.02.19 and small variants were called using Starling v2.4.7.[48] Aggregation of single-sample gVCFs was performed using the Illumina software gVCF genotyper v2019.02.29; normalisation/decomposition was implemented by vt version 0.57721[49]. The multi-sample VCF was then split into 1,371 roughly equal chunks to allow faster processing and the loci of interest were queried using bcftools v1.9[50] (see https://research-help.genomicsengland.co.uk/display/GERE/ for further information). Only variants that passed all provided site quality control criteria were processed. In addition, we filtered out genotypes with: genotype score <20; read depth <10; allele balance <0.2 and >0.8 for heterozygotes; allele balance >0.1 or <0.9 for homozygotes (reference and alternate, respectively). Genomic annotation was performed using Ensembl VEP[51]; one additional annotation was included: presence of a variant in HGMD v2021.2[35] with a “disease-causing” (DM) label.

Ancestry inference was performed in this cohort using principal component analysis. Data from the 1000 genomes project (phase 3) dataset was used and five broad super-populations were projected (European, African, Admixed American, East Asian, South Asian) (further information on this can be found online at https://research-help.genomicsengland.co.uk/display/GERE/Ancestry+inference).

We focused on a pre-determined subset of the Genomics England 100,000 Genomes Project dataset that includes only unrelated probands (n=29,556). Of these, 105 had a diagnosis of albinism, *i.e.* the ICD-10 term “Albinism” [E70.3] and/or the HPO terms “Albinism” [HP:0001022], “Partial albinism” [HP:0007443] or “Ocular albinism” [HP:0001107] were assigned. Together with the University Hospital of Bordeaux cases, these 105 probands formed the ‘case’ cohort. The remaining 29,451 probands had no recorded diagnosis or phenotypic features of albinism and formed the ‘control’ cohort. We note

that in-depth ophthalmic phenotyping was not routinely undertaken in Genomics England 100,000 Genomes Project participants. Thus, we cannot be certain that a small number of individuals with mild/subclinical forms of albinism are not included in the control cohort.

Case-control analysis to estimate the probability of an albinism diagnosis

The effect of heterozygosity for the *TYR:*c.1205G>A (p.Arg402Gln) [rs1126809] and *OCA2:*c.1327G>A (p.Val443Ile) [rs74653330] variant combination on the probability of receiving a diagnosis of albinism (*i.e.* the albinism risk) was estimated using data from the University Hospital of Bordeaux albinism cohort and the Genomics England 100,000 Genomes Project dataset. A case-control analysis of a binary trait (presence/absence of albinism) was conducted assuming a recessive model. Logistic regression using the Firth bias reduction method[33,34] was utilised (as implemented in "logistf" R package). The following parameters were considered: number of rare albinism-implicated variants, number of common albinism-associated alleles, ancestry and sex.

Rare albinism-implicated variants: Focusing on 19 albinism-related genes *(TYR, OCA2, TYRP1, SLC45A2, SLC24A5, C10ORF11, GPR143, HPS 1* to *10, LYST, SLC38A8*), we identified rare albinism-implicated changes by selecting variants that:

- are labelled as disease-causing (DM) in HGMD v2021.2.
- are included in the following HGMD v2021.2 "mutation type" categories: missense/nonsense, splicing, small deletions, small insertions or small indels; gross deletions, gross insertions/duplications and complex rearrangements were not analysed.
- have "total" MAF <1% in gnomAD v2.1.1[32].

Although *OCA2:*c.1327G>A qualified for inclusion in this group, it was excluded as it was part of the evaluated genotype (*i.e.* heterozygosity for the *TYR:*c.1205G>A and *OCA2:*c.1327G>A variant combination).

Common albinism-associated alleles: We recently reported three common changes in *TYR* that have a role in the genetic architecture of albinism.[18] These are:

- *TYR* c.-301C>T [rs4547091], located in the *TYR* promoter; the derived, non-ancestral allele c.-301C (or c.-301=) has a minor allele frequency (MAF) of ~60% in non-Finnish European (NFE) populations in the Genome Aggregation Database (gnomAD v2.1.1)[32];

- *TYR* c.575C>A (p.Ser192Tyr) [rs1042602], which has a MAF of ~36% in gnomAD NFE populations;
- *TYR* c.1205G>A (p.Arg402Gln) [rs1126809], which has a MAF of ~27% in gnomAD NFE populations; this variant was excluded from this group as it was part of the evaluated genotype.

It is highlighted that the *TYR* c.-301C and the *TYR* c.575C>A (p.Ser192Tyr) alleles are frequently found in linkage disequilibrium so the presence of the latter would, in most cases, suggest the presence of the former on the same haplotype.

Ancestry: Ancestry was factored in our analyses but our approach was imperfect as it was not possible to objectively determine genetic ancestry in the University Hospital of Bordeaux albinism cohort (as mentioned above, self-identified ethnicity was instead used as a surrogate). Confounding by ancestry (*i.e* population stratification) is therefore a possibility.[52] We attempted to quantify this by calculating the genomic inflation factor lambda (λGC). A total of 35 single-nucleotide variants were selected using a previously described approach (see Supplementary Table 5 and [18]). Case-control comparisons were then made for each of these 35 λGC markers using Firth regression analysis. The resulting test statistics were used to calculate the median value of λGC.[18]

Direct case-control comparison ("single genotype analysis"): Our first analysis focused on estimating the effect size of heterozygosity for the *TYR*:c.1205G>A and *OCA2*:c.1327G>A variant combination. A comparison was made between participants who carried this genotype and those who did not. Initially, we focused on individuals who were estimated to have European ancestries and who carried no other rare albinism-implicated variants (n=361 cases and n=20,304 controls). Firth regression analysis was performed using sex and the number of common albinism-associated alleles as covariates. Subsequently, we repeated the analysis on the full, multi-ancestry case and control cohorts (n=1,121 cases and n=29,451 controls) using sex, the number of common albinism-associated variants, the number of rare albinism-implicated variants and ancestry as covariates (the odds ratio for this analysis was 6.5, the 95% confidence interval was 4 – 10.3 and the p-value was 4.4 x 10^{-11}).

Case-control comparisons between selected TYR/OCA2 genotype groups ("genotype group analysis"): Our second analysis focused on performing case-control comparisons between groups of

genotypes that include different alleles at positions *TYR*:c.1205 and *OCA2*:c.1327. These groups were:

- *group A* (reference group): homozygous for *TYR*:c.1205= and *OCA2*:c.1327=
- *group B* (single heterozygote group): heterozygous for either the *TYR*:c.1205G>A and *OCA2*:c.1327= variant combination or for the *TYR*:c.1205= and *OCA2*:c.1327G>A variant combination.
- *group C* (dual heterozygote group): heterozygous for the *TYR*:c.1205G>A and *OCA2*:c.1327G>A variant combination.
- *group D* (homozygous combination group): either homozygous for *TYR*:c.1205G>A and heterozygous for *OCA2*:c.1327G>A; or heterozygous for *TYR*:c.1205G>A and homozygous for *OCA2*:c.1327G>A; or homozygous for both changes.

Group A was used as the reference group to which all other groups were compared using Firth regression analysis (i.e. the odds ratio for this group was fixed at 1). Comparisons with two additional groups which included homozygotes for *TYR*:c.1205G>A (group E) and homozygotes for *OCA2*:c.1327G>A (group F) were also performed. Initially, we focused on individuals who were estimated to have European ancestries and who carried no other rare albinism-implicated variants (n=361 cases and n=20,304 controls); sex and the number of common albinism-associated alleles were used as covariates (Fig.2; Supplementary Fig.1a; Supplementary Table 2). Subsequently, we repeated the analysis on the full, multi-ancestry case and control cohorts (n=1,121 cases and n=29,451 controls) using sex, the number of common albinism-associated alleles, the number of rare albinism-implicate variants and ancestry as covariates (Supplementary Fig.1b; Supplementary Table 6).

Replication of findings in the UK Biobank cohort

The effect of heterozygosity for the *TYR:*c.1205G>A and *OCA2:*c.1327G>A variant combination was studied in the UK Biobank. UK Biobank is a biomedical resource containing in-depth genetic and health information from >500,000 individuals from across the UK.[40] A subset of UK Biobank volunteers underwent enhanced phenotyping including visual acuity testing (131,985 individuals) and imaging of the central retina (84,748 individuals)[53]; the latter was obtained using optical coherence tomography (OCT), a non-invasive imaging test that rapidly generates cross-sectional retinal scans at

micrometre-resolution.[54] All UK Biobank volunteers whose data were analysed as part of this study were imaged using the 3D OCT-1000 Mark II device (Topcon, Japan); the relevant methodology has been previously described.[53] Notably, only 24 UK Biobank participants have been assigned a diagnosis of albinism (data field 41270; ICD-10 term “Albinism” [E70.3]) of which only 7 had visual acuity measurements and none had OCT imaging; 19 additional individuals had a diagnosis of albinism in their primary care record data (resource 591).

Aiming to assess the robustness of the findings in the combined University Hospital of Bordeaux and Genomics England 100,000 Genomes Project dataset, we performed case-control analyses in UK Biobank volunteers.

First, genotyping array data were used to obtain genotypes for the *TYR:*c.1205G>A variant (data field 22418 including information from the Applied Biosystems UK Biobank Axiom Array containing 825,927 markers). In contrast to this change, the *OCA2:*c.1327G>A variant was not directly captured by the array. However, high-quality (>99.9%) imputation data were available (data field 22828). Only individuals successfully genotyped at both sites were retained.

At first, Firth regression analysis was used to directly compare albinism susceptibility between those UK Biobank participants who are heterozygous for the *TYR*:c.1205G>A and *OCA2*:c.1327G>A variant combination and those who are not. Only individuals who were estimated by principal component analysis to have European ancestries (data field 22006) were included in this analysis; sex and the number of common albinism-associated alleles were used as covariates. Subsequently, we assigned UK Biobank volunteers of European ancestries into *TYR/OCA2* genotype groups (as described above) and performed comparisons between group A (which was used as the reference group) and all the other groups; sex and the number of common albinism-associated alleles were used as covariates (Fig.3a; Supplementary Table 4).

Given that reduced visual acuity and increased central retinal thickness (due to underdevelopment of the fovea) are two hallmark features of albinism we investigated the impact of selected *TYR/OCA2* variant combinations on these quantitative traits.

We first obtained data on the left LogMAR visual acuity for UK Biobank volunteers (data field 5201, “instance 0” datasets). These visual acuity measurements were then used to compare visual

performance between groups of volunteers with selected *TYR/OCA2* genotype combinations. As the obtained distributions deviated from normality, the Kruskal-Wallis test was used. Pair-wise comparisons were performed and the p-values were adjusted using the Benjamini-Hochberg method (Fig.3b; Supplementary Table 4).

We then obtained central retinal thickness measurements from UK Biobank OCT images using the "macular thickness at the central subfield" data field (27802; defined as the average distance between the hyperreflective bands corresponding to the RPE and the internal limiting membrane (ILM), across the central 1 mm diameter circle of the ETDRS grid) for UK Biobank volunteers.[55] The obtained measurements were subsequently used to compare central retinal thickness between groups of UK Biobank volunteers with selected *TYR/OCA2* genotype combinations. As some of the obtained distributions deviated from normality, the Kruskal-Wallis test was used. Pair-wise comparisons were performed and the p-values were adjusted using the Benjamini-Hochberg method (Fig.3c; Supplementary Table 4).

It is noted that we opted to perform all UK Biobank analyses on left eye data. This is because, as per study protocol[53], the left eye was tested after the right eye, potentially making the measurements from the left side less prone to artifacts (as the participants were more familiar with the test).

Ethics approval

Informed consent was obtained from all participants or their parents in the case of minors. The study of individuals from the University Hospital of Bordeaux albinism cohort has been approved by the relevant local ethics committee (Comité de Protection des Personnes Sud-Ouest et Outre Mer III, Bordeaux, France). The informed consent process for the Genomics England 100,000 Genomes Project has been approved by the National Research Ethics Service Research Ethics Committee for East of England – Cambridge South Research Ethics Committee. The UK Biobank has received approval from the National Information Governance Board for Health and Social Care and the National Health Service North West Centre for Research Ethics Committee (Ref: 11/NW/0382). All investigations were conducted in accordance with the tenets of the Declaration of Helsinki.

DATA AVAILABILITY

Genomics England 100,000 Genomes Project data are available under restricted access through a procedure described at https://www.genomicsengland.co.uk/about-gecip/for-gecip-members/data-and-data-access. UK Biobank data are available under restricted access through a procedure described at http://www.ukbiobank.ac.uk/using-the-resource/. All other data supporting the findings of this study are available within the article (including its supplementary information files).

CODE AVAILABILITY

The scripts used to analyse the datasets included in this study are available at https://github.com/davidjohngreen/tyr-oca2.

REFERENCES

1. Boycott, K.M. *et al*. A diagnosis for all rare genetic diseases: The horizon and the next frontiers. *Cell*.**177**, 32-37 (2019). doi: 10.1016/j.cell.2019.02.040.
2. Splinter, K. *et al*. Effect of genetic diagnosis on patients with previously undiagnosed disease. *N Engl J Med*. **379**, 2131-2139 (2018). doi: 10.1056/NEJMoa1714458.
3. Posey, J.E. *et al*. Insights into genetics, human biology and disease gleaned from family based genomic studies. *Genet Med*. **21**, 798-812 (2019). doi: 10.1038/s41436-018-0408-7.
4. Zurek, B. *et al*. Solve-RD: systematic pan-European data sharing and collaborative analysis to solve rare diseases. *Eur J Hum Genet.* **29**, 1325-1331 (2021). doi: 10.1038/s41431-021-00859-0
5. 100,000 Genomes Project Pilot Investigators. 100,000 Genomes Pilot on rare-disease diagnosis in health care - Preliminary report. *N Engl J Med*. **385**, 1868-1880 (2021). doi: 10.1056/NEJMoa2035790.
6. Boycott, K.M. *et al*. Care4Rare Canada: Outcomes from a decade of network science for rare disease gene discovery. *Am J Hum Genet*. **109**, 1947-1959 (2022). doi: 10.1016/j.ajhg.2022.10.002.
7. Schäffer, A.A. Digenic inheritance in medical genetics. *J Med Genet*. **50**, 641-652 (2013). doi: 10.1136/jmedgenet-2013-101713.
8. Katsanis, N. The continuum of causality in human genetic disorders. *Genome Biol*. **17**, 233 (2016). doi: 10.1186/s13059-016-1107-9.
9. Deltas C. Digenic inheritance and genetic modifiers. *Clin Genet*. **93**, 429-438 (2018). doi: 10.1111/cge.13150.
10. Nachtegael, C. *et al*. Scaling up oligogenic diseases research with OLIDA: the Oligogenic Diseases Database. *Database (Oxford).* **2022**, baac023 (2022). doi: 10.1093/database/baac023.
11. Cooper, D.N., Krawczak, M., Polychronakos, C., Tyler-Smith, C. & Kehrer-Sawatzki, H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum Genet.* **132**, 1077-1130 (2013). doi: 10.1007/s00439-013-1331-2.
12. Kousi, M. & Katsanis, N. Genetic modifiers and oligogenic inheritance. *Cold Spring Harb Perspect Med*. **5**, a017145 (2015). doi: 10.1101/cshperspect.a017145.
13. Costanzo, M. *et al*. Global genetic networks and the genotype-to-phenotype relationship. *Cell*. **177**, 85-100 (2019). doi: 10.1016/j.cell.2019.01.033.
14. Mukherjee, S. *et al*. Identifying digenic disease genes via machine learning in the Undiagnosed Diseases Network. *Am J Hum Genet.* **108**, 1946-1963 (2021). doi: 10.1016/j.ajhg.2021.08.010.
15. Papadimitriou, S. *et al*. Predicting disease-causing variant combinations. *Proc Natl Acad Sci U S A*. **116**:11878-11887 (2019). doi: 10.1073/pnas.1815601116.
16. Brehm, A. *et al*. Additive loss-of-function proteasome subunit mutations in CANDLE/PRAAS patients promote type I IFN production. *J Clin Invest*. **125**, 4196-4211 (2015). doi: 10.1172/JCI81260.
17. Bakker, R. *et al*. The retinal pigmentation pathway in human albinism: Not so black and white. *Prog Retin Eye Res*. **91**, 101091 (2022). doi: 10.1016/j.preteyeres.2022.101091.
18. Michaud, V. *et al*. The contribution of common regulatory and protein-coding TYR variants to the genetic architecture of albinism. *Nat Commun*. **13**, 3939 (2022). doi: 10.1038/s41467-022-31392-3.
19. Campbell, P. *et al*. Clinical and genetic variability in children with partial albinism. *Sci Rep*. **9**, 16576 (2019). doi: 10.1038/s41598-019-51768-8.
20. Monfermé, S. *et al*. Mild form of oculocutaneous albinism type 1: phenotypic analysis of compound heterozygous patients with the R402Q variant of the *TYR* gene. *Br J Ophthalmol*. **103**, 1239-1247 (2019). doi: 10.1136/bjophthalmol-2018-312729
21. Lasseaux, E. *et al*. Molecular characterization of a series of 990 index patients with albinism. *Pigment Cell Melanoma Res*. **31**, 466-474 (2018). doi: 10.1111/pcmr.12688.
22. Caulfield, M. *et al*. The National Genomic Research Library. https://doi.org/10.6084/m9.figshare.4530893.v7 (2017).
23. Okazaki, A. & Ott, J. Machine learning approaches to explore digenic inheritance. *Trends Genet*. **38**, 1013-1018 (2022). doi: 10.1016/j.tig.2022.04.009.

24. Jagirdar, K. *et al*. Molecular analysis of common polymorphisms within the human tyrosinase locus and genetic association with pigmentation traits. *Pigment Cell Melanoma Res*. **27**, 552-564 (2014). doi: 10.1111/pcmr.12253.
25. Dolinska, M.B. *et al*. Oculocutaneous albinism type 1: link between mutations, tyrosinase conformational stability, and enzymatic activity. Pigment Cell Melanoma Res. **30**, 41-52 (2017). doi: 10.1111/pcmr.12546.
26. Sviderskaya, E.V. *et al*. Complementation of hypopigmentation in p-mutant (pink-eyed dilution) mouse melanocytes by normal human P cDNA, and defective complementation by OCA2 mutant sequences. *J Invest Dermatol*. **108**, 30-34 (1997). doi: 10.1111/1523-1747.ep12285621
27. Bellono, N.W., Escobar, I.E., Lefkovith, A.J., Marks, M.S. & Oancea, E. An intracellular anion channel critical for pigmentation. *Elife*. **3**, e04543 (2014). doi: 10.7554/eLife.04543.
28. Bellono, N.W., Escobar, I.E. & Oancea, E. A melanosomal two-pore sodium channel regulates pigmentation. *Sci Rep*. **6**, 26570 (2016). doi: 10.1038/srep26570.
29. Pavan W.J. & Sturm, R.A. The genetics of human skin and hair pigmentation. *Annu Rev Genomics Hum Genet*. **20**, 41-72 (2019). doi: 10.1146/annurev-genom-083118-015230.
30. Landrum, M.J. *et al*. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062-D1067 (2018). doi: 10.1093/nar/gkx1153. [accessions VCV000003779.25 & VCV000000955.52].
31. Buniello, A. *et al*. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005-D1012 (2019). doi: 10.1093/nar/gky1120. [entries for rs1126809 and rs74653330]
32. Karczewski, K.J. *et al*. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. **581**, 434-443 (2020). doi: 10.1038/s41586-020-2308-7. [entries for 11-89017961-G-A (GRCh37) and 15-28228553-C-T (GRCh37)]
33. Heinze, G. & Schemper, M. A solution to the problem of separation in logistic regression. *Stat Med.* **21**, 2409-19 (2002). doi: 10.1002/sim.1047.
34. Firth, D. Bias reduction of maximum likelihood estimates. *Biometrika*. **80**, 27-38 (1993). doi: 10.1093/biomet/80.1.27
35. Stenson, P.D. *et al*. The Human Gene Mutation Database (HGMD®): optimizing its use in a clinical diagnostic or research setting. *Hum Genet*. **139**, 1197-1207 (2020). doi: 10.1007/s00439-020-02199-3.
36. Fahed, A.C. *et al*. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat Commun*. **11**, 3635 (2020). doi: 10.1038/s41467-020-17374-3.
37. Li, S. *et al*. Cancer risks associated with *BRCA1* and *BRCA2* pathogenic variants. *J Clin Oncol*. **40**, 1529-1541 (2022). doi: 10.1200/JCO.21.02112.
38. Han, X. *et al*. Myocilin gene Gln368Ter variant penetrance and association with glaucoma in population-based and registry-based studies. *JAMA Ophthalmol*. **137**, 28-35 (2019). doi: 10.1001/jamaophthalmol.2018.4477.
39. Zhou, W. *et al*. Global Biobank Meta-analysis Initiative: Powering genetic discovery across human disease. *Cell Genom* **2**, 100192 (2022). doi: 10.1016/j.xgen.2022.100192
40. Bycroft, C. *et al*. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. **562**, 203-209 (2018). doi: 10.1038/s41586-018-0579-z.
41. Masson, E. *et al*. Expanding ACMG variant classification guidelines into a general framework. *Hum Genomics*. **16**, 31 (2022). doi: 10.1186/s40246-022-00407-x.
42. Richards, S. *et al*. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. **17**, 405-424 (2015). doi: 10.1038/gim.2015.30.
43. Freedman, M.L. *et al*. Assessing the impact of population stratification on genetic association studies. *Nat Genet*. **36**, 388-93 (2004). doi: 10.1038/ng1333.
44. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics*. **55**, 997-1004 (1999). doi: 10.1111/j.0006-341x.1999.00997.x
45. Dadd, T., Weale, M.E. & Lewis, C.M. A critical evaluation of genomic control methods for genetic association studies. *Genet Epidemiol*. **33**, 290-298 (2009). doi: 10.1002/gepi.20379.
46. Turnbull, C. *et al.* The 100,000 Genomes Project: bringing whole genome sequencing to the NHS. *BMJ* **361**, 1687 (2018).
47. Köhler, S. *et al.* The Human Phenotype Ontology in 2021. *Nucleic Acids Res.* **49**, D1207–D1217 (2021).

48. Raczy, C. *et al.* Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics* **29**, 2041–2043 (2013).
49. Tan, A., Abecasis, G.R. & Kang, H.M. Unified representation of genetic variants. *Bioinformatics* **31**, 2202–2204 (2015).
50. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
51. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, (2016).
52. Marchini, J., Cardon, L.R., Phillips, M.S. & Donnelly, P. The effects of human population structure on large genetic association studies. *Nat Genet.* **36**, 512–517 (2004).
53. Chua, S.Y.L., *et al.* Cohort profile: design and methods in the eye and vision consortium of UK Biobank. *BMJ Open* **9**, e025077 (2019).
54. Drexler, W. & Fujimoto, J.G. State-of-the-art retinal optical coherence tomography. *Prog Retin Eye Res*. **27**, 45–88 (2008).
55. Currant, H. *et al*. Genetic variation affects morphological retinal phenotypes extracted from UK Biobank optical coherence tomography images. *PLoS Genet*. **17**, e1009497 (2021).

ACKNOWLEDGEMENTS

We acknowledge the following sources of funding: the Wellcome Trust (224643/Z/21/Z, Clinical Research Career Development Fellowship to P.I.S.; 200990/Z/16/Z, Transforming Genetic Medicine Initiative to G.C.B.); the UK National Institute for Health Research (NIHR) Clinical Lecturer Programme (CL-2017-06-001 to P.I.S.); Retina UK and Fight for Sight (GR586, RP Genome Project - UK Inherited Retinal Disease Consortium to G.C.B.); Christopher Green (D.J.G.); the French Albinism Association (Genespoir) and the French National Research Agency (Agence Nationale de la Recherche; ANR-21-CE17-0041-01 to B.A.).

The UK Biobank Eye and Vision Consortium is supported by funding from the NIHR Biomedical Research Centre at Moorfields Eye Hospital and UCL Institute of Ophthalmology, the Alcon Foundation and the Desmond Foundation. The complete list of members of this Consortium can be found in the Supplementary Information

This research was made possible through access to the data and findings generated by the 100,000 Genomes Project. The 100,000 Genomes Project is managed by Genomics England Limited (a wholly owned company of the Department of Health

and Social Care). The 100,000 Genomes Project is funded by the NIHR and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure. The 100,000 Genomes Project uses data provided by patients and collected by the National Health Service (NHS) as part of their care and support. We acknowledge the contribution of the Genomics England Research Consortium to the 100,000 Genomes Project. The complete list of members of this Consortium can be found in the Supplementary Information.

Lastly, we acknowledge the help of Cécile Courdier at the University Hospital of Bordeaux Molecular Genetics Laboratory, Sophie Javerzat at the University of Bordeaux, Dave Gerrard at the University of Manchester, and Claire Hardcastle, Chris Campbell, Steph Barton and Simon Ramsden at the North West of England Genomic Laboratory Hub.

AUTHOR CONTRIBUTIONS STATEMENT

P.I.S. conceived and designed the experiments. The UK Biobank Eye and Vision Consortium, T.F., E.B., B.A., G.C.B. and P.I.S provided datasets and analytical tools. D.J.G., V.M., E.L., C.P., T.F., B.A. and P.I.S. analysed the data. P.I.S. wrote the manuscript with support from D.J.G. All authors critically revised and approved the manuscript.

G.C.B. and B.A. contributed equally to this work.

COMPETING INTERESTS STATEMENT

E.B. is a paid consultant and equity holder of Oxford Nanopore, a paid consultant to Dovetail, and a non-executive director of Genomics England, a limited company wholly owned by the UK Department of Health and Social Care. All other authors declare no competing interests.

UK Biobank Eye and Vision Consortium

Graeme C. Black[1,5], Panagiotis I. Sergouniotis[1,4,5,6]

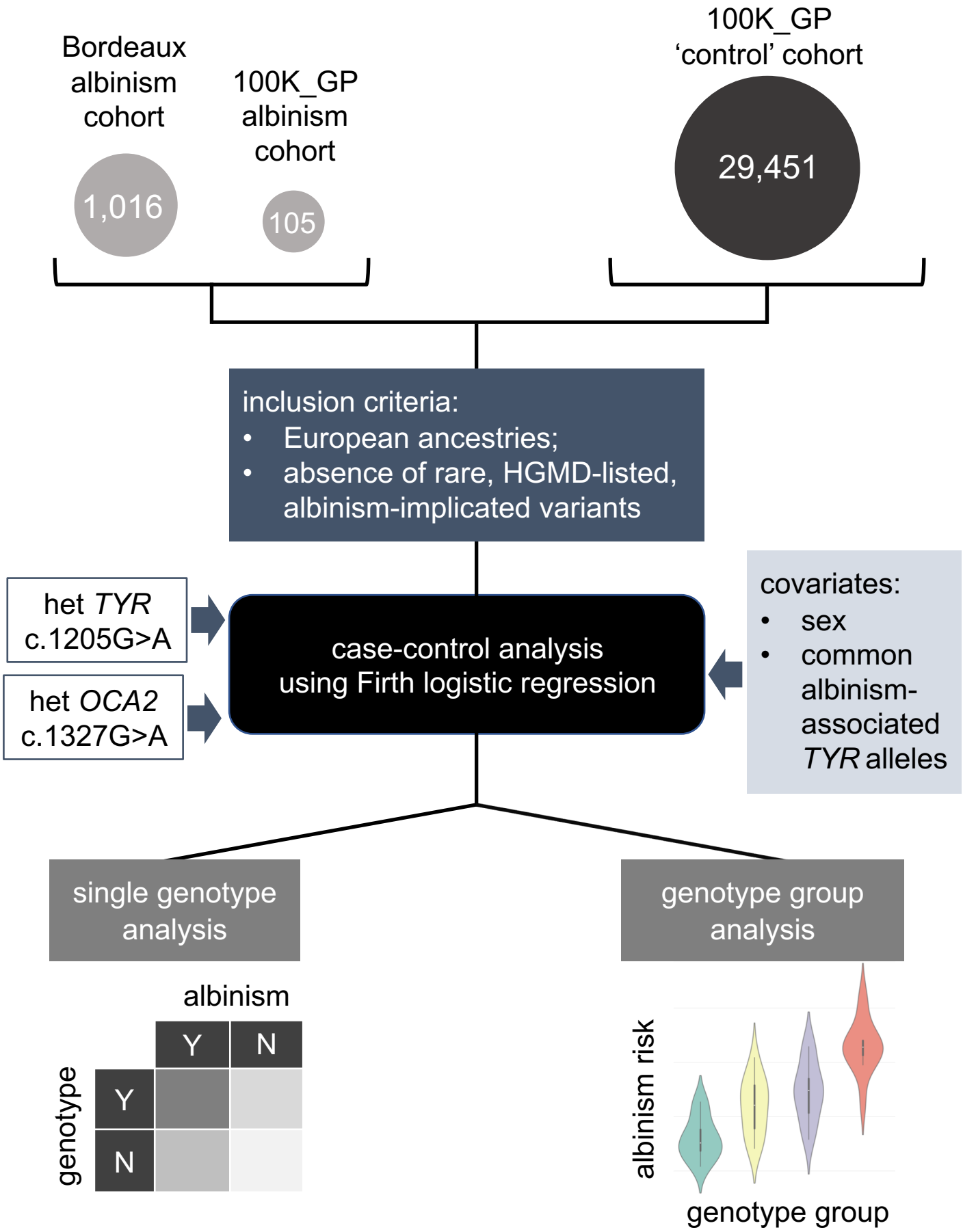
[1] Division of Evolution, Infection and Genomics, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK.

[4] European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Cambridge, UK.

[5] Manchester Centre for Genomic Medicine, Saint Mary's Hospital, Manchester University NHS Foundation Trust, Manchester, UK.

[6] Manchester Royal Eye Hospital, Manchester University NHS Foundation Trust, Manchester, UK.

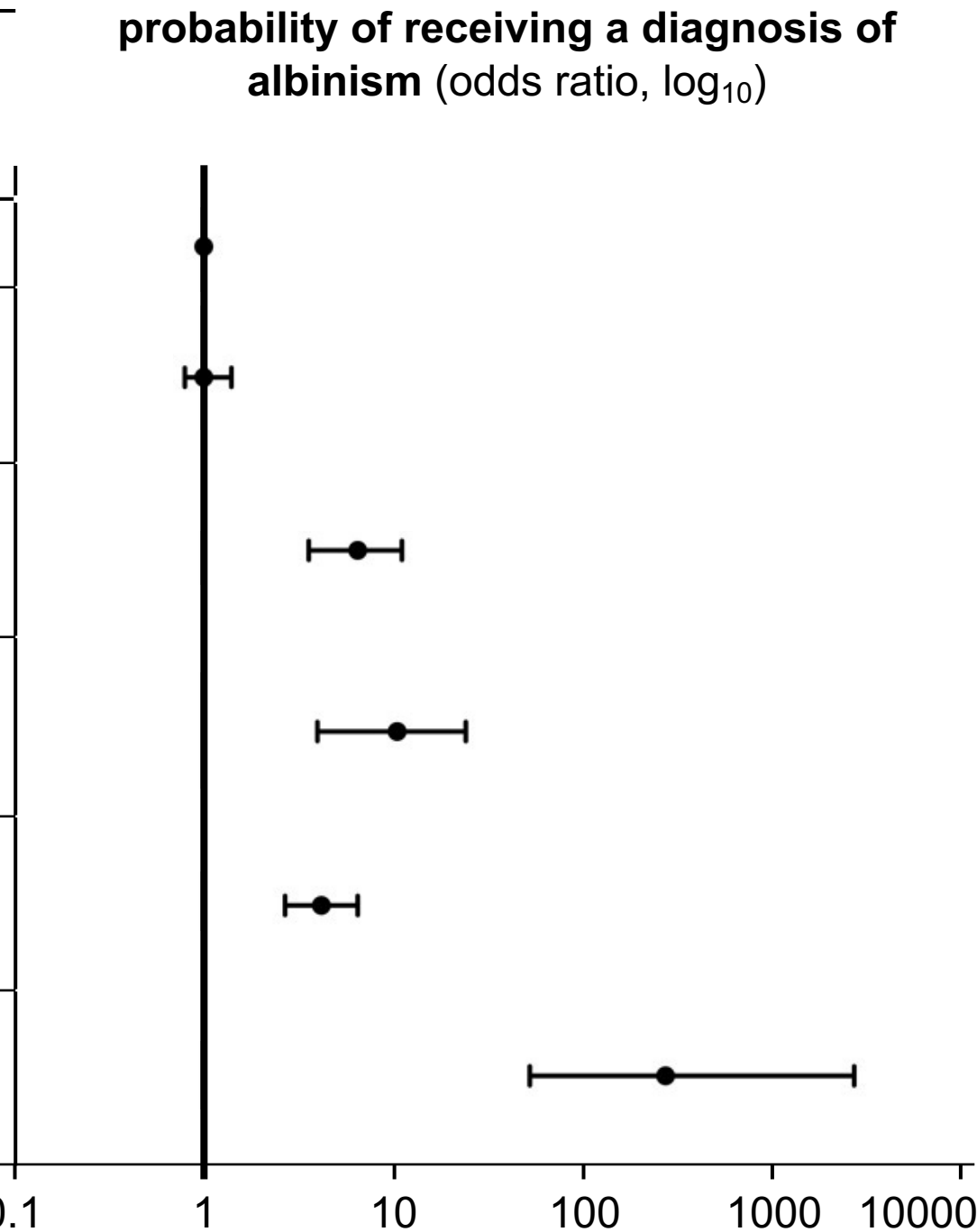
A full list of members appears in the Supplementary Information.

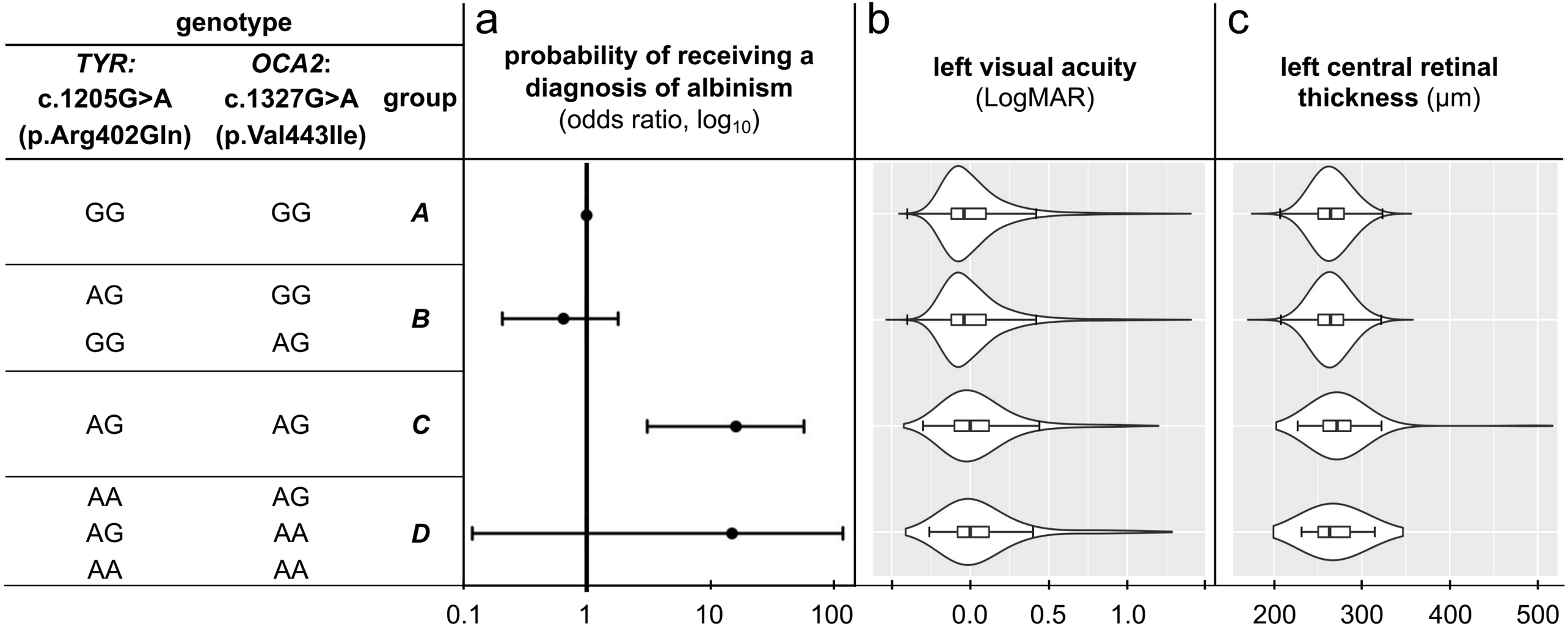


validation:

- analyses repeated using different inclusion criteria & covariates
- analyses repeated & expanded in an independent cohort (UKB)

genotype		
TYR: **c.1205G>A (p.Arg402Gln)**	***OCA2*:** **c.1327G>A (p.Val443Ile)**	**group**
GG	GG	***A***
AG GG	GG AG	***B***
AG	AG	***C***
AA AG AA	AG AA AA	***D***
AA	GG	***E***
GG	AA	***F***




genotype
TYR: c.1205G>A (p.Arg402Gln)
OCA2: c.1327G>A (p.Val443Ile)
group
GG GG A
AG GG
GG AG B
AG AG C
AA AG
AG AA
AA AA D
a
probability of receiving a diagnosis of albinism (odds ratio, log10)
0.1 1 10 100
b
left visual acuity (LogMAR)
0.0 0.5 1.0
c
left central retinal thickness (µm)
200 300 400 500