Title Page

Title:
A comparative GWAS of eye colour in light and dark eye genetic backgrounds defined by *HERC2* rs12913832 polymorphism

Authors:
Cristina L Abbatangelo, MSc[1,5*], Frida Lona Durazo[2,3], PhD, Melissa Edwards, PhD[4], Esteban J Parra, PhD[5]

Author affiliations:
[1]Department of Anthropology, Faculty of Arts and Science, University of Toronto, Toronto, ON, Canada
[2]Montréal Heart Institute, Montréal, QC, Canada
[3]Université de Montréal, Montréal, QC, Canada
[4]Brain Tumour Research Center, SickKids Research Institute, Peter Gilgan Center for Research and Learning, Toronto, ON, Canada
[5]Department of Anthropology, University of Toronto Mississauga, Mississauga, ON, Canada

*** Corresponding author:**
Cristina Laura Abbatangelo
c.abbatangelo@mail.utoronto.ca

Short title: *HERC2* comparative eye colour GWAS

Keywords: rs12913832, eye colour, genome-wide association study (GWAS), *HERC2*, *OCA2,* quantitative eye colour measures

Abstract

rs12913832, a polymorphism located in an enhancer within the *HERC2* gene, which is known to regulate *OCA2* transcription, is heavily relied upon as a predictor of light versus dark eyes. Individuals with the GG genotype are projected to have blue eyes, while individuals with the AA or AG genotypes are projected to have darker eye colours (primarily brown). However, eye colour is a polygenic trait, and previous studies have revealed that a significant proportion of individuals self-report an eye colour that is not concordant with their genotype at rs12913832. Herein, we address the question: What common markers are influencing eye colour in individual's whose self-reported phenotype does not correspond to the expected phenotype based on their rs12913832 genotype? Building upon our prior investigation of iris pigmentation genetics in individuals with an expected "blue eye" background (rs12913832:GG genotype) in a sample of the Canadian Partnership for Tomorrow's Health (CanPath) cohort, this study extends the analysis to include individuals with an expected "brown eye" background (rs12913832:AA+AG). We identified variants in *SLC45A2*, *TYRP1*, *TYR*, *SLC24A4* and *TSPAN10*, which may influence eye colour presentation in individuals with the rs12913832:GG genotype and variants in *IRF4*, *TYRP1* and *OCA2*, which may influence eye colour presentation in individuals with the rs12913832:AA+AG genotype. These markers include well-known pigmentation-associated single nucleotide polymorphisms, such as rs16891982 (*SLC45A2*), rs1126809 (*TYR*), rs12203592 (*IRF4*), rs1800407 (*OCA2*) and rs6420484 (*TSPAN10*). Several of these loci were replicated using independent quantitative eye colour measures, including heterochromia and CIELAB colour dimensions. This research highlights the importance of gene-gene interactions and the polygenic nature of pigmentation traits, emphasizing modifying effects that can sometimes counteract the

60 dominant influence of rs12913832, contributing to advancements in pigmentation genetics and
61 forensic applications.

62 Introduction

63 Human pigmentation traits, such as hair, skin, and eye colour, represent some of the most striking
64 examples of visible genetic variation. These traits are highly polygenic, with complex gene-gene
65 and gene-environment interactions contributing to their phenotypic expression. Understanding the
66 genetic architecture underlying these traits not only informs biological mechanisms but also
67 enhances applications such as predictive modeling, reconstruction of ancestral phenotypes and
68 human identification. Of particular interest is the genetic basis underlying iris pigmentation,
69 which remains incompletely understood despite significant advances in large-scale genome-wide
70 association studies (GWAS) (1).

71 A key genetic determinant of human eye colour is the single nucleotide polymorphism (SNP)
72 rs12913832, located in a conserved enhancer region within the *HERC2* gene that regulates
73 transcription of the neighbouring *OCA2* gene (2–5). *OCA2* plays a central role in melanin
74 biosynthesis, and its expression levels are strongly influenced by allelic variation at rs12913832.
75 The ancestral A allele facilitates transcription factor binding and promotes long-range chromatin
76 looping, enabling physical interaction between the *OCA2* promoter and the *HERC2* enhancer. This
77 configuration enhances *OCA2* expression, leading to increased melanin production and resulting
78 in darker eye colours, such as brown and hazel (4,6). Conversely, the derived G allele reduces
79 chromatin looping efficiency, thereby decreasing *OCA2* expression and melanin production. This
80 reduction is associated with lighter iris pigmentation, characteristic of blue eyes (4,6). As such,

81 rs12913832 has become a key marker in pigmentation research and forensic DNA phenotyping
82 (FDP), due to its strong predictive value for distinguishing between light and dark eye colours
83 (2,4,7). Individuals with the GG genotype are typically predicted to have blue eyes, while those
84 with AG or AA genotypes are generally expected to exhibit brown eyes (**Fig 1**). However, while
85 these patterns hold true for most individuals, particularly within European populations where eye
86 colour diversity is greatest, genotype-phenotype mismatches have been reported.

87 **Fig 1**. **Expected eye colour phenotype based on rs12913832 genotype.**

88 Specifically, previous research has illustrated that some rs12913832:GG individuals present with
89 eye colours other than blue, and some rs12913832:AA or AG individuals show eye colours other
90 than brown (8–10). A recent publication found a notable discrepancy in the Canadian Partnership
91 for Tomorrow's Health (CanPath) cohort where 33% of individuals with the rs12913832:GG
92 genotype self-reported eye colours other than blue (11). This finding challenges the notion that
93 the rs12913832:GG genotype exclusively corresponds to blue eyes and highlights the presence of
94 additional genetic or regulatory factors that may modulate melanin production, resulting in a
95 broader spectrum of iris pigmentation. In a previous study (12), we stratified participants of the
96 CanPath cohort based on rs12913832 genotypes and identified a number of SNPs that seem to
97 modify the effect of the rs12913832 GG genotype, contributing to an explanation as to why
98 individuals self-report "non-blue" eye colour in spite of having severely reduced expression of the
99 *OCA2* gene, which usually results in blue eye colour.

100 The present study builds on this foundation by extending the 2023 study in two important ways:
101 1) Identifying in the same CanPath sample used in the 2023 study potential genetic modifiers that
102 may explain individuals with rs12913832:AA+AG genotypes reporting eye colours other than

103 brown and hazel, and 2) Replicating the same markers in another independent sample of
104 individuals of European ancestry, for whom detailed quantitative iris pigmentation measures are
105 available (including measures capturing central heterochromia, defined as the presence of different
106 eye colours in the inner and outer portions of the iris). By incorporating additional GWAS and
107 post-GWAS analyses, this work aims to address the following research question: **What common**
108 **markers are influencing eye colour variation in individuals whose self-reported phenotype**
109 **does not correspond to the expected phenotype associated with their genotype at rs12913832?**
110 Analyzing both self-reported data and detailed quantitative measures of iris colour, we describe
111 polymorphisms that act to counteract the strong influence of rs12913832 on iris colour.

112 # Methods

113 *Study Dataset and Design*

114 This research was conducted under the approval of the University of Toronto Research Ethics
115 Board (REB), adhering to Human Research Protocol #36429. The study dataset was drawn from
116 the Canadian Partnership for Tomorrow's Health (CanPath) biobank, Canada's largest collection
117 of population health data, containing genomic and phenotypic information from over 300,000
118 Canadians of predominantly European ancestry (13,14). Access to the data was granted through
119 CanPath, under application number DAO-034431 (Research Project Number 329058, 2018-10-01
120 start date). All analyses were conducted using de-identified data. The authors did not participate
121 in data collection and had no access to personally identifiable information at any stage of the
122 research. All participants provided written informed consent at the time of data collection by
123 CanPath.

The study utilized a subset of 5,481 individuals for which self-reported eye colour was available (N_{GG}=2,724 and N_{AA+AG}=2,757), encompassing various Canadian provinces. **Fig. 2** illustrates the provincial cohorts represented in the biobank. Among this group, 56.29% were female, and the average age was approximately 55 years (standard error ± 0.12).

CanPath participants underwent genotyping between 2012 and 2018. Two genotyping chips were utilized: the Axiom 2.0 UK Biobank Array (Affymetrix) (UKBB) for (*N*=3,212) individuals and the Global Screening Array (GSA) 24v1+MDP (*N*=2,429) participants, conducted by CanPath. The genotyping arrays analyzed between 658,296 and 813,168 SNPs. Imputation was carried out previously on the Sanger Imputation Server using the Haplotype Reference Consortium dataset as a reference for imputation (11,12).

Fig 2. Study cohorts and provinces participating in the Canadian Partnership for Tomorrow's Health (CanPath). Additional details about the study and participating cohorts can be found in (14) and at canpath.ca.

Among the 2,724 participants with rs12913832:AA or AG genotypes and available self-reported eye colour data, 512 reported eye colours other than brown or hazel. This observation suggests that the rs12913832:AA+AG genotype alone does not fully account for dark eye pigmentation and points to the possibility that other genetic variants may influence eye colour in individuals with the AA+AG genotype.

To investigate, we piggybacked off the 2023 rs12913832:GG study (12) and applied the same methods for AA+AG individuals in the CanPath sample, as well as a separate replication dataset. **Fig 3** illustrates the step-by-step computational methodology.

145 **Fig3**. **Step-by-step computational methodology**. Flow chart illustrating the computational steps used to perform a
146 genome-wide association study (GWAS) to identify genetic variants associated with eye colour in rs12913832:AA+AG
147 individuals.

148

149 *Preliminary Statistics and Quality Control (QC)*

150 Participants with rs12913832:AA+AG genotypes were extracted from CanPath arrays (UKBB and
151 GSA). We retained only those individuals who self-reported as having European-related ancestry
152 and for which self-reported eye colour data were available. Individuals who self-reported amber
153 eyes were excluded due to the low frequency of individuals who reported this colour category.

154 Pre-imputation QC was performed as described in Lona-Durazo et al (2022) (11) . Post-imputation
155 QC began with the removal of all variants other than bi-allelic SNPs (e.g., insertion/deletions,
156 multi-allelic markers, and other types of polymorphisms) using bcftools (version 1.13) (15,16).
157 Additionally, bcftools was also used to ensure that only SNPs were represented in the imputed
158 genotype files, no insertion/deletions (indels) or other types of polymorphisms, as well as to
159 remove markers with low imputation accuracy (`INFO>0.8`) and duplicated markers. Variants
160 with strong deviations from Hardy-Weinberg equilibrium (HWE) (`--hwe 1E-6`) were filtered
161 from the imputed files using `vcftools` (version 0.1.16.) (17). Pre-imputation and post-
162 imputation QCs were performed separately on each array.

163 An initial principal component analysis (PCA) was performed using a pruned (`--`
164 `indeppairwise 50 1 0.1`) set of unimputed variants from both CanPath arrays projected
165 onto the 1000s Genomes Project (1KG) Phase 3 samples using PLINK (version 1.9) (18,19) to
166 identify any outliers that did not cluster within the European superpopulation of the 1KGP based
167 on the first three principal components. After removal of outliers, 1,412 samples were retained

168 from the GSA array and 1,312 from the UKBB array (for a total of 2,724 AA/AG individuals). A
169 second PCA was then performed for each array using PLINK (version 1.9) (18,19) and the first 10
170 principal components were retained as covariates and added to the phenotype files corresponding
171 to each array using R studio (version 4.2.0) (20).

172 *Genome-wide Association Study (GWAS)*

173 A genome-wide association study (GWAS) targeting eye colour was independently conducted on
174 individuals with AA+AG genotypes from each array, employing both linear and logistic mixed
175 models via SAIGE (version 1.0.5) (21). In the linear model, eye colour is coded as follows: blue=1,
176 green=2, hazel=3, and brown=4. The logistic model was coded as follows: blue=1 versus non-
177 blue=0 (brown and hazel). The linear model used for categorical eye colours followed an additive
178 framework, wherein the estimated effect size corresponds linearly to the count of effect alleles.

179 The GWAS model included covariates such as age, sex, and the first ten principal components as
180 fixed effects and a Genetic Relationship Matrix (GRM) as a random effect, alongside imputed
181 genotype data.

182 *Meta-analysis*

183 A meta-analysis was carried out using the beta coefficients and standard errors from the linear
184 GWAS results corresponding to the two distinct CanPath arrays, along with data from each logistic
185 GWAS using the software Metasoft (version 2.0.1) (22). Metasoft assesses input GWAS using two
186 types of models, 1) fixed effects (FE), most appropriate when effect sizes are assumed to be
187 consistent across studies, and 2) an enhanced random effects model (RE2), which is more suitable
188 when there is variability or heterogeneity between datasets (23).

189 Metasoft computes two key indicators of heterogeneity: Cochran's Q and I^2. The Q statistic reflects
190 the weighted squared differences between individual study effects and the combined effect
191 estimate; a P-value below 0.10 typically indicates significant heterogeneity. The I^2 metric (24,25)
192 estimates the proportion of variation across studies that exceeds what would be expected by
193 chance, calculated as $I^2 = 100\% \times (Q - df)/Q$.

194 When I^2 exceeds 50%, this may suggest that a random effects model is more appropriate than a
195 fixed effects model (26,27). In the absence of heterogeneity, both models are expected to produce
196 similar outcomes. Metasoft also generates Bayesian posterior probabilities indicating the
197 likelihood that an association exists in each study included in the meta-analysis (23).

198 *Conditional and Joint Analysis (COJO) Using GWAS Summary Statistics and Fine-Mapping*

199 The combined summary statistics from the linear GWAS meta-analysis were subsequently
200 processed using the COJO method included in the GCTA suite (28). COJO is used to determine
201 whether genome-wide significant signals (P-value<5E-8) identified in a GWAS (or meta-analysis)
202 are attributable to a single variant or represent multiple independent signals. The analysis was
203 based on fixed-effects (FE) model P-values from the meta-analysis and incorporated weighted
204 average effect allele frequencies derived from both contributing datasets. A linkage disequilibrium
205 (LD) reference was generated from the GSA array as it had the larger sample size of the two arrays
206 for AA and AG individuals.

207 Statistical fine-mapping was performed using LDSTORE2 (v2.0) and FINEMAP (v1.4) (29),
208 focusing on regions highlighted by the COJO analysis. This approach helps pinpoint variants most
209 likely to causally influence eye colour. FINEMAP uses a Bayesian model to integrate GWAS or

210 meta-analysis summary statistics with SNP LD data, producing posterior probabilities for putative
211 causal variants.

212 In the first step, LDSTORE2 was used to generate LD correlation matrices from the GSA
213 24v1MDP array data within 1 Mb windows (±500 kb from lead SNPs). Next, these LD matrices
214 were input into FINEMAP along with the defined genomic windows. The analysis assumed up to
215 10 causal variants per region, generating 95% credible sets – groups of SNPs with the highest
216 posterior probability of being causal. Each set is summarized by its posterior inclusion
217 probabilities and a k-value indicating the most probable number of causal variants.

218 *SNP Annotation*

219 To annotate the genome-wide significant variants (P<5E-08) identified through meta-analysis, we
220 employed SNPnexus, a web-based annotation tool (https://www.snp-nexus.org/v4/) (30–34).
221 Gene and variant classification were based on data from the UCSC Genome Browser and Ensembl
222 databases using the hg19 human genome build. If, through SNPNexus, we were able to determine
223 if a variant was listed in ClinVar (35), then the clinical annotation was also documented. The
224 potential functional impact of non-synonymous coding variants was evaluated using SIFT (36)
225 and PolyPhen (37) prediction scores, which provide qualitative outputs such as "probably
226 damaging", "possibly damaging/deleterious-low confidence", or "benign/tolerated". For non-
227 coding variants, we applied the CADD scoring system (38), which ranks variant deleteriousness
228 across the genome. A CADD score of 20 or higher suggests the variant is among the top 1% of the
229 most deleterious substitutions (34). Furthermore, we investigated protein-level expression of
230 implicated genes using STRING (39) and searched for eQTLs (expression Quantitative Trait Loci)
231 in GTEx (https://www.gtexportal.org/home/) (40).

232 *Replication in Independent Sample of Individuals of European Ancestry with Quantitative*
233 *Measures of Iris Colour*

234 The SNPs identified in the CanPath GWAS were replicated in a sample of individuals of European
235 ancestry living in Canada, for whom detailed eye colour quantitative measures based on high-
236 resolution iris pictures are available. The sample and the methods used to measure iris
237 pigmentation have been described in a previous paper (41). The sample includes 549 individuals
238 with ancestry from different regions in Europe. High-resolution iris pictures were taken using a
239 Miles Professional Iris Camera and quantitative measures of iris colour using the CIE 1976 L*a*b*
240 (CIELAB) colour space were obtained using a web-based application (41). The L* coordinate
241 represents the lightness dimension and ranges from 0 to 100, with 0 being black and 100 being
242 white. The a* and b* coordinates represent variation in colour, with negative values of a*
243 indicating green and positive values of a* indicating red, and negative values of b* indicating blue
244 and positive values of b* indicating yellow. Individuals with blue eye colour are characterized by
245 negative a* and b* values, and higher L* values than those observed for individuals with non-blue
246 eyes. However, there is considerably variation in L*, a* and b* dimensions within individuals
247 reporting "blue" or "brown" eye colour (41). In addition to L*, a* and b* values, the amount of
248 central heterochromia (i.e., differences in iris colour between the inner and outer portions of the
249 iris) was measured by a colour metric (ΔE) that represents differences in the average colour of the
250 ciliary zone (outer iris region) and the average colour of the pupillary zone Inner iris region). Two
251 different cameras were used during the study, and camera was used as a covariate in the statistical
252 analysis. Participants in the study also self-reported eye colour in 5 categories: 0=Light blue, Light
253 gray or light green; 1=Blue, gray or green, 2=Hazel or light brown, 3=Dark brown, 4=Brownish
254 black.

255 Genotyping was carried out with Illumina's Infinium Multi-Ethnic Global Array (MEGA) at the
256 Clinical Genomics Centre (Mount Sinai Hospital, Toronto, Ontario, Canada) using standard
257 protocols. QC steps were implemented to remove samples and markers, according to the following
258 criteria, Sample QC: 1) Removal of samples with missing call rates <0.9, 2) Removal of samples
259 that were outliers in Principal Component Analysis (PCA) plots, 3) Removal of samples with sex
260 discrepancies, 4) Removal of samples that were outliers for heterozygosity, and 5) Removal of
261 related individuals (pi-hat> 0.2). Marker QC: 1) Removal of markers with genotype call rate <0.95,
262 2/removal of markers with Hardy-Weinberg Es <10−6, 3) Removal of Insertion/Deletion (Indel)
263 markers, 4) Removal of markers with allele frequencies <0.01, 5) Removal of markers not present
264 in the 1000 Genomes reference panel, or that do not match on chromosome, position and alleles,
265 6) Removal of A/T or G/C SNPs with MAF >40% in the 1000 Genomes European reference
266 samples, and 7) Removal of SNPs with allele frequency differences >20% between the study
267 sample and the 1000 Genomes European reference sample. After performing the QC steps
268 described above, the samples were phased using the program SHAPEIT2 and imputed at the
269 Sanger Imputation Service, using the Positional Burrows-Wheeler Transform (PBWT) algorithm
270 (42), and the samples of the 1000 Genomes as reference haplotypes. All the markers included in
271 this analysis have imputation INFO scores higher than 0.9, indicating good imputation confidence.

272 Following the same approach used for the CanPath sample, individuals were stratified in two
273 categories based on *HERC2* rs12913832 genotype (GG and AA+AG). Association of the selected
274 markers with the quantitative traits were tested using linear regression with sex, camera, and 4
275 principal components as covariates. Age was not included as a covariate because there was a
276 narrow range in participant's age (18-35 years) and age did not have a significant effect on the
277 traits. Given that some quantitative traits (a*, b* and ΔE) showed non-normal distributions, an

278 inverse normal transformation of the residuals from a regression of a*, b* and ΔE on sex, camera
279 and 4 principal components was carried out, and these normalized values were used as dependent
280 variables in the linear regression. A linear model was also used to evaluate association of the
281 markers with the five self-reported eye colour categories (0=Light blue, Light gray or light green;
282 1=Blue, gray or green, 2=Hazel or light brown, 3=Dark brown, 4=Brownish black).

283 Results

284 *Preliminary Statistics and Quality Control (QC)*

285 After applying all QC filters, genotypes and self-reported eye colour phenotypes were retained
286 from a total of 2,724 AA+AG individuals (1,412 from the GSA array and 1,312 from the UKBB
287 array). Eye colour categories included blue, green, hazel and brown. Of the 2,724
288 rs12913832:AA+AG participants, 512 did not self-report brown or hazel eyes, representing
289 41.16%. As was reported in 2023, of the 2,757 rs12913832:GG participants for which self-
290 reported eye colour data were available, 904 did not report blue eyes. **Fig 4** illustrates eye colour
291 frequency stratified by genotype at the rs12913832 locus. In total, 5,481 CanPath participants are
292 represented when considering all rs12913832 genotypes combined.

293 **Fig 4**. **Eye colour distribution among 2,757 rs12913832:GG and 2,724 rs12913832:AA+AG Canadian**
294 **Partnership for Tomorrow's Health (CanPath) participants genotyped with the UKBB array and GSA array.**
295 Self-reported eye colour categories retained for investigation were described as blue, green, hazel, and brown.
296 Individuals with missing phenotype information or who self-reported their eye colour as amber were excluded.

297

298 After QC and imputation 8,822,293 variants were retained from AA+AG individuals genotyped
299 with the GSA array for further analysis and 7,104,980 variants were retained from AA+AG
300 individuals genotyped with the UKBB array.

301 *Genome-Wide Association Study (GWAS) and Meta-Analysis*

302 As was performed on the GG individuals in the 2023 study, two types of GWAS were performed
303 on AA+AG individuals from each array using SAIGE (v1.0.5) (21). A linear mixed model was
304 applied where eye colour was coded as blue=1, green=2, hazel=3, and brown=4 in accordance
305 with an additive genetic model, as well as a logistic model comparing blue versus non-blue (hazel
306 and brown) eye colour. The first ten genetic principal components generated during the QC stage
307 were included as covariates, along with age and sex.

308 The summary statistics (P-values, betas, and SE) from the linear and logistic models conducted in
309 each CanPath array GWAS were then integrated in a meta-analysis using Metasoft (v2.0.1) (22).
310 The resulting Manhattan plot from the linear model is illustrated in **Fig 5** (Supplementary **Figs S1**
311 and **S2** present the QQ plots for each meta-analysis and the Manhattan plot from the logistic
312 model). Significant peaks are observed in chromosomes 6 (*IRF4*), 9 (*TYRP1*) and 15 (*OCA2*).

313 **Fig 5**. **Manhattan plot of eye colour meta-analysis based on linear model GWAS in rs12913832:AA+AG**
314 **individuals (*N*=2,724) from the Canadian Partnership for Tomorrow's Health (CanPath) project genotyped**
315 **with the Axiom UKBB and GSA arrays.** GWAS were performed with SAIGE (21). Categorical eye colours coded
316 as blue=1, green=2, hazel=3, and brown=4. The meta-analysis was conducted with Metasoft (22). The black horizontal
317 line indicates the genome-wide threshold (P<5E-8). Significant peaks are observed in chromosomes 6 (*IRF4*), 9
318 (*TYRP1*) and 15 (*OCA2*).

319

320 *Comparing GWAS Lead SNPs from GG and AA+AG Individuals.*

321 **Table 1** illustrates the top SNPs from the meta-analysis according to the fixed effects (FE) linear
322 model in each significant genomic region. Several SNPs had evidence of heterogeneity among the
323 two studies (indicated by I^2 and Cochran's Q values in the meta-analysis), however, they all
324 retained genome-wide significant P-values in the random effects (RE2) model too, which considers

heterogeneity among studies. Given that heterogeneity between studies did not impact SNP significance, the fixed effects (FE) model was prioritized for downstream analyses. Please note that that in the 2023 publication (12) the effect alleles in the tables were incorrect, and a correction was submitted.

Table 1. Top SNP from each significant genomic region in meta-analysis of GG stratified individuals and AA+AG stratified individuals using a linear model (N_{GG}=2,747, N_{AA+AG}=2,724).

Marker ID	Chr	A1	A2	A1 Frequency	Gene	P-value (FE)	Beta	SE
GG Stratified Individuals								
rs35389	5	**G**	A	0.046	*SLC45A2*	1.18E-12	0.352	0.049
rs2733836	9	**T**	G	0.438	*TYRP1*	1.37E-13	0.133	0.018
rs1126809	11	A	**G**	0.272	*TYR*	4.58E-13	-0.179	0.019
rs12896399	14	T	**G**	0.401	*SLC24A4*	2.05E-35	-0.221	0.018
rs62075722	17	A	**G**	0.364	*TSPAN10*	6.36E-10	-0.116	0.019
AA+AG Stratified Individuals								
rs12203592	6	**T**	C	0.164	*IRF4*	3.13E-26	-0.318	0.030
rs7853779	9	A	**G**	0.389	*TYRP1*	3.12E-8	0.129	0.023
rs1800407	15	**T**	C	0.128	*OCA2*	4.10E-15	-0.292	0.037

Based on the categorical coding of the model (1=blue, 2=green, 3=hazel, 4=brown), a positive beta can be interpreted as an indication of eye colour "darkening" and a negative beta can be interpreted as an indication of eye colour "lightening" for the effect allele (A1). We highlight alleles increasing pigmentation in the rs12913832 GG (e.g. blue-eye) background with red font and alleles decreasing pigmentation in the rs12913832 AA+AG (e.g. non blue-eye) background in blue font.

Marker ID, reference SNP cluster ID;
Chr, chromosome;
A1, effect allele: the allele whose effects in relation to eye colour are being studied;
A2, the alternative allele;
A1 frequency, fraction of chromosomes in population that carry the effect allele;
Gene, the gene the SNP is located in, or the gene the SNP is nearest to;
Beta, per unit increase (if positive) or decrease (if negative) in the outcome;
P-value (FE), parameter of statistical significance used to determine the certainty of an association, based off the fixed effects meta-analysis model;
SE, standard error.

Conditional and Joint Analysis (COJO)

Results from the COJO analysis using GCTA (28) for the GG individuals were previously described in 2023 (12), but they are summarized in **Table 2** to easily compare the different signals

350 coming from each genetic background. The COJO analysis of the linear model conducted on
351 AA+AG individuals identified one significant genome-wide SNP driving the signals on
352 chromosomes 6 and 9, while two genome-wide significant SNPs were observed driving the signal
353 on chromosome 15. Among these, rs12203592 in *IRF4* is a well-established causal pigmentation
354 variant, previously associated with both skin and hair pigmentation, as well as melanoma
355 susceptibility (6,43,44). The SNP rs7853779 on chromosome 9, located near *TYRP1*, had prior
356 associations with pigmentation traits and melanoma susceptibility (1,45). Two SNPs on
357 chromosome 15, rs1800407 and rs4778138, both situated in or near the *OCA2* gene, have been
358 linked to variation in eye colour in earlier research (46–48).

359 **Table 2. Independent signals identified by COJO (28) for genome-wide significant peaks illustrated in the meta-**
360 **analysis.**

CHR	Marker ID	POS	A1	A2	Beta	SE	P-value (FE)	Joint beta	Joint SE	Joint P-value	LD (r^2)	Nearest gene	MAF
GG individuals													
5	rs35389	33954880	**G**	A	0.352	0.049	1.180E-12	0.352	0.051	4.550E-12	0	*SLC45A2*	0.046
9	rs2733836	12692252	**T**	G	0.133	0.018	1.370E-13	0.133	0.018	6.410E-13	0	*TYRP1*	0.438
11	rs1126809	89017961	A	**G**	-0.179	0.02	1.200E-19	-0.179	0.021	3.950E-18	0	*TYR*	0.272
14	rs12896399	92773663	T	**G**	-0.221	0.018	2.050E-35	-0.221	0.019	4.020E-30	0	*SLC24A4*	0.401
17	rs62075722	79611271	A	**G**	-0.116	0.019	6.360E-10	-0.116	0.019	1.390E-09	0	*TSPAN10*	0.365
AA+AG individuals													
6	rs12203592	396321	**T**	C	-0.318	0.03	3.130E-26	-0.318	0.031	2.170E-24	0	*IRF4*	0.164
9	rs7853779	12612893	A	**G**	0.129	0.023	3.120E-08	0.129	0.024	4.340E-08	0	*TYRP1*	0.389
15	rs1800407	28230318	**T**	C	-0.292	0.037	4.100E-15	-0.274	0.038	8.150E-13	0.082	*OCA2*	0.128
15	rs4778138	28335820	G	**A**	0.179	0.027	3.040E-11	0.164	0.028	2.710E-09	0	*OCA2*	0.273

361 Alleles increasing pigmentation in individuals with rs12913832 GG background are highlighted in red, and alleles
362 decreasing pigmentation in individuals with rs12913832 AA+AG background are highlighted in blue.
363 **Chr**, chromosome;
364 **Marker ID**, Reference SNP cluster ID;
365 **POS**, Physical position (hg19/GRCh37);
366 **A1**, effect allele: the allele whose effects in relation to eye colour are being studied;
367 **A2**, alternative allele;
368 **Beta**, per unit increase (if positive) or decrease (if negative) in the outcome;
369 **SE**, Standard error;
370 **P-value (FE)**, P-value, parameter of statistical significance used to determine the certainty of an association
371 based off the fixed effects meta-analysis model;

372 **Joint beta, SE** and **P-value**, effect size, standard error and P-value from a joint analysis of all the selected
373 SNPs;
374 **LD(r^2)**, Linkage disequilibrium correlation between the SNP i and SNP i + 1 for the SNPs on the list;
375 **Nearest gene**, the gene the SNP is located in, or the gene the SNP is nearest to;
376 **MAF**, minor allele frequency, fraction of chromosomes in population that carry the minor allele (frequency
377 <0.5).

378 *Fine-mapping*

379 Fine-mapping results are represented in **Table 3**. Only SNPs with $\log_{10}$ of Bayes Factor ($\log_{10}$BF)
380 greater than 2 have been included in the table. For some regions (*TYRP1*, *TYR*, *SLC24A4* and
381 *TSPAN10* for the GG stratified individuals), the *k* value (most likely number of causal SNPs) is
382 greater than one, however, only SNPs within one credible set surpassed a $\log_{10}$BF>2. Thus,
383 although FINEMAP suggests the presence of multiple signals in some instances, the markers in
384 the additional credible sets did not always have sufficiently high $\log_{10}$BF, so we focused on the
385 markers that surpassed the thresholds for "considerable evidence" as recommended by the
386 FINEMAP documentation (29).

387 For the GG stratified individuals, the markers with the top PIP in the FINEMAP analysis coincide
388 with the lead SNPs identified in the COJO analysis: *SLC45A2* rs35389, *TYRP1* rs2733836, *TYR*
389 rs1126809, *SLC24A4* rs12896399 (there is another marker with the same PIP in this region,
390 rs12896471), and *TSPAN10* rs62075722. The non-synonymous variant rs1126809 is highlighted
391 as the causal variant in the *TYR* region, with PIP=0.960. All the other variants have PIP values
392 lower than 0.5.

393 For the AA+AG stratified individuals, the FINEMAP analysis identifies a single credible set in
394 the *IRF4* and *TYRP1* regions. The marker rs12203592 is highlighted as the putative causal SNP
395 within the *IRF4* region, with PIP=1.00. In contrast, there are several markers in high LD in the
396 credible set of the *TYRP1* region, with PIP values<0.10. The top SNPs identified by FINEMAP

397 coincide with the lead signals in the COJO analysis. Finally, the FINEMAP analysis identifies 3
398 independent credible sets in the *OCA2/HERC2* region. In the first credible set, FINEMAP
399 identifies rs1800407 as the causal variant, with PIP=0.997. In the second credible set, FINEMAP
400 highlights two SNPs: rs7164220 (PIP=0.472) and rs4778138 (PIP=0.341). The latter was the lead
401 SNP identified in the COJO analysis. The variant with the top PIP in the third credible set is
402 rs79380392 (PIP=0.135), which is located within the *HERC2* gene. In contrast to FINEMAP,
403 which points to three independent causal sets, COJO only identified two independent regions in
404 the *OCA2/HERC2* region. **Fig S3** illustrates LD matrices for SNPs in the *TYRP1* region and
405 *HERC2/OCA2* region. **Fig S4** illustrates examples of regional plots for signals represented in
406 **Table 3**.

407 **Table 3. Fine-mapping results. Fine-mapping was performed using 1Mb windows (500 kb on each side of**
408 **significant SNP).**

Marker ID	Chr	A1	A2	Nearest gene	Beta	SE	P-value	MAF	PIP	*k*
GG individuals										
rs35389	5	**G**	A	*SLC45A2*	0.352	0.049	1.180E-12	0.046	0.408	1
rs35395	5	T	**C**	*SLC45A2*	-0.357	0.050	1.475E-12	0.034	0.335	
rs16891982	5	C	**G**	*SLC45A2*	-0.328	0.047	3.481E-12	0.038	0.155	
rs35407	5	A	**G**	*SLC45A2*	-0.345	0.051	9.106E-12	0.034	0.066	
rs2733836	9	**T**	G	*TYRP1*	0.133	0.018	1.368E-13	0.438	0.068	2
rs1126809	11	A	**G**	*TYR*	-0.179	0.020	1.196E-19	0.272	0.960	2
rs12896399	14	T	**G**	*SLC24A4*	-0.221	0.018	2.052E-35	0.401	0.154	2
rs12896471	14	C	**G**	*SLC24A4*	-0.221	0.018	2.059E-35	0.425	0.154	
rs56137766	14	T	**A**	*SLC24A4*	-0.222	0.018	2.089E-35	0.430	0.136	
rs35983729	14	T	**G**	*SLC24A4*	-0.223	0.018	2.205E-35	0.432	0.128	
rs4144266	14	A	**G**	*SLC24A4*	-0.223	0.018	2.367E-35	0.432	0.119	
rs4904866	14	C	**T**	*SLC24A4*	-0.221	0.018	3.456E-35	0.425	0.089	
rs746586	14	C	**T**	*SLC24A4*	-0.221	0.018	3.994E-35	0.426	0.075	
rs12883151	14	A	**G**	*SLC24A4*	-0.222	0.018	3.960E-35	0.432	0.070	
rs62075722	17	A	**G**	*TSPAN10*	-0.116	0.019	6.357E-10	0.365	0.174	2
rs6420484	17	A	**G**	*TSPAN10*	-0.116	0.019	6.602E-10	0.366	0.169	
rs34635363	17	**G**	A	*NPLOC4*	0.114	0.019	1.528E-09	0.364	0.079	
AA+AG individuals										

rs12203592	6	**T**	C	*IRF4*	-0.318	0.030	3.134E-26	0.164	1.000	1
rs7853779	9	A	**G**	*TYRP1*	0.129	0.023	3.123E-08	0.389	0.074	1
rs10960723	9	**C**	A	*TYRP1*	-0.128	0.023	3.674E-08	0.391	0.064	
rs713596	9	**A**	G	*TYRP1*	-0.128	0.023	3.749E-08	0.391	0.063	
rs2762457	9	**G**	T	*TYRP1*	-0.126	0.023	4.855E-08	0.402	0.050	
rs2762460	9	**T**	C	*TYRP1*	-0.125	0.023	4.975E-08	0.409	0.049	
rs2733831	9	**A**	G	*TYRP1*	-0.125	0.023	5.343E-08	0.408	0.046	
rs1800407	15	**T**	C	*OCA2*	-0.292	0.037	4.100E-15	0.128	0.997	3
rs7164220	15	**C**	A	*OCA2*	-0.179	0.030	1.545E-09	0.220	0.472	
rs4778138	15	G	**A**	*OCA2*	0.179	0.027	3.040E-11	0.273	0.341	
rs79380392	15	C	**T**	*HERC2*	0.249	0.055	5.354E-06	0.049	0.135	

SNPs with a log10 of Bayes Factor greater than 2 are recorded in the table. Alleles increasing pigmentation in individuals with rs12913832 GG background are highlighted in red, and alleles decreasing pigmentation in individuals with rs12913832 AA+AG background are highlighted in blue. A thick border indicates separation of chromosomal fine-mapped regions, a thin border indicates separation of credible sets (the collection of variants that together have a 95% probability of containing the true causal SNP).

Marker ID, reference SNP cluster ID;
Chr, chromosome;
A1, effect allele: the allele whose effects in relation to eye colour are being studied;
A2, alternative allele;
Nearest gene, the gene the SNP is located in, or the gene the SNP is nearest to;
Beta, per unit increase (if positive) or decrease (if negative) in the outcome
SE, standard error;
P-value, parameter of statistical significance used to determine the certainty of an association;
MAF, minor allele frequency;
PIP, posterior inclusion probability;
k, most likely number of causal SNPs.

SNP Annotations

Beyond the rs12913832 genotype, acting as a genetic background in this study, three other SNPs within *OCA2* and one within *HERC2* were nominated as candidate causal loci in AA+AG stratified individuals. One of them is a non-synonymous SNP (rs1800407) and three are intronic variants (rs4778138, rs7164220, rs79380392). These results are similar to the conditional analysis with GCTA-COJO, where the program picked up multiple independent signals coming from this region. We annotated the putative regulatory function of the fine-mapped SNPs with a $\log_{10}$ of Bayes Factor greater than 2 using diverse databases (e.g. ENCODE, Roadmap Epigenomics Project) encompassed within SNPnexus (30–34). A summary of the output can be found in Supplementary **Table S1**. All SNPs in the table were identified as either overlapping with a protein-coding gene

435 or were located near a protein coding gene by SNPNexus (30–34). Annotation details for the lead

436 SNPs identified by COJO are described in **Table 4** below.

437 **Table 4. Annotation summary of lead SNPs.**

GG Stratified				
SNP	**Overlapped or nearest gene**	**Protein Consequence**	**Eye Colour Darkening Allele**	**Annotation**
rs35389	*SLC45A2*, overlapped	Protein coding; Intronic	**G**	eQTL in Thyroid (P=8.3E-5) (GTEx database). Low CADD PHRED score of 0.530. Confirmed associations with skin pigmentation, eye colour and hair colour according to GAD. Classified as Benign by ClinVar.
rs2733836	*TYRP1*, overlapped	Protein coding; Intronic	**T**	eQTL for many brain tissues (GTEx database). Low CADD score (2.549). No confirmed associations (unconfirmed association with Type 2 Diabetes).
rs1126809	*TYR*, overlapped	Protein coding; Non-synonymous	**G**	eQTL in Skin – Sun Exposed (Lower leg) (P=2.35E-5) and Testis (P=1.5E-4) (GTEx database). Reported as "Deleterious" by SIFT (Score=0.30) and "Probably Damaging" by PolyPhen (Score=0.994). Confirmed associations with multiple phenotypes including many pigmentation related phenotypes such as type IB (yellow) oculocutaneous albinism and type I-A (tyrosinase negative) oculocutaneous albinism according to GAD. Previously reported as significant in GWAS of melanoma, squamous cell carcinoma, optic disc size, basal cell carcinoma, hair colour, eye colour, sunburn, tanning response and vitiligo based on SNPnexus' GWAS catalogue. This SNP has conflicting interpretations of pathogenicity in ClinVar. Many reports have associated the variant with conditions like albinism and general abnormalities in pigmentation-related metabolic processes (49–51).
rs12896399	*SLC24A4*, nearest downstream gene	Protein coding	**G**	eQTL in Skin – Not Sun Exposed (Suprapubic) (P=1.1E-8), Skin – Sun Exposed (Lower leg) (P=5.0E-6), Whole blood (P=2.1E-5) and Muscle – Skeletal (P=4.1E-5) (GTEx database). CADD PHRED score of 5.795. Previously reported as significant in GWAS of hair colour, hair morphology traits and eye colour (blue vs. green eyes) based on SNPnexus' GWAS catalogue. The G allele increases expression and is associated with darker skin. This means that increased expression of this gene may counteract the effects of lower *OCA2* expression in GG individuals.
rs62075722	*TSPAN10*, overlapped	Non-coding, intronic	**G**	eQTL in Skin – Not Sun Exposed (Suprapubic) (P=4.1E-18) and Skin – Sun Exposed (Lower leg) (P=6.9E-16) (GTEx database). CADD PHRED score of 6.286. Confirmed associations with coronary artery disease, Parkinson's disease, and eye colour. Previously reported as significant in GWAS of brown vs. black hair and corneal astigmatism based on SNPnexus' GWAS catalogue. A

				G allele is associated with increased expression and darker iris colour.
AA+AG Stratified				
SNP	**Overlapped or nearest gene**	**Protein Consequence**	**Eye Colour Lightening Allele**	**Annotation**
rs12203592	*IRF4*, overlapped	Protein coding; Intronic	**T**	eQTL in Skin – Sun exposed (Lower leg) (P=6.5E-7) and many other tissues in GTEx. CADD PHRED score of 11.780. Confirmed associations with hair colour and eye colour in GAD. Previously reported as significant in GWAS of facial pigmentation, squamous cell carcinoma, tanning, sunburns, non-melanoma skin cancer, hir colour, eye colour and freckling based on SNPnexus' GWAS catalogue. Marked as "Affects" by ClinVar, reaffirming its influence on pigmentation phenotypes. The T allele decreases expression of *IRF4* and is associated with lighter iris colour. This means that lower expression of *IRF4* counteracts the effect of AA+AG, promoting lighter eye colours.
rs7853779	*TYRP1*, downstream	Protein coding	**G**	CADD PHRED score of 7.814. No confirmed associations in GAD, SNPnexus' GWAS catalogue or ClinVar.
rs1800407	*OCA2*, overlapped	Protein coding; Non-synonymous	**T**	Reported as "Deleterious" by SIFT (Score=0.20) and "Probably Damaging" by PolyPhen (Score=0.986). Confirmed associations with hair colour, eye colour Parkinson's disease and socioeconomic factors in GAD. Previously reported as significant in GWAS of squamous cell carcinoma, cataracts, skin pigmentation, eye colour, macular thickness and tanning response based on SNPnexus' GWAS catalogue. Classified as benign or likely benign in ClinVar, it is related to both oculocutaneous albinism, pigmentation variation and melanoma susceptibility. Variant has a haploinsufficiency (HI) score of 30, and homozygous mutations lead to oculocutaneous albinism type 2 (4,52,53).
rs4778138	*OCA2,* overlapped	Protein coding; Intronic	**A**	eQTL in Whole Blood (P=5.8E-20), Nerve-Tibial (P=2.9E-7) and Artery-Aorta (P=8.2E-5) (GTEx database). CADD PHRED score of 1.982. Confirmed associations with hair colour, eye colour, Parkinson's disease and socioeconomic factors in GAD. Previously reported as significant in GAS of squamous cell carcinoma and melanoma.

SNPs were annotated using SNPnexus (https://www.snp-nexus.org/v4/) (30–34) and GTEx (https://gtexportal.org/home/) (40). Note that GTEx does not have iridial tissue information, so focus was placed on skin tissue information (sun exposed and not-exposed). Integrated in SNPnexus are additional annotation tools. We utilized CADD PHRED scores (38), SIFT (36) and PolyPhen (37) to assess the deleteriousness of a variant, GAD (53) and ClinVar (35) reports to learn about associated diseases and other phenotypes.

GTEx, Genotype-Tissue Expression Project;
eQTL, expression Quantitative Trait Locus;
sQTL, splicing Quantitative Trait Locus;
CADD PHRED, Combined Annotation Dependent Depletion scaled score (represents the probability of a variant being deleterious relative to all other possible variants);
SIFT, Sorting Intolerant from Tolerant, predicts whether an amino acid substitution affects protein function based on sequence homology (applied to naturally occurring nonsynonymous polymorphisms and laboratory induced missence mutations);

451 **PolyPhen,** Polymorphism Phenotyping prediction tool used to asses the potential impact of an amino acid substitution
452 on a protein's structure and function
453 **GAD,** Genetic Association Database;
454 **ClinVar,** public archive of reports of human variations classified for diseases and drug responses, with supporting
455 evidence.

456

457 Supplementary **Fig S5** and **Table S2** illustrate the STRING network and summarize the functional
458 enrichments within the networks. There are slight differences in the biological processes enriched for
459 genes which were associated with GG individuals versus AA+AG individuals. In GG individuals,
460 genes can be seen enriched for melanin biosynthetic processes and developmental pigmentation, while
461 in AA+AG individuals, genes were enriched in melanosome membrane. Interestingly, under the
462 "Human Phenotype (Monarch)" STRING domain, the genes associated with the AA+AG, or brown
463 eye background, the phenotype with the strongest signal (FDR=1.96E-5) was blue irides, or a markedly
464 blue colouration of the iris.

465 *Replication in European ancestry cohort with quantitative iris pigmentation measures*

466 Replication results for the six different analyses are reported in Supplementary **Table S3**
467 (Supplementary **Fig S6** illustrates the distributions of rs12913832 genotype across self-reported
468 eye colour category for this replication cohort). Of the five independent signals from GG
469 individuals, three markers survived Bonferroni correction in at least one quantitative measure (*TYR*
470 rs1126809 in the L and b* dimensions, *SLC24A4* rs12896399 in the b* dimension, and *TSPAN10*
471 rs62075722 in the b* dimension). Of the four independent signals from AA+AG individuals, three
472 markers passed Bonferroni correction in multiple dimensions (*IRF4* rs12203592 in the L, a*, b*
473 and ΔE dimensions, *OCA2* rs1800407 in the L and a* dimensions, and *OCA2* rs4778138 in the L
474 and a* dimensions). Two markers demonstrated nominal significance in other quantitative

475 measurements (*SLC24A4* rs12896399 for the GG stratified sample and *TYRP1* rs7853779 for the
476 AA+AG stratified sample,). The effects appear strongest in the b* dimension (Blue [-] to yellow
477 [+]) for GG stratified individuals and in the L dimension (Lightness dimension: 0 (black) to 100
478 (white)) for AA+AG stratified individuals. We also conducted a linear regression using the self-
479 reported eye colour categories in the independent replication dataset, and the results were
480 concordant with the quantitative measures. In GG stratified individuals, *TYR* rs1126809 was
481 nominally significant, and in AA+AG stratified individuals, all markers, with the exception of
482 *TYRP1* rs7853779, survived Bonferroni correction. Importantly, rs12913832 was also assessed in
483 the replication dataset in AA+AG stratified individuals. This marker illustrated significance in the
484 L, a* and ΔE dimensions, as well as the linear model – indicative that eye colour is dosage
485 dependent with respect to genotype at rs12913832. In particular, rs12913832 is very strongly
486 associated with ΔE (p=2.97E-07), suggesting that having one copy of the G allele increases the
487 likelihood of central heterochromia (e.g. differences in iris colour between the inner and outer
488 rings of the pupil).

489 Discussion

490 This study expands upon previous work investigating genetic modifiers of iris pigmentation by
491 analyzing individuals whose self-reported eye colour does not align with the expected phenotype
492 based on their *HERC2* rs12913832 genotype. Specifically, we focused on those with a
493 homozygous derived genotype (rs12913832:GG), which typically predicts blue eyes, and
494 heterozygous or homozygous ancestral genotypes (rs12913832:AG or AA), which are more often
495 associated with brown eyes. Despite the heavy reliance on this single SNP in forensic and
496 anthropological models, exceptions in phenotype suggest additional modifiers. Our genome-wide

497 approach successfully identified several SNPs that modify eye colour in these "non-concordant"
498 individuals and replicated key findings using high-resolution quantitative iris colour data in an
499 independent European sample.

500 Through our GWAS meta-analyses and fine-mapping analyses we identified loci in known
501 pigmentation-associated regions including *SLC45A2*, *TYRP1*, *TYR*, *SLC24A4*, and *TSPAN10* in
502 the GG group (where blue eyes is expected) and *IRF4*, *TYRP1*, and *OCA2* in the AA+AG group
503 (where brown eyes is expected).

504 *Genotype-stratified Associations with Eye Colour in Recent Studies*

505 In 2020, Meyer and colleagues (10) conducted a candidate gene study investigating the association
506 between brown eye colour in rs12913832:GG individuals and SNPs in previously known
507 pigmentation genes (*TYR*, *TYRP1* and *SLC24A4*). The group identified four variants to be the most
508 promising candidates for the explanation of the genotype-phenotype discordances observed in GG
509 individuals: *TYRP1* rs35866166:C, *TYRP1* rs62538956:C, *SLC24A4* rs1289469:C, and *TYR*
510 rs1126809:G. Of these four markers, three were replicated in our fine-map analyses: *TYRP1*
511 rs35866166:C, *TYRP1* rs62538956:C and *TYR* rs1126809:G. The *TYR* rs1126809 SNP is
512 represented in **Table 3** and has a high PIP of 0.96, strongly suggesting it is the causal variant, with
513 the G allele strongly associated with darker iris pigmentation in rs12913832:GG individuals. The
514 *TYRP1* rs35866166 and *TYRP1* rs62538956 markers are present in our 95% credible set, however,
515 they each had very low PIP values (0.001 and 0.003 respectively) and our fine-mapping analysis
516 prioritize a different variant within this region (rs2733836). That being said, we did find that the
517 effect direction was the same in our study, with the rs35866166:C and rs62538956:C alleles
518 corresponding to positive beta values (in the direction of eye colour darkening with our linear

519 model). The authors did note that none of the variants identified in the candidate gene study
520 showed statistically significant associations with eye colour after traditional multiple testing
521 correction, but in our GWAS we found rs1126809 and rs62538956 to surpass genome-wide
522 significance (P=1.200E-19 and P=5.846E-9 respectively), and rs35866166 to be nominally
523 significant (P=2.950E-6). Three years later, in 2023, Salvo and colleagues (54) performed an
524 additional candidate gene study investigating the association between variants in the
525 *HERC2/OCA2* region and blue eye colour in rs12913832:AA+AG individuals. They identified
526 five novel variants to be the most promising candidates for explanation of blue eye colour in
527 individuals with a brown-eye background: rs74409036:A (*OCA2*), rs78544415:T (*OCA2*),
528 rs72714116:T (*OCA2*), rs191109490:C (*HERC2*) and rs551217952:C (*HERC2*). None of these
529 SNPs are genome-wide significant in our analysis, and our analysis points to different markers
530 within this region. We describe in more detail the main regions and polymorphisms identified in
531 our stratified analysis below.

532 *Eye Colour Darkening in GG Stratified Individuals*

533 ***SLC24A4*** and ***SLC45A2***: *SLC24A4* encodes a sodium/calcium/potassium exchange protein
534 (GeneCards, 2025). Variants in this gene have been previously implicated in skin, hair and eye
535 pigmentation, although its precise function in melanogenesis remains unclear (55,56). Our lead
536 SNP in this region is rs12896399, which is situated approximately 15,000 base pairs upstream of
537 the *SLC24A4* transcription start site and has been included in established eye colour prediction
538 tools, including IrisPlex (57–59) and Snipper 3.5 (60). The T allele of rs12896399 is associated
539 with blue eye colour (59), a relationship also observed in this study. In our preliminary 2023
540 investigation of rs12913832:GG individuals (12), we observed strong linkage disequilibrium

541 among SNPs within the *SLC24A4* credible set, with pairwise comparisons yielding r^2 values
542 greater than 0.9. eQTL data in GTEx indicates that a G allele at rs12896399 is associated with
543 increased *SLC24A4* expression, suggesting that higher expression of this transporter is able to
544 compensate for the reduced expression of *OCA2* in rs12913832:GG individuals. Our results point
545 to rs12896399:G, or closely linked variants, as strong candidates for darker eye colours in GG
546 individuals.

547 *SLC45A2* encodes another transmembrane transporter involved in melanin biosynthesis. This gene
548 is expressed in melanoma cell lines and is implicated in oculocutaneous albinism type 4. Its
549 polymorphisms are most commonly associated with variation in skin and hair pigmentation (61).
550 The *SLC45A2* variant rs16891982 is included in both the IrisPlex (57–59) and The Snipper 3.5
551 (60) models, and appeared in our COJO and fine-mapping analyses of rs12913832:GG individuals.
552 However, rs16891982 was not the top candidate within the credible set (PIP=0.155). Instead,
553 rs35389 emerged as the most likely causal variant, with a PIP of 0.408. These two SNPs are in
554 linkage disequilibrium, as confirmed by LDpair (D'=1.0, R^2=0.6964, χ^2=700.577, P<0.001 with
555 EUR reference) (Machiela & Chanock, 2015). While rs35389 has not been previously associated
556 with eye colour, it has been evaluated as an ancestry informative marker (62) and has shown
557 significant associations with methylation patterns in the context of population stratification (63).

558 ***TYRP1***: Notably, *TYRP1* emerged as a consistent locus across both genetic backgrounds. *TYRP1*
559 encodes the tyrosinase-related protein 1, which is a member of the tyrosinase family and is
560 involved in the synthesis of eumelanin (64). Specifically, it is responsible for the stabilization of
561 tyrosinase, which is an important enzyme in the initiation of melanin synthesis (65) and contributes
562 to melanosomal structure and maturation (66). *TYRP1* is primarily expressed in melanocytes and

563 the retinal pigment epithelium (RPE) (67). Only one marker, rs2733836, illustrated association in
564 the GG background. There are no *TYRP1* variants currently included in eye colour prediction tools
565 specifically, although rs683 appears for the prediction of both hair and eye colour (57–59). In
566 2009, Liu et al. (55) compared predictive power of a six-SNP model (*HERC2* rs12913832, *OCA2*
567 rs1800407, *SLC24A4* rs12896399, *SLC45A2* rs16891982, *TYR* rs1393350, and *IRF4*
568 rs122203592) with a 15 SNP model where two *TYRP1* variants were included (rs1408799 and
569 rs683) and found that the inclusion of the additional SNPs provided very minimal additive effects
570 and did not significantly improve predictive accuracy. rs2733836 (associated in this study),
571 rs1408799 and rs683 are in linkage disequilibrium (D'>0.7) (Supplementary **Fig S3**), yet
572 rs2733836 has not previously been associated with eye colour. Not much is known of the variant's
573 consequence apart from the fact that it is located 2 kilobases upstream of *TYRP1*; the variant itself
574 is not a protein coding variant and is therefore not expressed, but the gene is, and can be found
575 most highly expressed in skin (30–34).

576 ***TYR***: The primary *TYR* variant identified in our study was rs1126809, a coding SNP in exon 4 of
577 *TYR,* and most likely causal variant in this region (PIP=0.960). The alternative allele (A) induces
578 a missense mutation that renders the tyrosinase enzyme thermosensitive and less efficient,
579 reducing melanin synthesis (68,69). Given that *TYR* is the rate-limiting enzyme in melanogenesis
580 (70,71), individuals carrying the A allele are expected to exhibit lighter pigmentation. Our results
581 indicate that the presence of the rs1126809 G allele can oppose/counteract the effects of reduced
582 *OCA2* expression driven by the rs12913832:GG genotype

583 There is currently only one *TYR* variant included in eye colour phenotyping models: rs1393350
584 (57–60,72,72–76). rs1393350 and rs1126809 are in near-perfect linkage disequilibrium (D'=1.0,

585 R^2=0.9532, χ^2=958.8784, P<0.001, with EUR reference), as reported by LDpair (77). Our findings
586 echo that of Meyer et al., 2020 (10) and raise the possibility that rs1126809 may be a more
587 biologically informative marker than rs1393350. Both variants reached genome-wide significance
588 in our GWAS and were identified in fine-mapping analyses; however, rs1393350 demonstrated a
589 notably low posterior inclusion probability (PIP=0.017) despite its strong P-value (P=5.81E-18),
590 suggesting it may not be the causal variant but rather tags a signal in high LD. Importantly, while
591 rs1393350 is known to associate with iris colour (with the A allele enriched in green eyes and the
592 G allele in blue) (56), its moderating function remains poorly defined. In contrast, rs1126809 is a
593 functional missense variant that reduces tyrosinase enzymatic activity through thermosensitivity,
594 directly impacting melanin production.

595 ***TSPAN10***: *TSPAN10*, or tetraspanin-10, has recently been associated with both quantitative and
596 categorical measurements of eye colour (1,63). It is known to be expressed in the eye, including
597 iris, ciliary body and RPE (78), however, its specific contribution to pigmentation variation has
598 yet to be fully identified. There are no *TSPAN10* (or *NPLOC4*) variants currently included in any
599 DNA phenotyping tools. In our analysis, rs62075722 illustrated the highest PIP (0.174) of the
600 credible set and is an eQTL in GTEx (40), although it should be noted that a PIP of 0.174 is low
601 and high LD among variants in the credible set makes it difficult to identify a singular putative
602 causal variant. The expression data indicates that a G allele at this locus is associated with
603 increased expression and a darker iris colour in GG stratified individuals. This is concordant with
604 the other two variants identified by FINEMAP, rs6420284 and rs34635363, which are also
605 significantly expressed in skin tissue in GTEx. Both the rs6420284:G allele and the rs34635363:G
606 allele are associated with increased expression and darker iris pigmentation.

607 *Eye Colour Lightening in AA+AG Stratified Individuals*

608 ***IRF4***: The transcription factor *IRF4* has been shown to enhance *TYR* gene expression through
609 interaction with MITF, a key regulator in melanocyte development (6). One specific *IRF4* variant,
610 rs12203592, has been demonstrated to influence *IRF4* expression in melanocytes (79), and is
611 included in the IrisPlex prediction panel (57,58). In the present study, we detected a significant
612 association between eye colour and rs12203592 in individuals carrying the AA or AG genotypes
613 at rs12913832. Prior research also suggests that interactions between *IRF4* and *OCA2* may
614 modulate eye pigmentation. For instance, both Laino et al. (80) and Lona-Durazo et al. (11)
615 reported that individuals with rs12913832 AA or AG genotypes – generally linked to darker eye
616 colour – can still present with blue eyes if they carry one or two copies of the rs12203592:T allele,
617 which has been associated with lighter pigmentation (80). The T allele at rs12203592 leads to
618 decreased *IRF4* expression (40). Phenotypically, carriers of the T allele tend to exhibit lighter eye
619 colour (e.g., blue or green), particularly among those with the rs12913832:AA or AG genotypes,
620 where the effect of the T allele can partially override the darker eye phenotype typically associated
621 with these genotypes.

622 ***TYRP1***: The observation of *TYRP1* variants (rs7853779, rs10960723, rs713596, rs276257,
623 rs2762460, rs2733831) in the AA+AG genetic background is both interesting, and more difficult
624 to interpret as there are no other studies, candidate gene or otherwise, which have investigated the
625 role of *TYRP1* SNPs in individuals with an rs12913832:AA+AG genetic background. Of the six
626 *TYRP1* variants with a $\log10_{BF}>2$, only three have published associations in scientific literature.
627 rs2762460 has been previously associated with breast cancer (81), and epidemiological studies
628 have proposed that melanoma and breast cancer may share genetic factors (82,83), positioning

629 *TYRP1* as a pleiotropic candidate gene. rs2762457 was used as a tag SNP in a candidate gene study
630 of individuals with a blue eye genetic background, however, it was determined not to be associated
631 with eye colour, but rather in strong LD with other SNPs associated with eye colour (10). Of note,
632 our study indicates that variants within this gene, which is known to play an important role in
633 melanin synthesis, may be responsible for darker eye colours in rs12913832:GG individuals and
634 lighter eye colours in rs12913832:AA+AG individuals.

635 It is also interesting to note that the *TYRP1* variant identified in the GG genetic background is in
636 high LD with the *TYRP1* variants identified in the AA+AG genetic background (all pairwise
637 relationships have a D'>0.9 and R^2>0.69, as confirmed by LDmatrix (with EUR reference) (77)).
638 Variants in this region may make good candidates for SNP-SNP interaction tests in future studies,
639 however, it would be important to consider that formal tests of epistatic interactions often suffer
640 from statistical power and multiple hypothesis testing limitations.

641 ***HERC2/OCA2***: *OCA2* encodes the P protein, a critical transmembrane component of the
642 melanosome that plays a vital role in melanin biosynthesis and pigmentation (84). In addition to
643 transporting melanosomal content, the P protein helps regulate melanosomal pH, which in turn
644 influences both the amount and quality of melanin deposited in melanocytes (5,7). Dysfunction or
645 loss of this protein, observed in both mice and humans, results in oculocutaneous albinism,
646 underscoring its essential function in pigmentation pathways (7,85). Regulation of *OCA2*
647 expression occurs via a long-range enhancer located within intron 86 of the neighbouring *HERC2*
648 gene. Reduced expression of *OCA2*, mediated by variation in this enhancer, disrupts melanosome
649 maturation and limits the effective processing and transport of melanin. This results in lighter
650 pigmentation (such as that observed in blue eyes) due to compromised P protein availability (5).

651 In our study, multiple *HERC2/OCA2* variants were observed to be associated with eye colour
652 variation in AA+AG individuals. By applying a Bayesian fine-map approach we found that the
653 most likely number of causal variants for this genomic region was three, for which three candidates
654 were located in *OCA2* (rs1800407, rs4778138/rs7164220) and one candidate in *HERC2*
655 (rs79380392). rs1800407 is a missense variant currently included in forensic phenotyping tools
656 (57–59). The rs1800407:T allele encodes the amino acid glutamate and having just one copy of
657 the T allele is associated with an increased probability of having green and hazel eye colours in
658 European populations (52,53). Given this, the presence of the T allele at rs1800407 in our study
659 may help explain the elevated frequency of self-reported green and hazel eye colours by
660 individuals with rs12913832:AA or AG genotypes (**Fig 4**). With a PIP of 0.997, and the only SNP
661 with a $\log_{10}$BF>2 in this credible set, rs1800407 appears to be the first of three causal SNPs in this
662 region. In the second credible set (rs7164220 and rs4778138), it is difficult to parse which variant
663 is the causal variant. rs4778138 was identified as an independent signal in the COJO analysis,
664 however, rs7164220 possesses the slightly higher PIP in the FINEMAP analysis (PIP=0.472).
665 rs4778138 has been previously identified as being involved in a three-SNP haplotype system
666 (rs4778138, rs4778241, and rs7495174) associated with blue eyes (86). In addition to rs4778138,
667 we also observed rs7495174 from the blue eye haplotype in our study; although it was only
668 nominally significant (PIP=004, P=1.87E-7), it may help explain the instances of blue eyes in the
669 AA+AG cohort. Finally, rs79380392 appears to be the causal SNP in the third credible set as the
670 only SNP with a $\log_{10}$BF>2. It is important to note that, although highlighted by FINEMAP, this
671 SNP did not reach genome-wide significance in our study (Table 3), so this putative causal variant
672 must be interpreted with caution.

673 *Replication*

674 The replication analyses provided strong confirmation for the majority of associations identified
675 in the discovery cohort. Among GG stratified individuals, three out of five independent signals
676 (*TYR* rs1126809, *SLC24A4* rs12896399 and *TSPAN10* rs62075722) replicated with Bonferroni-
677 corrected significance in at least one quantitative dimension. These associations were most
678 commonly observed in the b* dimension (blue-yellow). Among AA+AG individuals, three of four
679 loci (*IRF4* rs12203592, *OCA2* rs1800407, and *OCA2* rs4778138) survived Bonferroni correction
680 across multiple quantitative measures. These associations were most commonly observed in the
681 L* (lightness) dimension, consistent with the decrease in eumelanin concentration that
682 accompanies a transition from brown eye colours to intermediate and blue eye colours. The results
683 from the linear model using self-reported eye colour categories were also quite consistent with the
684 CanPath sample results, with the replicated variants showing the same effect directions as those
685 observed in linear CanPath analysis.

686 Importantly, rs12913832 was included in the AA+AG replication because there is variation in this
687 group (1 versus 2 copies of the ancestral allele). The results indicate that having one derived allele
688 (G) increases the likelihood of lighter eyes, in other words, the marker does not behave in an
689 entirely dominant fashion. In fact, the effect of this marker is particularly strong for ΔE, indicating
690 that heterozygous individuals are more likely to have central heterochromia. This observation has
691 been reported in previous studies, where each copy of the ancestral allele (A) decreased the colour
692 difference between the ciliary and pupillary zones (41). The same study also notes that rs12913832
693 may have both dominant/recessive and additive effects on iris colour variation, depending on
694 which dimensions of colour space are being measured (41).

695 *Conclusion*

696 A limitation of our study is the reliance on self-reported eye colour categories provided by
697 participants in the CanPath cohorts. While practical, such categorical data may not reflect the full
698 spectrum of iris pigmentation as accurately as quantitative measurement approaches (41,63,87).
699 Additionally, participants may report their eye colour inaccurately or may not know which
700 category to report (consider an individual with central heterochromia where the outer portion of
701 their eye appears blue but there is a brown ring around the pupil). In this regard, we were fortunate
702 to be able to replicate many of our signals in an independent European dataset for which
703 quantitative iris colour measurements and a quantitative measurement of heterochromia were
704 available. It is also important to note that the main goal of our project was to identify potential
705 modifiers of the effect of rs12913832 polymorphism on eye colour, and not to evaluate the
706 improvement on eye colour prediction based on the addition of these regions to current eye colour
707 prediction panels, which ideally would require having access to additional large datasets with well-
708 curated eye colour and genotype data. Finally, we fully acknowledge that solely focusing on
709 individuals with European genetic ancestry limits the conclusions that can be drawn. Future studies
710 would benefit from both increased size and diversity of study samples. Not only would it be
711 interesting to see replication of results in other genetic ancestry groups, but an expansion of sample
712 size and diversity may also draw attention to other genomic regions that are involved in the large
713 spectrum of observed eye colour variation. In particular, recently admixed populations would
714 provide an excellent opportunity to study in more detail gene-gene interactions shaping eye colour
715 variation. The presence of uncharacterized alleles and epistatic effects may explain in part why
716 forensic prediction tools for pigmentary traits have poorer performance in recently admixed
717 populations than in European groups (88–90).

718 Together, these findings emphasize that the genetic modifiers of eye colour differ substantially
719 between individuals with the rs12913832:GG and rs12913832:AA+AG backgrounds. When
720 combined with expression information from GTEx, a clear pattern emerges from our GWAS data.
721 For those with GG genotypes, who typically present with blue eyes, variants in core melanin
722 synthesis genes (*TYR*, *TYRP1*), transporter proteins located on melanosomes (*SLC24A4*,
723 *SLC45A2*), and potentially *TSPAN10* appear to contribute to darker-than-expected eye colour.
724 Specifically, we observe that rs1126809:G (*TYR*), rs12896399:G (*SLC24A4*), and rs62075722:G
725 (*TSPAN10*) all lead to increased expression of their respective genes (as reported in GTEx), and
726 demonstrated eye colour darkening effects in GG stratified individuals. Conversely, among
727 AA+AG individuals, lighter iris colours appear to result from a) other markers disrupting the
728 expression or function of the *OCA2* gene and its product, independently of the *HERC2* rs12913832
729 polymorphism, and b) modulation by *IRF4*, which interacts with MITF and TFAP2α (transcription
730 factor AP2α) to regulate *TYR* expression (91). For the two lead SNPs which had eQTLs in GTEx,
731 we see that the rs12203592:T (*IRF4*) and rs1800407:T allele (*OCA2*) decrease expression in their
732 respective genes, contributing to eye colour lightening in the AA+AG stratified individuals.
733 Interestingly, *TYRP1* emerges in both genotype stratified groups. It seems that *TYRP1* may work
734 both to counteract low expression *OCA2* levels in individuals with GG genotypes, and also
735 decrease melanin levels in individuals with AA+AG genotypes.

736 In summary, this study advances our understanding of eye colour variability by identifying genetic
737 variants that may modulate pigmentation in individuals with discordant *HERC2* rs12913832
738 genotypes. The identification of shared and background-specific loci underscores the polygenic
739 and regulatory complexity of iris pigmentation. By integrating categorical self-report data with
740 quantitative phenotype replication, our findings provide new insights into how moderating genetic

741 factors contribute to subtle (and sometimes more pronounced) deviations from expected eye
742 colour.

743 Acknowledgements

744 The Canadian Partnership for Tomorrow's Health (CanPath) research is only possible with the
745 commitment of its research participants, its staff and its funders. The data used in this research
746 were made available by CanPath with contributions from CARTaGENE, Alberta's Tomorrow
747 Project, the Ontario Health Study, the BC Generations Project, and Atlantic PATH. Computational
748 analyses were carried out using the high-performance computing and storage resources of the
749 Digital Research Alliance of Canada.

750 References

751 1. Simcoe M, Valdes A, Liu F, Furlotte NA, Evans DM, Hemani G, et al. Genome-wide
752 association study in almost 195,000 individuals identifies 50 previously unidentified genetic
753 loci for eye color. Sci Adv [Internet]. 2021 Mar 12 [cited 2025 Jul 14];7(11). Available from:
754 https://www.science.org/doi/10.1126/sciadv.abd1239

755 2. Eiberg H, Troelsen J, Nielsen M, Mikkelsen A, Mengel-From J, Kjaer KW, et al. Blue eye
756 color in humans may be caused by a perfectly associated founder mutation in a regulatory
757 element located within the HERC2 gene inhibiting OCA2 expression. Hum Genet. 2008
758 Mar;123(2):177–87.

759 3. Sturm RA, Larsson M. Genetics of human iris colour and patterns. Pigment Cell Melanoma
760 Res. 2009 Oct;22(5):544–62.

761 4. Sturm RA, Duffy DL, Zhao ZZ, Leite FPN, Stark MS, Hayward NK, et al. A Single SNP in
762 an Evolutionary Conserved Region within Intron 86 of the HERC2 Gene Determines Human
763 Blue-Brown Eye Color. Am J Hum Genet. 2008 Feb;82(2):424–31.

764 5. Visser M, Kayser M, Palstra RJ. *HERC2* rs12913832 modulates human pigmentation by
765 attenuating chromatin-loop formation between a long-range enhancer and the *OCA2*
766 promoter. Genome Res. 2012 Mar;22(3):446–55.

767 6. Visser M, Palstra RJ, Kayser M. Allele-specific transcriptional regulation of IRF4 in
768 melanocytes is mediated by chromatin looping of the intronic rs12203592 enhancer to the
769 IRF4 promoter. Hum Mol Genet. 2015 May 1;24(9):2649–61.

770 7. Walsh S, Liu F, Ballantyne KN, Van Oven M, Lao O, Kayser M. IrisPlex: A sensitive DNA
771 tool for accurate prediction of blue and brown eye colour in the absence of ancestry
772 information. Forensic Sci Int Genet. 2011 Jun;5(3):170–80.

773 8. Andersen JD, Pietroni C, Johansen P, Andersen MM, Pereira V, Børsting C, et al.
774 Importance of nonsynonymous *OCA 2* variants in human eye color prediction. Mol Genet
775 Genomic Med. 2016 Jul;4(4):420–30.

776 9. Andersen JD, Johansen P, Harder S, Christoffersen SR, Delgado MC, Henriksen ST, et al.
777 Genetic analyses of the human eye colours using a novel objective method for eye colour
778 classification. Forensic Sci Int Genet. 2013 Sep;7(5):508–15.

779 10. Meyer OS, Lunn MMB, Garcia SL, Kjærbye AB, Morling N, Børsting C, et al. Association
780 between brown eye colour in rs12913832:GG individuals and SNPs in TYR, TYRP1, and
781 SLC24A4. Parine NR, editor. PLOS ONE. 2020 Sep 11;15(9):e0239131.

782 11. Lona-Durazo F, Thakur R, Pairo-Castineira E, Funderburk K, Zhang T, Kovacs MA, et al.
783 Investigating the genetic architecture of eye colour in a Canadian cohort. iScience. 2022
784 Jun;25(6):104485.

785 12. Abbatangelo CL, Lona Durazo F, Wendt FR, Parra EJ, Novroski NMM. From Genetic
786 Association to Forensic Prediction: Computational Methods and Tools for Identifying
787 Phenotypically Informative Single Nucleotide Polymorphisms. Forensic Genomics. 2023 Jun
788 1;3(2):47–68.

789 13. Borugian MJ, Robson P, Fortier I, Parker L, McLaughlin J, Knoppers BM, et al. The
790 Canadian Partnership for Tomorrow Project: building a pan-Canadian research platform for
791 disease prevention. Can Med Assoc J. 2010 Aug 10;182(11):1197–201.

792 14. Dummer TJB, Awadalla P, Boileau C, Craig C, Fortier I, Goel V, et al. The Canadian
793 Partnership for Tomorrow Project: a pan-Canadian platform for research on chronic disease
794 prevention. Can Med Assoc J. 2018 Jun 11;190(23):E710–7.

795 15. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of
796 SAMtools and BCFtools. GigaScience [Internet]. 2021 Jan 29 [cited 2025 Jul 14];10(2).
797 Available from:
798 https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/giab008/6137722

799 16. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and
800 population genetical parameter estimation from sequencing data. Bioinformatics. 2011 Nov
801 1;27(21):2987–93.

802 17. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call
803 format and VCFtools. Bioinformatics. 2011 Aug 1;27(15):2156–8.

18. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience [Internet]. 2015 Dec 1 [cited 2025 Jul 14];4(1). Available from: https://academic.oup.com/gigascience/article/doi/10.1186/s13742-015-0047-8/2707533

19. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet. 2007 Sep;81(3):559–75.

20. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing:; 2022.

21. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet. 2018 Sep;50(9):1335–41.

22. Zhao J. METASOFT [Internet]. Available from: https://github.com/jinghuazhao/software-notes/tree/master/docs/files/METASOFT

23. Han B, Eskin E. Random-Effects Model Aimed at Discovering Associations in Meta-Analysis of Genome-wide Association Studies. Am J Hum Genet. 2011 May;88(5):586–98.

24. Higgins JPT. Measuring inconsistency in meta-analyses. BMJ. 2003 Sep 6;327(7414):557–60.

25. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002 Jun 15;21(11):1539–58.

26. Higgins JPT, Little J, Ioannidis JPA, Bray MS, Manolio TA, Smeeth L, et al. Turning the Pump Handle: Evolving Methods for Integrating the Evidence on Gene-Disease Association. Am J Epidemiol. 2007 Aug 28;166(8):863–6.

27. Kavvoura FK, Ioannidis JPA. Methods for meta-analysis in genetic association studies: a review of their potential and pitfalls. Hum Genet. 2008 Feb;123(1):1–14.

28. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis. Am J Hum Genet. 2011 Jan;88(1):76–82.

29. Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics. 2016 May 15;32(10):1493–501.

30. Chelala C, Khan A, Lemoine NR. SNPnexus: a web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. Bioinformatics. 2009 Mar 1;25(5):655–61.

31. Oscanoa J, Sivapalan L, Gadaleta E, Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). Nucleic Acids Res. 2020 Jul 2;48(W1):W185–92.

32. Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). Nucleic Acids Res. 2012 Jul 1;40(W1):W65–70.

33. Dayem Ullah AZ, Lemoine NR, Chelala C. A practical guide for the functional annotation of genetic variations using SNPnexus. Brief Bioinform. 2013 Jul 1;14(4):437–47.

34. Dayem Ullah AZ, Oscanoa J, Wang J, Nagano A, Lemoine NR, Chelala C. SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. Nucleic Acids Res. 2018 Jul 2;46(W1):W109–13.

35. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014 Jan;42(D1):D980–5.

36. Ng PC, Henikoff S. Predicting Deleterious Amino Acid Substitutions. Genome Res. 2001 May 1;11(5):863–74.

37. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010 Apr;7(4):248–9.

38. Schubach M, Maass T, Nazaretyan L, Röner S, Kircher M. CADD v1.7: using protein language models, regulatory CNNs and other nucleotide-level scores to improve genome-wide variant predictions. Nucleic Acids Res. 2024 Jan 5;52(D1):D1143–54.

39. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2023 Jan 6;51(D1):D638–46.

40. The GTEx Consortium, Aguet F, Anand S, Ardlie KG, Gabriel S, Getz GA, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020 Sep 11;369(6509):1318–30.

41. Edwards M, Cha D, Krithika S, Johnson M, Cook G, Parra EJ. Iris pigmentation as a quantitative trait: variation in populations of European, East Asian and South Asian ancestry and association with candidate gene polymorphisms. Pigment Cell Melanoma Res. 2016 Mar;29(2):141–62.

42. Durbin R. Efficient haplotype matching and storage using the positional Burrows–Wheeler transform (PBWT). Bioinformatics. 2014 May 1;30(9):1266–72.

43. Peña-Chilet M, Blanquer-Maceiras M, Ibarrola-Villava M, Martinez-Cadenas C, Martin-Gonzalez M, Gomez-Fernandez C, et al. Genetic variants in PARP1 (rs3219090) and IRF4(rs12203592) genes associated with melanoma susceptibility in a Spanish population.

873 BMC Cancer [Internet]. 2013 Dec [cited 2025 Jul 14];13(1). Available from:
874 https://bmccancer.biomedcentral.com/articles/10.1186/1471-2407-13-160

875 44. Jacobs LC, Liu F, Pardo LM, Hofman A, Uitterlinden AG, Kayser M, et al. IRF4, MC1R and
876 TYR genes are risk factors for actinic keratosis independent of skin color. Hum Mol Genet.
877 2015 Jun 1;24(11):3296–303.

878 45. Landi MT, Bishop DT, MacGregor S, Machiela MJ, Stratigos AJ, Ghiorzo P, et al. Genome-
879 wide association meta-analyses combining multiple risk phenotypes provide insights into the
880 genetic architecture of cutaneous melanoma susceptibility. Nat Genet. 2020 May;52(5):494–
881 504.

882 46. Branicki W, Brudnik U, Wojas-Pelc A. Interactions Between *HERC2* , *OCA2* and *MC1R*
883 May Influence Human Pigmentation Phenotype. Ann Hum Genet. 2009 Mar;73(2):160–70.

884 47. Donnelly MP, Paschou P, Grigorenko E, Gurwitz D, Barta C, Lu RB, et al. A global view of
885 the OCA2-HERC2 region and pigmentation. Hum Genet. 2012 May;131(5):683–96.

886 48. Wallpe C. Understanding the relationship between HERC2 and OCA2 variants and iris
887 pigmentation genetics. [Internet]. Purdue University; 2021. Available from:
888 https://hammer.purdue.edu/articles/thesis/UNDERSTANDING_THE_RELATIONSHIP_BE
889 TWEEN_HERC2_AND_OCA2_VARIANTS_AND_IRIS_PIGMENTATION_GENETICS/
890 15057078?file=28957086

891 49. Ghodsinejad Kalahroudi V, Kamalidehghan B, Arasteh Kani A, Aryani O, Tondar M,
892 Ahmadipour F, et al. Two Novel Tyrosinase (TYR) Gene Mutations with Pathogenic Impact
893 on Oculocutaneous Albinism Type 1 (OCA1). Anderson MG, editor. PLoS ONE. 2014 Sep
894 12;9(9):e106656.

895 50. Michaud V, Lasseaux E, Green DJ, Gerrard DT, Plaisant C, UK Biobank Eye and Vision
896 Consortium, et al. The contribution of common regulatory and protein-coding TYR variants
897 to the genetic architecture of albinism. Nat Commun [Internet]. 2022 Jul 8 [cited 2025 Jul
898 14];13(1). Available from: https://www.nature.com/articles/s41467-022-31392-3

899 51. Monfermé S, Lasseaux E, Duncombe-Poulet C, Hamel C, Defoort-Dhellemmes S, Drumare
900 I, et al. Mild form of oculocutaneous albinism type 1: phenotypic analysis of compound
901 heterozygous patients with the R402Q variant of the TYR gene. Br J Ophthalmol. 2019
902 Sep;103(9):1239–47.

903 52. Jannot AS, Meziani R, Bertrand G, Gérard B, Descamps V, Archimbaud A, et al. Allele
904 variations in the OCA2 gene (pink-eyed-dilution locus) are associated with genetic
905 susceptibility to melanoma. Eur J Hum Genet. 2005 Aug;13(8):913–20.

906 53. Rebbeck TR, Kanetsky, PA, Walker AH, Holmes R, Halpern AC, Schuchter LM, et al. P
907 Gene as an Inherited Biomarker of Human Eye Color. Cancer Epidemiol Biomarkers Prev.
908 2002;11(8):782–4.

54. Salvo NM, Andersen JD, Janssen K, Meyer OL, Berg T, Børsting C, et al. Association between Variants in the OCA2-HERC2 Region and Blue Eye Colour in HERC2 rs12913832 AA and AG Individuals. Genes. 2023 Mar 11;14(3):698.

55. Liu F, Van Duijn K, Vingerling JR, Hofman A, Uitterlinden AG, Janssens ACJW, et al. Eye color and the prediction of complex phenotypes from genotypes. Curr Biol. 2009 Mar;19(5):R192–3.

56. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Magnusson KP, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat Genet. 2007 Dec;39(12):1443–52.

57. Chaitanya L, Breslin K, Zuñiga S, Wirken L, Pośpiech E, Kukla-Bartoszek M, et al. The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. Forensic Sci Int Genet. 2018 Jul;35:123–35.

58. Walsh S, Chaitanya L, Breslin K, Muralidharan C, Bronikowska A, Pospiech E, et al. Global skin colour prediction from DNA. Hum Genet. 2017 Jul;136(7):847–63.

59. Walsh S, Chaitanya L, Clarisse L, Wirken L, Draus-Barini J, Kovatsi L, et al. Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. Forensic Sci Int Genet. 2014 Mar;9:150–61.

60. Ruiz Y, Phillips C, Gomez-Tato A, Alvarez-Dios J, Casares De Cal M, Cruz R, et al. Further development of forensic eye color predictive tests. Forensic Sci Int Genet. 2013 Jan;7(1):28–40.

61. www.genecards.org [Internet]. 2025 [cited 2025 Jun 30]. SLC45A2 gene - GeneCards summary for uniprot function. Available from: ttps://www.genecards.org/cgi-bin/carddisp.pl?gene=SLC45A2

62. Zeng X. Selection of Highly Informative Markers for Apportionment of Ancestry and Population Affiliation [Internet]. University of North Texas Health Science Center at Fort Worth; 2016. Available from: http://myaccess.library.utoronto.ca/login?qurl=https%3A%2F%2Fwww.proquest.com%2Fdissertations-theses%2Fselection-highly-informative-markers%2Fdocview%2F1777348958%2Fse-2%3Faccountid%3D14771

63. Liu J, Hutchison K, Perrone-Bizzozero N, Morgan M, Sui J, Calhoun V. Identification of Genetic and Epigenetic Marks Involved in Population Structure. Ravasi T, editor. PLoS ONE. 2010 Oct 7;5(10):e13209.

64. www.genecards.org [Internet]. 2025 [cited 2025 Jun 30]. TYRP1 gene - GeneCards summary for uniprot function. Available from: https://www.genecards.org/cgi-bin/carddisp.pl?gene=TYRP1

65. Gautron A, Migault M, Bachelot L, Corre S, Galibert M, Gilot D. Human *TYRP1*: Two functions for a single gene? Pigment Cell Melanoma Res. 2021 Sep;34(5):836–52.

946 66. Sarangarajan R, Boissy RE. Tyrp1 and Oculocutaneous Albinism Type 3. Pigment Cell Res.
947 2001 Dec;14(6):437–44.

948 67. Murisier F, Beermann F. Genetics of pigment cells: lessons from the tyrosinase gene family.
949 Histol Histopathol. 2006;(21):567–78.

950 68. Berson JF, Frank DW, Calvo PA, Bieler BM, Marks MS. A Common Temperature-sensitive
951 Allelic Form of Human Tyrosinase Is Retained in the Endoplasmic Reticulum at the
952 Nonpermissive Temperature. J Biol Chem. 2000 Apr;275(16):12281–9.

953 69. Jagirdar K, Smit DJ, Ainger SA, Lee KJ, Brown DL, Chapman B, et al. Molecular analysis
954 of common polymorphisms within the human *Tyrosinase* locus and genetic association with
955 pigmentation traits. Pigment Cell Melanoma Res. 2014 Jul;27(4):552–64.

956 70. Cooksey CJ, Garratt PJ, Land EJ, Pavel S, Ramsden CA, Riley PA, et al. Evidence of the
957 Indirect Formation of the Catecholic Intermediate Substrate Responsible for the
958 Autoactivation Kinetics of Tyrosinase. J Biol Chem. 1997 Oct;272(42):26226–35.

959 71. Biosynthesis of Melanins. In: Melanins and Melanosomes [Internet]. 1st ed. Wiley; 2011
960 [cited 2025 Jul 14]. p. 87–116. Available from:
961 https://onlinelibrary.wiley.com/doi/10.1002/9783527636150.ch4

962 72. Eduardoff M, Gross TE, Santos C, De La Puente M, Ballard D, Strobl C, et al. Inter-
963 laboratory evaluation of the EUROFORGEN Global ancestry-informative SNP panel by
964 massively parallel sequencing using the Ion PGM™. Forensic Sci Int Genet. 2016
965 Jul;23:178–89.

966 73. Phillips C, Salas A, Sánchez JJ, Fondevila M, Gómez-Tato A, Álvarez-Dios J, et al. Inferring
967 ancestral origin using a single multiplex assay of ancestry-informative marker SNPs.
968 Forensic Sci Int Genet. 2007 Dec;1(3–4):273–80.

969 74. Phillips C, Parson W, Lundsberg B, Santos C, Freire-Aradas A, Torres M, et al. Building a
970 forensic ancestry panel from the ground up: The EUROFORGEN Global AIM-SNP set.
971 Forensic Sci Int Genet. 2014 Jul;11:13–25.

972 75. Inference of Ancestry in Forensic Analysis I: Autosomal Ancestry-Informative Marker Sets.
973 In: Methods in Molecular Biology [Internet]. New York, NY: Springer New York; 2016
974 [cited 2025 Jul 14]. p. 233–53. Available from: http://link.springer.com/10.1007/978-1-4939-
975 3597-0_18

976 76. Inference of Ancestry in Forensic Analysis II: Analysis of Genetic Data. In: Methods in
977 Molecular Biology [Internet]. New York, NY: Springer New York; 2016 [cited 2025 Jul 14].
978 p. 255–85. Available from: http://link.springer.com/10.1007/978-1-4939-3597-0_19

979 77. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-
980 specific haplotype structure and linking correlated alleles of possible functional variants.
981 Bioinformatics. 2015 Nov 1;31(21):3555–7.

982 78. www.unitpro.org [Internet]. 2025 [cited 2025 Jun 15]. TSN10_HUMAN (Q9H1Z9).
983 Available from: https://www.uniprot.org/uniprotkb/Q9H1Z9/entry#expression

984 79. Terzian T, Torchia EC, Dai D, Robinson SE, Murao K, Stiegmann RA, et al. p53 prevents
985 progression of nevi to melanoma predominantly through cell cycle regulation. Pigment Cell
986 Melanoma Res. 2010 Dec;23(6):781–94.

987 80. Laino AM, Berry EG, Jagirdar K, Lee KJ, Duffy DL, Soyer HP, et al. Iris pigmented lesions
988 as a marker of cutaneous melanoma risk: an Australian case-control study. Br J Dermatol.
989 2018 May;178(5):1119–27.

990 81. Wang S, Huo D, Kupfer S, Alleyne D, Ogundiran TO, Ojengbede O, et al. Genetic variation
991 in the vitamin D related pathway and breast cancer risk in women of African ancestry in the
992 root consortium. Int J Cancer. 2018 Jan;142(1):36–43.

993 82. Fuhrman B, Cardenas V. Melanocytic Nevi as Biomarkers of Breast Cancer Risk. PLoS
994 Med. 2014 Jun 10;11(6):e1001661.

995 83. Kvaskoff M, Bijon A, Mesrine S, Vilier A, Baglietto L, Fournier A, et al. Association
996 between Melanocytic Nevi and Risk of Breast Diseases: The French E3N Prospective
997 Cohort. Holmberg L, editor. PLoS Med. 2014 Jun 10;11(6):e1001660.

998 84. www.genecards.org [Internet]. 2025 [cited 2025 Jun 30]. OCA2 gene – GeneCards summary
999 for uniprot function. Available from: https://www.genecards.org/cgi-
1000 bin/carddisp.pl?gene=OCA2&keywords=OCA2

1001 85. Walsh S, Wollstein A, Liu F, Chakravarthy U, Rahu M, Seland JH, et al. DNA-based eye
1002 colour prediction across Europe with the IrisPlex system. Forensic Sci Int Genet. 2012
1003 May;6(3):330–40.

1004 86. Duffy DL, Montgomery GW, Chen W, Zhao ZZ, Le L, James MR, et al. A Three–Single-
1005 Nucleotide Polymorphism Haplotype in Intron 1 of OCA2 Explains Most Human Eye-Color
1006 Variation. Am J Hum Genet. 2007 Feb;80(2):241–52.

1007 87. Norton HL, Edwards M, Krithika S, Johnson M, Werren EA, Parra EJ. Quantitative
1008 assessment of skin, hair, and iris variation in a diverse sample of individuals and associated
1009 genetic variation. Am J Phys Anthropol. 2016 Aug;160(4):570–81.

1010 88. Beleza S, Johnson NA, Candille SI, Absher DM, Coram MA, Lopes J, et al. Genetic
1011 Architecture of Skin and Eye Color in an African-European Admixed Population. Spritz RA,
1012 editor. PLoS Genet. 2013 Mar 21;9(3):e1003372.

1013 89. Freire-Aradas A, Ruiz Y, Phillips C, Maroñas O, Söchtig J, Tato AG, et al. Exploring iris
1014 colour prediction and ancestry inference in admixed populations of South America. Forensic
1015 Sci Int Genet. 2014 Nov;13:3–9.

1016 90. Marano LA, Andersen JD, Goncalves FT, Garcia ALO, Fridman C. Evaluation of HIrisplex-
1017 S system markers for eye, skin and hair color prediction in an admixed Brazilian population.
1018 Forensic Sci Int Genet Suppl Ser. 2019 Dec;7(1):427–8.

1019 91. Chhabra Y, Yong HXL, Fane ME, Soogrim A, Lim W, Mahiuddin DN, et al. Genetic
1020 variation in IRF4 expression modulates growth characteristics, tyrosinase expression and
1021 interferon-gamma response in melanocytic cells. Pigment Cell Melanoma Res. 2018
1022 Jan;31(1):51–63.

1023

1024

1025 Authors' Roles

1026

1027 CLA: Formal Analysis, Writing – original draft, Writing – review and editing. FLD: Methodology,
1028 Writing – review and editing. ME: Data collection, Writing – review and editing. EJP:
1029 Conceptualization, Methodology, Writing – review and editing, Supervision.

1030

1031 Competing Interests

1032

1033 N/A

1034

1035

1036

1037

1038

1039

1040

1041

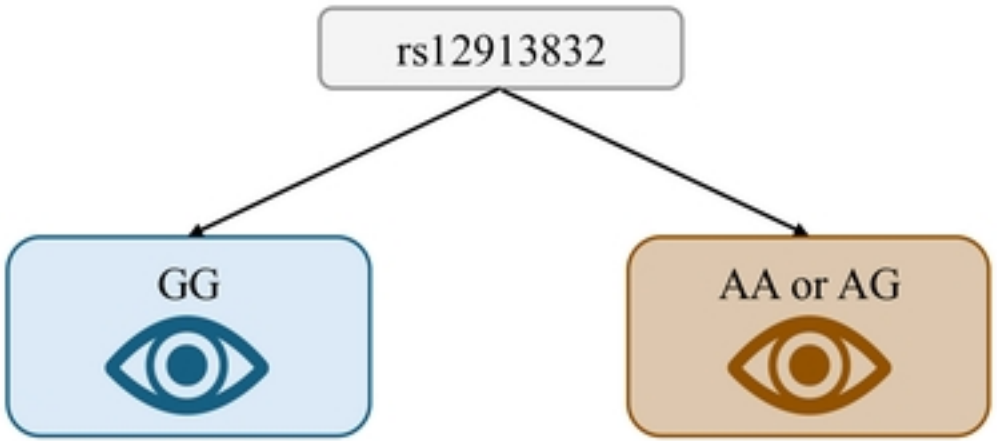

Fig 1

6 COHORTS
9 PROVINCES
300,000+ CANADIANS
CanPath
Canadian Partnership for Tomorrow's Health
ALBERTA'S TOMORROW PROJECT
BC GENERATIONS PROJECT
CARTOGENE
CancerCare Manitoba
ActionCancerManitoba
THE MANITOBA TOMORROW PROJECT
Atlantic PATH
Ontario Health Study
Étude sur la santé Ontario

Fig 2

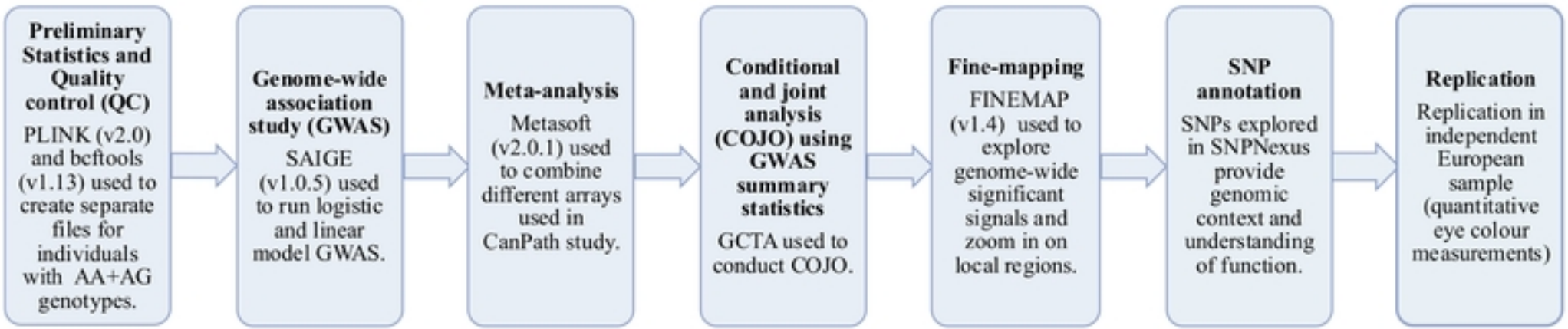
Preliminary Statistics and Quality control (QC)
PLINK (v2.0) and bcftools (v1.13) used to create separate files for individuals with AA+AG genotypes.
Genome-wide association study (GWAS)
SAIGE (v1.0.5) used to run logistic and linear model GWAS.
Meta-analysis
Metasoft (v2.0.1) used to combine different arrays used in CanPath study.
Conditional and joint analysis (COJO) using GWAS summary statistics
GCTA used to conduct COJO.
Fine-mapping
FINEMAP (v1.4) used to explore genome-wide significant signals and zoom in on local regions.
SNP annotation
SNPs explored in SNPNexus provide genomic context and understanding of function.
Replication
Replication in independent European sample (quantitative eye colour measurements)

Fig 3

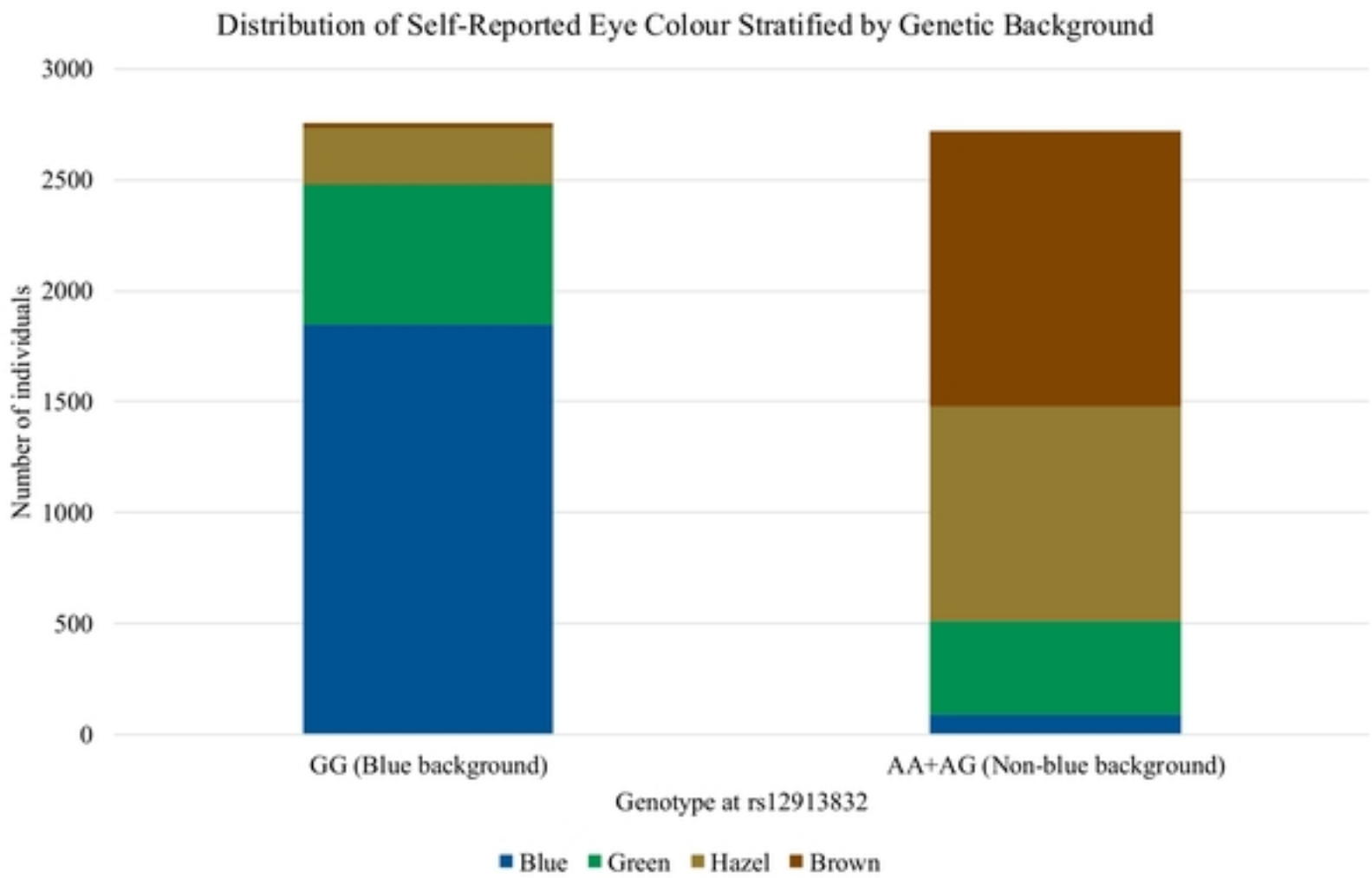
Distribution of Self-Reported Eye Colour Stratified by Genetic Background
3000
2500
2000
1500
1000
500
0
Number of individuals
GG (Blue background)
AA+AG (Non-blue background)
Genotype at rs12913832
Blue
Green
Hazel
Brown

Fig 4

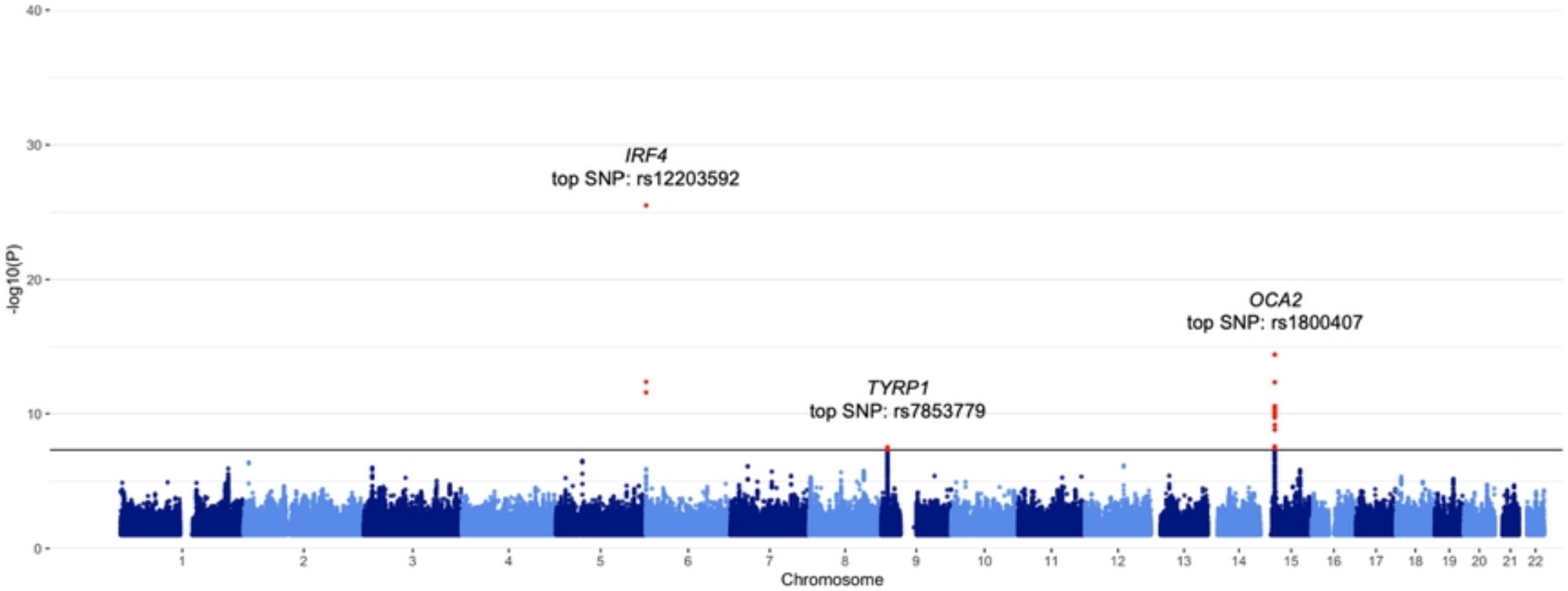
40
30
20
10
0
-log10(P)
IRF4
top SNP: rs12203592
TYRP1
top SNP: rs7853779
OCA2
top SNP: rs1800407
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

Fig 5