

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Skin Phototype and Disease. A Comprehensive Genetic Approach to Pigmentary Traits Pleiotropy using PRS in the GCAT Cohort

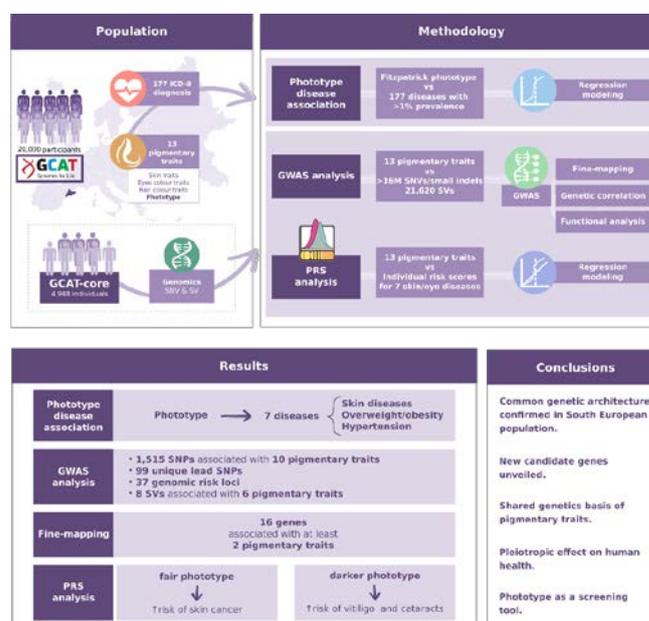
Xavier Farré ^{1#}, Natalia Blay ^{1#}, Beatriz Cortés ^{1#}, Anna Carreras ¹, Susana Iraola-Guzmán ¹ and Rafael de Cid ^{1*}

¹. Genomes for Life-GCAT lab. Germans Trias i Pujol Research Institute (IGTP), Badalona, Spain;

These authors contributed equally and should be considered first authors;

* Correspondence: rdecid@igtp.cat;

Abstract: Human pigmentation has been largely associated with different disease prevalence among populations, but most of these studies are observational and inconclusive. Known to be genetically determined, pigmentary traits have been largely studied by GWAS, mostly in Caucasian ancestry cohorts from North Europe, identifying robustly, several loci involved in many of the pigmentary traits. Here, we conduct a detailed analysis by GWAS and PRS of 13 pigmentary-related traits in a South European cohort of Caucasian ancestry (n=20,000). We observed fair phototype strongly associated with non-melanoma skin cancer and other dermatoses and confirmed by PRS-approach the shared genetic basis with skin and eye diseases, such as melanoma (OR=0.95), non-melanoma skin cancer (OR=0.93), basal cell carcinoma (OR=0.97) and darker phototype with vitiligo (OR=1.02), cataracts (OR=1.04). Detailed genetic analyses revealed 37 risk loci associated with 10 out of 13 analyzed traits, and 16 fine-mapped genes significantly associated with at least two pigmentary traits. Some of them widely reported, such as *MC1R*, *HERC2*, *OCA2*, *TYR*, *TYRP1*, *SLC45A2*, and unveiling three novel candidates *RP11-1084j3.4*, *C1QTNF3* and *C17orf112*, not reported in GWAS Catalog. These results highlight the importance of assess phototype as a genetic proxy of skin functionality and disease when evaluating open mixed populations.



Keywords: Population genetics; Pigmentary traits; Phototype; Pleiotropy; Skin disease; Eye disease

1. Introduction

Skin, hair, and eye pigmentation are determined by the type and amount, of melanin in the basal layer of the epidermis, hair follicles, and the front layers of the iris, respectively [1-5]. In the human skin, the production of melanin, takes place through the melanogenesis within the melanocytes [6, 7], and is promoted by exposure ultraviolet radiation (UVR), which darkens the skin, the eyes [5] and promotes freckles development [8]. Melanin pigmentation protects the skin from the damaging effects of UVR, but exerts, a wide variety of functions, and harbors different structures, and presentations [9], being not only responsible for color, but also, found in the light-sensitive tissue of the retina, playing a role in normal vision [10].

Pigmentary traits are genetically determined, heritable [11], and with a polygenic architecture [12]. Some traits, such as, skin or eye color are constitutive, while other traits, as ease of skin tanning or skin burning, are facultative, and more related to the functionality of the skin, in an interplay between genetics and environmental factors [13, 14]. In dermatology practice, the structure and functional behavior of the skin is condensed by the phototype, a skin classification not only defined by the skin color, but also including other pigmentary traits (hair and eye color, presence of freckles and sunlight sensitivity).

Several studies have untangled the genetic component of these traits, identifying the main genes associated with several pigmentary traits. *MC1R* was associated with red hair, fair skin, or freckles [15, 16]; *OCA2* was previously associated with hair and eye color [17, 18]; *SLC24A5*, linked to pigmentation in zebrafish [19], was also confirmed to play an important role in human pigmentation. Specifically, a non-synonymous variant in the third exon of this gene (rs1426654), explains between 25 to 38% of the difference between Europeans and Africans in skin melanin index [19]; *IRF4* was associated with skin, hair, and eye pigmentation [20, 21] and *TYR* was associated with eye color type [22, 23]. However, most of the comprehensive genetics on Caucasian Ancestry for pigmentary traits are conducted on North European populations [24] or small studies [25, 26].

Ancestry background is important because natural selection shapes the overall genetic architecture of skin pigmentation [27]. The latitude hypothesis, explained by favoring vitamin D synthesis in regions far from the equator, is accepted as the main force during ancestral human population spreading [28], but not the only explanation. Recent migrations and cultural factors have brought many people into UVR regimes different from those experienced by their ancestors, and accordingly, exposed them to new disease risks [29]. Interestingly, those genes, while playing an important role in skin pigmentation, are also known to be associated with the prevalence and incidence of skin-colored related diseases, such as melanoma [30-32], cutaneous squamous cell carcinoma [33], vitiligo [34], or indirectly, involved in the cancer aggressiveness, where abnormal melanin physiology is involved in the immunomodulation of tumor microenvironment, sustaining melanoma progression and metastasis [35]. Furthermore, modification of the production and packaging of melanin in the skin affects vitamin D levels [36], which is essential for a wide range of physiological processes, including immune function and calcium homeostasis.

In this study, we report a comprehensive analysis of the GCAT cohort, with the identification of phototype associated-diseases in a total of nearly 20,000 subjects and 177 ICD-9 diagnoses, and genome-wide analysis, including structural variants, across 13 pigmentary traits, using polygenic risk scores, in the largest South European population analyzed so far.

2. Materials and Methods

Graphical pipeline including all the material and methods can be found in Supplementary Scheme I.

2.1. Study population

Participants included in this study, belong to the GCAT|Genomes for life cohort, a population-based cohort from Southern Europe (Catalonia, NE Spain) [37]. The GCAT

cohort comprises 20,000 volunteers, recruited between 2014-2018, from Catalan general population, aged between 40 and 65 years old at the time of recruitment, and 59.16% of women. All the participants completed a detailed, self-reported, baseline questionnaire, including self-reported pigmentary-traits definition and family ethnicity. All participants gave their consent, and all procedures were carried out in accordance with ethical standards. This study was approved by the ethics committee of the Hospital Germans Trias i Pujol (Protocol CEIC PI-13-020).

2.2. Phenotype data collection

Pigmentary traits were collected for all subjects with a self-report epidemiological questionnaire, during the recruitment (including 19,205 valid responses). These data include single pigmentation-related phenotypes, comprising: three binary phenotypes for hair color, including red hair (red hair as cases, and blond, brown and black hair as controls), blond hair (blond hair as cases, and brown and black hair as controls), and brown hair (brown hair as cases and black hair as controls); eye color (i.e., light blue or green, blue or green, hazel, brown and dark brown); skin color (i.e., black, dark, medium, white and very white); freckles (i.e., no, occasional, few, some and abundant); ease sunburn (i.e., never, rarely, sometimes, moderately and easily); ease suntan (i.e., never, slightly, moderate and strongly); as well as, sun habits, such as, the use of sunscreen in exterior (i.e., never or rarely, sometimes, almost all the time, always or not in the exterior during sunny moments); the use of sunscreen when suntan (i.e., never or rarely, sometimes, almost all the time, always or not in the exterior during sunny moments) and weekly hours spent outside (Table S1). For phototype study, we included Fitzpatrick scale, this is a phototype variable calculated from the aggregation of self-reported variables based on the widely used Fitzpatrick skin phototype classification [38], including information about skin color, hair color, eye color, freckles, ethnicity, sunburn, and suntan. This information was obtained from a questionnaire, where a score was calculated (Table S2), and transformed into the six phototype categories (from I to VI) corresponding to Fitzpatrick skin phototype classification (see description in Table S3). A second variable, phototype score, based on the raw score from the questionnaire was also included in the study for analysis purposes. This score ranges from 0 to 86 in our study, being the higher scores the darker skins with higher sunlight tolerance. A detailed description of all included variables can be found in Table S1 and Figure S1.

2.3. Genetic variants

Genetic analysis was conducted in a subsample of the GCAT cohort (GCATcore) that is fully genotyped by SNP-array and imputation. This subsample included 4,988 participants with Iberian ethnicity, determined by self-described ethnicity, and supervised ancestry inference by Principal Components Analysis (PCA), as previously described by Galván Femenía et al. [39]. Genotyping was done using the Infinium Expanded Multi-Ethnic Genotyping Array (MEGAEX) (ILLUMINA, San Diego, California, USA). Then, we imputed both SNVs and SVs, with IMPUTE2, and using GCAT|Panel as reference [40], an ancestry-geographically matched panel generated by whole-genome sequencing. The final core set included a total of 10,216,971 variants with MAF ≥ 0.01 and INFO score ≥ 0.7 (21,620 SVs, and 10,195,351 SNVs and small indels). All data is available at EGA (Study ID EGAS00001003018).

2.4. Electronic health records disease prevalence and phototype

Medical diagnoses of the whole cohort were obtained from Electronic Health Records (EHR) of the Public Healthcare System of Catalonia. All diagnostic information (2010-current) is linked to each participant, within the framework of the PADRIS Program as provided by the Agency of Health Quality and Assessment of Catalonia (AQuAS) [41]. We selected all ICD-9 minimum basic data set (MBDS) diagnoses from the period 2012-2017, grouped them into three digits' level, and selected the ones with a prevalence of at least

1%, 177 diagnoses were finally included in the analysis. Then, a regression analysis was conducted for each disease with Fitzpatrick scale (phototypes I-VI), correcting by age and gender. A Bonferroni correction by the number of ICD-9 chapters ($n = 15$) was applied, accounting for redundancy. Results are reported as odds ratios (OR) with 95% confidence intervals (CI) and p-value (p).

2.5. Genome-wide association analysis of phototype

Independent genome-wide association analysis (GWAS) for the 13 pigmentary-related traits (as three binary and ten continuous variables) were performed, including a total of 16,967,118 SNVs and small indels, and 21,620 SVs with a Minor Allele Frequency (MAF) ≥ 0.01 and INFO score ≥ 0.7 . Linear and logistic regression models were used for continuous and binary traits respectively using PLINK (version 2.0) [42], assuming an additive genetic model and including age and sex as covariates, as well as the first 10 PCA to control the potential population stratification. Standard genome-wide threshold ($p < 5e-8$) was kept for significance. SNPs were functionally annotated and mapped to the closest gene with SnpEff [40]. Manhattan and quantile-quantile (QQ) plots were generated to explore the results, as well as a 3D Manhattan representation that provides a visual representation of the common GWAS hits between the different traits. Regional Manhattan plots for the associated SVs were represented, including information from the genes in that region obtained from GENCODE (v40) [43]. Additionally, the association results were analyzed using MAGMA [44], a gene-based approach that aggregates multiple genetic markers into genes to determine their joint effect.

Then, for each GWAS, independent genomic loci were identified as described in Watanabe et al [45], using GCAT panel [40] as the reference panel to compute the linkage disequilibrium (LD) measure. First, we defined independent significant SNPs (PLINK software, $p_1 = 0.05$; $p_2 = 1$, linkage disequilibrium threshold $r_2 = 0.6$, and physical distance threshold for clumping 1000kb), and lead SNPs (PLINK software, $p_1 = 0.05$, $p_2 = 1$, $r_2 = 0.1$, and distance 1000kb). Then, LD blocks that were closer than 250kb were merged into genomic trait-associated loci.

To further elucidate the biological mechanisms underlying pigmentary traits, we performed a functional enrichment analysis using WebGestalt [46], which tests for an overrepresentation of the genes resulting from the gene-based analysis performed with MAGMA, in specific GO terms, pathways, and disease-associated gene sets.

2.6. Trait heritability and pairwise genetic correlations

We estimated the SNP heritability of each individual trait (h^2), and computed pairwise genetic correlations, using LDSC (v1.0.1) [47] with the precalculated LD scores for 1000 Genomes phase 3 for European population. For binary traits, SNP heritability and genetic correlations were computed at the liability scale, assuming that the sample prevalence is equal to the population prevalence.

2.7. EHR phenotype-wide association

A Phenotype Wide association study (PheWAS) was conducted with former identified lead SNPs using the publicly available data PheWeb [48] that includes GWAS results for 1,419 EHR-derived PheWAS codes in approximately 400,000 White British participants of the UK Biobank [49]. Effect sizes directions from the PheWAS analysis were referenced to the same effect allele as the one in our GWAS. For strand ambiguous SNPs, allele frequencies were checked to reference the data to the same effect allele, excluding those SNPs in which the allele frequency ranged from 0.4 to 0.6.

2.8. Polygenic risk scores association

Polygenic risk scores (PRS) were derived for selected relevant skin and eye diseases, based on disease-phototype association, GWAS results, and available genetic data from

public databases. Here we present data for three skin cancer phenotypes (including melanoma, basal cell carcinoma, and non-melanoma skin cancer), two eye diseases (glaucoma and cataracts), two immune-related dermatoses (psoriasis and vitiligo) and metabolic diseases (essential hypertension and overweight, obesity and other hyperalimentation). We used weights for relevant diseases from the PGS Catalog [50] giving priority to those polygenic scores in which the source of variant associations (GWAS) was from European ancestry (Table S4). We harmonized the data by removing strand ambiguous SNPs and discordant SNP alleles. Then, polygenic risk scores for each individual were computed using PLINK v.1.9 [42]. Associations between pigmentary traits and polygenic risk scores were assessed using linear or logistic regression, accounting for age and sex of the participants, as well as the first 10 PCAs to control for population stratification. Results are reported as odds ratios (OR) with 95% confidence intervals (CI), and p-value (p).

3. Results

3.1. Systematic search on clinical data (EHR) identifies diseases associated with phototype in the GCAT cohort.

Among the selected 177 EHR ICD-9 diagnoses, seven diseases were significantly associated with a fair Fitzpatrick scale phototype after Bonferroni correction, and eight were nominally (Figure S2 and Table S5). Among significantly associated diagnoses, *Diseases of the skin, and subcutaneous tissue* and *Neoplasms* were the two most represented ICD-9 chapters (71%). The most significantly associated diseases were: *Other and unspecified malignant neoplasm of skin* (ICD-9 code 173, OR = 0.60, CI = 0.5-0.71, $p = 1.97e-9$), and *Other dermatoses* (ICD-9 code 702, OR = 0.79, CI = 0.7-0.88, $p = 1.67e-5$). Other prevalent diseases included: *Benign neoplasm of skin* (ICD-9 code 216), *Contact dermatitis and other eczema* (ICD-9 code 692), and *Erythematous conditions* (ICD-9 code 695). Also, we found two common conditions associated with fair phototype: *Overweight, obesity and other hyperalimentation* (ICD-9 code 278) and *Essential hypertension* (ICD-9 code 401).

3.2. GWAS analysis identifies thousand variants associated with 10 pigmentary traits in the GCAT cohort.

We carried out a GWAS across >10M variants with a MAF ≥ 0.01 and INFO score ≥ 0.7 for each individual pigmentary trait, including SNVs and SVs (Figure S3). Single-variant analysis identified 3,562 associations at 1,515 SNPs in 10 pigmentary traits (red hair, blond hair, brown hair, eye color, freckles, ease sunburn, ease suntan, skin color phototype score and Fitzpatrick scale), the other three (sunscreen in exterior, sunscreen when suntan and weekly hours spent in the exterior) did not show any significant association. The 1,515 associated SNPs were grouped into 257 independent SNPs and 99 unique lead SNPs grouped into 37 genomic risk loci, mapping to 48 genes (Table S6). A summary of genome-wide significant variants and their associated loci is shown in Table 1. Detailed results regarding SVs are described later in section 3.5. Twelve out of the 48 genes corresponding to six loci, were pleiotropic, associated with more than one pigmentary trait (*SLC45A2*, *HERC2*, *OCA2*, *IRF4*, *MC1R*, *ADAMTS12*, *RXFP3*, *TYR*, *TYRP1*, *NOX4*, *AMACR*, *EXOC2*). Noteworthy, *SLC45A2*, a critical determinant of skin and eye pigmentation, was the most pleiotropic gene, associated with nine pigmentary traits.

Table 1. Genome-wide association study of pigmentary traits results overview. Overview of the results obtained in the study, summarizing the 13 analyzed traits, type of trait (constitutive, facultative or built), type of variable (continuous, discrete or binary), total number of SNPs identified, number of independent SNPs, number of lead SNPs, number of risk loci, and number of structural variants.

Trait	Classification	Type	SNPs (n)	Independent SNPs (n)	Lead SNPs (n)	Genomic risk loci (n)	SVs (n)
Red hair	Constitutive	Binary	181	54	28	23	4
Blond hair	Constitutive	Binary	209	17	5	3	0

Brown hair	Constitutive	Binary	143	22	8	4	0
Eye color	Constitutive	Discrete	829	97	28	6	2
Skin color	Constitutive	Discrete	478	66	20	7	2
Freckles	Facultative	Discrete	30	7	5	4	0
Ease sunburn	Facultative	Discrete	170	28	11	6	0
Ease suntan	Facultative	Discrete	190	25	9	4	2
Sunscreen in exterior	Behavioral	Discrete	0	0	0	0	0
Sunscreen when suntan	Behavioral	Discrete	0	0	0	0	0
Weekly hours spent in the exterior	Behavioral	Continuous	0	0	0	0	0
Phototype score	Built	Continuous	768	111	34	6	4
Fitzpatrick scale	Built	Discrete	564	82	22	5	4
Total			1,515	257	99	37	8

GWAS results were represented in a 3D Manhattan plot to highlight the pleiotropic nature of associated regions (Figure 1). *SLC45A2* (lead rs16891982) was associated with nine traits (blond hair, brown hair, eye color, ease sunburn, ease suntan, Fitzpatrick scale, phototype score, skin color and freckles); *IRF4* (lead rs12203592) show association with six traits (brown hair, eye color, ease sunburn, ease suntan, skin color and freckles); *HERC2* (lead rs12913832) was associated with seven traits (blond hair, brown hair, ease sunburn, ease suntan, Fitzpatrick scale, phototype score and skin color) and *MC1R* (lead rs1805007) presented association with four traits (red hair, ease suntan, Fitzpatrick scale and phototype score). Pleiotropy, as expected, was evident for the aggregated variables of phototype, which include dependent variables; but it was also higher for skin color, suntan, sunburn, and brown hair; and lower in red and blond hair color and eye color (Figure 2). Since there was a gene overlapping in the 37 genomic risk loci, we refined these results using a gene-based analysis for further steps (Section 3.4).

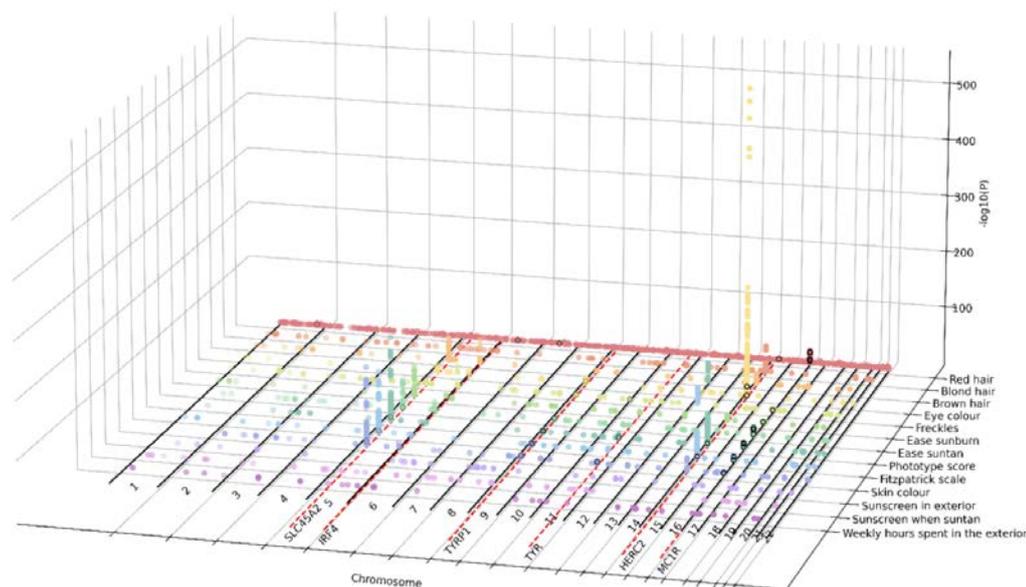


Figure 1. Combined GWAS results for the 13 pigmentation traits highlight their pleiotropic nature. Three-dimensional Manhattan plot depicting in the X axis the chromosomal positions of the genome-wide significant associations; in the Y axis, pigmentation traits represented by colors; and in the Z axis, the level of significance of the association, expressed as $-\log_{10}(p\text{-value})$. Colored dots represent the lead SNPs associated to one or more pigmentation traits. Black circles represent SVs, overlapping with significant GWAS hits. Red-dotted lines represent the mapping of genome-wide significant associated SNPs with several pigmentation traits mapping to *SLC45A2*, *IRF4*, *TYR*, *HERC2* and *MC1R* genes.

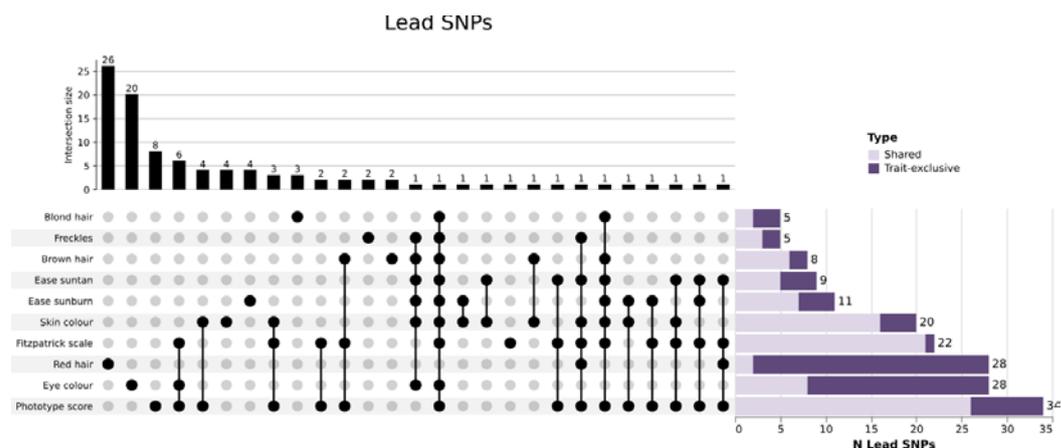


Figure 2. Combination analysis of leading SNPs per pigimentary trait. On top, bar-plot showing the total count of shared or trait-exclusive leading SNPs for each trait or combination of traits. Bottom-left, diagram representing the distribution of shared/trait-exclusive leading SNPs (black dots). Bottom-right, bar-plot representing the total count of leading SNPs both shared (light purple) and trait-exclusive (dark purple) per trait. The combination analysis indicates that red hair and eye color have the higher number of trait-exclusive leading SNPs, highlighting the unique genetic architecture of these two pigimentary traits compared with the other traits analyzed.

3.3. Pigimentary traits correlation are partly explained by common genetics.

SNP heritability of the pigimentary traits, as well as genetic correlations between them, were measured using Linkage Disequilibrium Score Regression (LDSC) technique. This allows us to have an overview of the phenotypic variance explained by the SNPs, as well as, to obtain a measure of the shared genetic basis between these traits. The higher heritability was observed for hair color (e.g. brown hair: $h^2=0.42$, SE: 0.189; blond hair: $h^2=0.67$, SE: 0.381; red hair: $h^2=1.916$, SE: 2.001), phototype score ($h^2=0.28$, SE: 0.112), and Fitzpatrick scale ($h^2=0.20$, SE: 0.197).

All the traits show a high genetic correlation (Table S8, Figure S4). As expected the analysis showed a high correlation between the two phototype measures; Fitzpatrick scale and phototype score ($rg: 0.96$, se: 0.08, $p: 1.1e-32$). Furthermore; phototype measures correlate with eye and hair measures and ease sunburn was highly correlated with eye colour and phototype measures. Regarding habits, sunscreen use in exterior was correlated with sunscreen when suntan and inversely sunscreen when suntan and weekly hours in the exterior. Red hair was highly correlated with sunscreen in, sunscreen when suntan and ease sunburn, However, the low number of red hair cases in the cohort difficult the accurate estimation (Table S7). Among all, four traits remain significant; a negative correlation among ease of sunburn and phototype score ($rg: -1.15$, se: 0.25, $p: 2.88e-6$), ease of sunburn and Fitzpatrick scale ($rg: -1.43$, se: 0.52, $p: 6.08e-3$); and among brown hair and light hair (meaning brown versus black trait) with phototype score ($rg: -0.85$, se: 0.29, $p: 3.89e-3$).

3.4. Genes containing SNPs associated with pigimentary traits, show an overrepresentation of pigimentary metabolic processes, skin and eye-related diseases

To further characterize the functional impact of the SNPs identified in the GWAS analysis, we performed a gene-based analysis, with aggregated SNPs into single genes. Then, we tested the association between genes and pigimentary traits using MAGMA software [42] (Figure 3A, Table S9). After Bonferroni correction, 22 genes had a significant association with any of the pigimentary traits. Two genes, *HERC2* and *SLC45A2*, out of the 22 showed associations with skin, hair, eye and phototype. *DPEP1* and *SPATA33* were associated with skin, red hair and phototype. *MC1R*, *TYR*, *ADAMTS12*, *AMACR*, *C1QTNF3*, *CHMP1A*, *GAS8*, *RP11-1084J3.4*, *RXFP3* and *SPIRE2* were associated with skin and phototype, and *OCA2* and *TYRP1* were associated with eye and phototype traits. Six identified genes did not show association with phototype, being exclusively associated

with one trait: *C17orf112*, *DEF8*, *VPS9D1* and *ZNF276* with red hair, *SLC25A37* associated with sunscreen, and *TRPS1* with skin color. *RP11-1084J3.4*, *C1QTNF3* and *C17orf11* were not previously reported as pigimentary traits in GWAS Catalog.

Functional enrichment analysis using WebGestalt [46] revealed an overrepresentation of genes involved in the melanogenesis pathway, tyrosine metabolism, monoxygenase and glycolipid binding functions. Moreover, this analysis revealed an enrichment of genes associated with diseases such as albinism, melanoma, basal cell carcinoma, vitiligo, or visual impairment, and with Hermansky-Pudlak Syndrome (Figure 3B, Table S10, Figure S5).

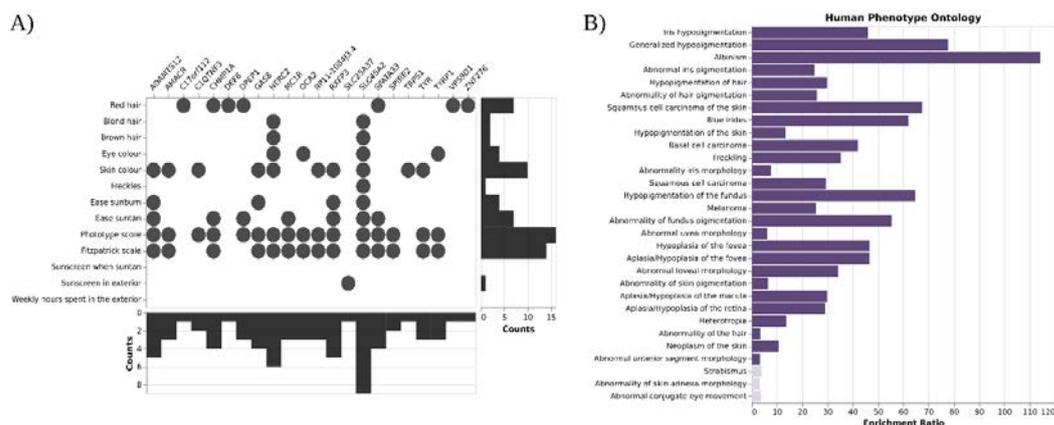


Figure 3. Results of the gene-based analysis and functional enrichment. (A) Scheme depicting the total count of associations (black dots) per gene (bottom bar plot) and trait (right-hand bar plot) identified by gene-based analysis with MAGMA software. *SLC45A2* was the most pleiotropic gene, being associated with 9 pigimentary traits, on the other hand, phototype traits were the ones with more associated genes (B) Functional enrichment analysis of the 22 aggregated genes with WebGestalt. Bar plot depicting enrichment ratio in the X axis, and the different Human Phenotype Ontologies in the Y axis. Dark color indicates that pass the FDR threshold. Pigimentary traits and diseases are significantly enriched in these genes.

3.5. Analysis of structural variants identifies small intronic deletions as the most common structural variation associated with eight pigimentary traits.

The use of the GCAT|Panel as the reference panel for imputation allowed us to impute not only SNV information, but also SVs to investigate functional consequences. A total of 15 SVs showed a suggestive significant association with at least a pigimentary trait (Figure S6, Table S11), and eight of them surpassed the genome-wide significance threshold. All Identified SVs were located in already identified risk loci and were not leading the association. For SVs impact interpretation, mapping genes were considered based on closest gene to SVs, and this could generate some discrepancy in the cited genes. Red hair and phototype (both phototype score and Fitzpatrick scale) harbors the higher number of significant loci; being the more significant, an intronic deletion in *SPIRE2* (16:89896057:G:) associated to red hair (OR=25.82, CI=13.79-48.35, $p=3e-24$), phototype score (OR=1.72e-3, CI=3.18e-4-9.39e-3, $p=2.04e-13$), Fitzpatrick scale (OR = 0.73, CI = 0.67 - 0.8, $p = 2.58e-12$) and also present in ease suntan (OR=0.62, CI=0.55-0.69, $p=3.84e-18$). Red hair was also associated with three other intronic deletions in *GAS8*, *CHMP1A*, and *CPNE7*. Other significant associated loci were two small intronic deletion associated with eye color in *HERC2* (15:28473597:A:) (OR=1.51, CI=1.39- 1.65, $p = 2e-21$) and *LU-RAP1L-AS1* (9:12756780:C:) OR=1.14, CI=1.09-1.19, $p = 2.56e-9$), both also associated with phototype score. And two intronic deletions associated with skin color; a small deletion in *BNC2* (9:16876769:G:) (OR=0.91, CI=0.88-0.94, $p = 1.05e-8$), and a middle-deletion in *SLC45A2* (5:33973937:G:<midDEL>) (OR = 0.89, CI = 0.86 - 0.92, $p = 1.6e-12$). In summary, most of the identified SV are deletions (13) and, only suggestively, one insertion

(Alu) in chromosome 6 (*DST*) and one translocation in chromosome 7 (*ZNF92*) were observed (Figure 4). The majority of the SVs map to intronic or intergenic regions (11 out of 15 SVs) (Table S12).

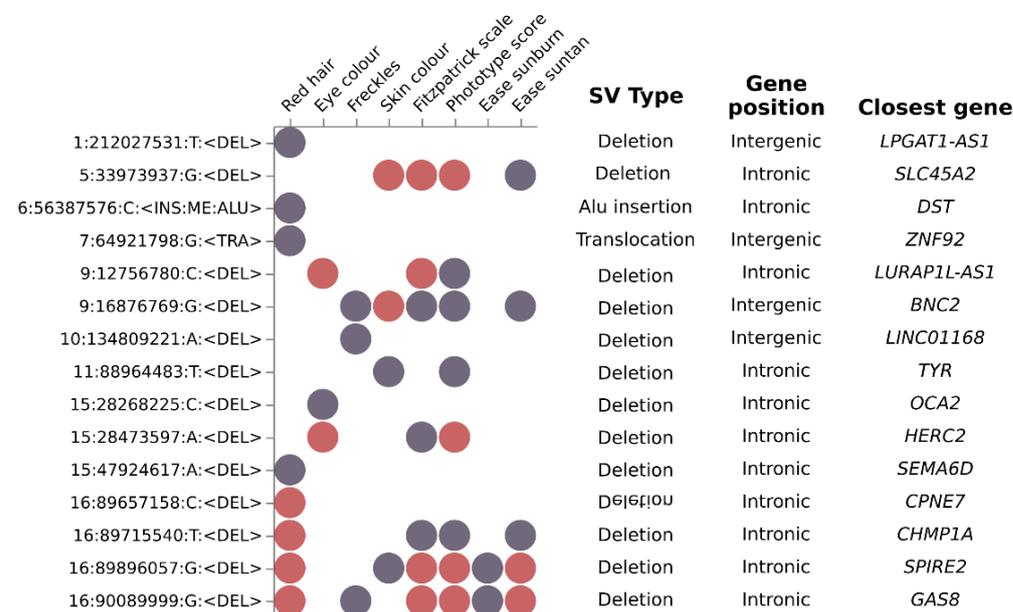


Figure 4. Genome-wide analysis identified structural variants significantly associated with a subset of pigimentary traits. Dot-plot depicting the suggestive (purple dots) and significant (red dots) association of fifteen SVs (Y axis) with a subset of eight pigimentary traits (X axis). Additional SV information are shown in SV type (deletion, Alu insertion and translocation), Gene position and Closest gene columns. Most Identified SV were deletions located in intronic regions.

3.6. PheWAS analysis confirmed the pleiotropy among pigimentary traits genes and related diseases

To assess whether the lead SNPs associated with pigimentary traits identified in our analysis have been previously associated with diseases, we performed a Phenotype Wide Association Study (PheWAS) analysis with these SNPs and using a PheWeb dataset that contains the results of associations carried out with the UK Biobank data, as described in Material and methods section. This analysis revealed that 17 out of the 99 SNPs had at least an association surpassing the suggestive significance threshold in the UKbiobank ($P=1e-5$). Amongst the diseases, those belonging to the neoplasm and dermatologic categories, were the most frequent ones (Table S13). Highly pleiotropic variants mapped in three fundamental pigimentary genes *IRF4*, *HERC2*, *TYR*; rs12203592, which is an intronic variant in *IRF4* gene, significantly associated with six of the pigimentary traits in our analysis (brown hair, ease sunburn, ease suntan, eye color, freckles and skin color), and with skin cancer and actinic keratosis. The allele T associates with fair skin pigmentation, and increasing ease of sunburn, but also increases the risk of melanoma or actinic keratosis (Figure 5A); rs12913832, an intronic variant in *HERC2* gene, significantly associated with eight of the pigimentary traits (blond hair, brown hair, ease sunburn, ease suntan, Fitzpatrick scale, phototype score and skin color), and diseases affecting the eyes, such as cataract or glaucoma. The allele G, associated with fair skin or eye color, decreases the risk of glaucoma and cataract (Figure 5B); and rs1126809, a missense variant in *TYR* gene significantly associated with three pigimentary traits (ease sunburn, Fitzpatrick scale and phototype score), and with neoplasms and bipolar disorder. Here, the allele A is associated with fair skin color, increased risk of ease of sunburn, and risk of skin cancer and bipolar disorder (Figure 5C).

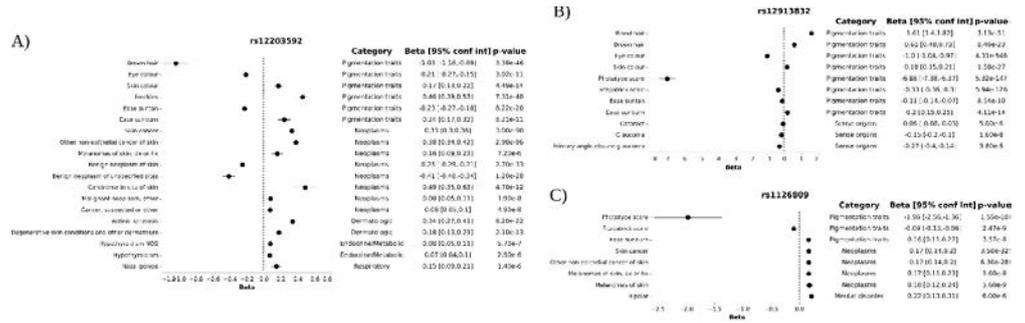


Figure 5. Pleiotropic SNPs associated with pigimentary traits and diseases. Forest plot depicting the effect size of disease-trait association of selected SNPs: (A) rs12203592 (*IRF4*), (B) rs12913832 (*HERC2*), and (C) rs1126809 (*TYR*). X axis indicated the effect size (beta) for any trait reaching suggestive threshold (Y axis). Additional information about trait category, beta, CI and p-value is shown.

3.7. Polygenic Risk Scores confirmed the association of pigimentary traits and eye diseases

To further confirm the shared genetics between pigimentary traits and diseases, we assessed the association between PRSs for these diseases with pigimentary traits. We observed significant associations between pigimentary traits and melanoma, non-melanoma skin cancer, basal cell carcinoma, vitiligo, cataracts, and overweight, obesity and hyperalimentation. Fitzpatrick scale, ease sunburn, ease suntan, sky color and freckles, shown the same association pattern for melanoma, non-melanoma skin cancer and basal cell carcinoma, and the opposite for vitiligo and cataracts. No clear pattern was observed for overweight. Individuals with a fair phenotype present a higher PRS for melanoma, non-melanoma skin cancer and basal cell carcinoma (Figure 6A, B and C), and a trend for increased risk overweight, obesity and hyperalimentation (Figure 6D). Individuals with increased risk overweight, obesity and hyperalimentation (Figure 6D). Individuals with dark phototype present a higher PRS for vitiligo and cataracts (Figure 6E and F). No significant associations were observed for glaucoma, psoriasis, and essential hypertension. Individuals that have more skin sensitivity (ease sunburn) have increased risk for melanoma, non-melanoma skin cancer, basal cell carcinoma, and overweight. An inverse association with ease suntan was also observed for all of them.

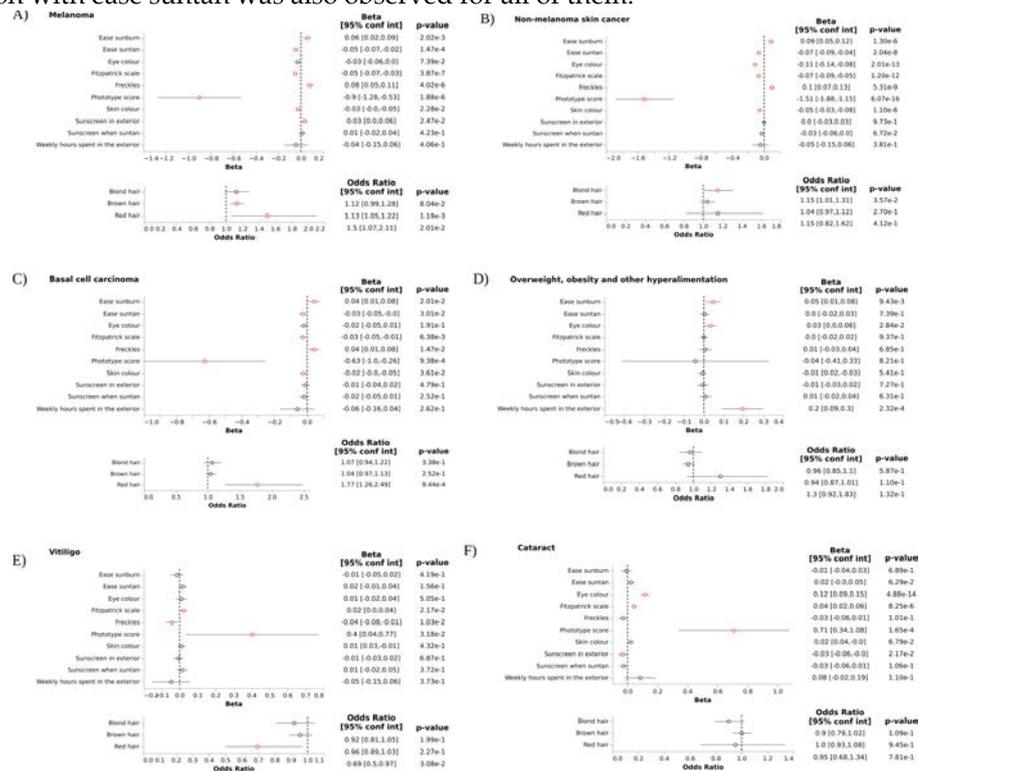


Figure 6. Association of PRS of skin and sense organs diseases with pigmentary traits. Forest plot depicting the effect size of significant PRS (A-F). Upper plots, depicting the beta value and the 95% CI of discrete and continuous variables, and bottom plots, depicting the odds ratio and the 95% CI of binary variables, for each regression between PRS and pigmentary traits. X axis indicates the effect size for each pigmentary trait (Y axis). Additional information about beta, CI and p-value is shown. Associations reaching the nominal significance threshold are highlighted in red.

4. Discussion

The fact that differences among individuals and populations are mainly due to the genetic variation component of the human genome is also extensible to pigmentary traits [51], making it plausible that certain pathologies, previously associated to different ancestries, found their cause in the onset of a single mutation, a deletion, or an inversion that is fixed over the time together with pigmentary traits. Most GWAS, of pigmentary traits, studied North European cohorts [52-54], remarking on the need for more studies in a geographically distinct population. In this context, we present a GWAS study of 13 pigmentary traits (constitutive/facultative pigmentation, and behavioral related traits) in a South European cohort from Catalonia (NE of Spain).

Our study confirmed previous findings in North European populations. Fair phototypes are significantly at more risk of developing skin-related diseases, such as: neoplasm and dermatosis; and to a lesser extent, benign neoplasms, dermatitis, eczema, and erythema. This predisposition has been largely related to the pheomelanin present in keratinocytes of fair phototypes, with a truthful mechanistic explanation by; (i) an overall increase of DNA damage that overloads the DNA repair machinery [55], (ii) release of reactive oxygen species (ROS) that induce more DNA damage [56] and (iii) immunosuppressive effect of UVR [57].

In addition, our study highlighted that fair skin phototype was also associated vitiligo, cataracts, glaucoma and psoriasis, and with two common conditions non primarily related with skin color: obesity and essential hypertension. Obesity represents a significant health problem and together with hypertension significantly increases the risks of suffering from other major diseases affecting vital organs as heart, brain, kidney, and other diseases. Several authors suggested that deficiency of vitamin D, may be linked to obesity, hypertension, hyperlipidemia, and diabetes, increasing metabolic and cardiovascular risk [58, 59], thus relating these conditions to functional ability to synthesize vitamin D. This hypothesis reinforces the role of geographical location and lifestyle factors (diet, physical activity, clothes, etc) related with the association observed, but also be conditioned by the skin phototype, and the close relation with vitamin D metabolism. Preliminary genetic analysis of vitamin D shown a genetically determined basis in the PheWas analysis, but no able to show its association with phototype in our cohort (data not shown). From our data, we cannot extrapolate such a conclusion, and more refined attention is needed to identify the exact mechanisms mediating this association.

When looking at genes, we observed 16 genes associated with one or more pigmentary traits and diseases. Six of them have been largely associated with melanogenesis, and pigmentary traits, in different ethnic groups (*MC1R*, *HERC2*, *OCA2*, *TYR*, *TYRP1*, *SLC45A2*) [39, 60, 61]; other have been previously reported to be related with: hair color, hair measurement, suntan, sunburn, but also with non-melanoma skin carcinoma, cutaneous melanoma, pigmentary retinopathy and oculocutaneous albinism (*ADAMTS12*, *AMACR*, *CHMP1A*, *DPEP1*, *GAS8*, *RXFP3*, *SPATA33* and *SPIRE2*). Our analysis identified for the first time, to our knowledge three novel candidate genes (*RP11-1084J3.4*, *C1QTNF3* and *C17orf112*). *RP11-1084J3.4*, mapping on chromosome 5, also known as *C1QTNF3-AMACR*, is a read-through transcript including exons from genes *C1QTNF3* and *AMACR*, and it is not likely to produce a protein product [62]. This gene shares locus with its parent gene *C1QTNF3* (Complement C1q Tumor Necrosis Factor-Related Protein 3), both being associated with skin color and phototype in the present study; and with *AMACR* (alpha-methylacyl-CoA racemase), the other parent gene which is reported to be associated with hair color [17], and we found it to be associated with skin color and phototype, revealing

an interesting locus related with pigmentation traits to further study. *C17orf112* located in chromosome 17, also known as *LINC02876*, is a long non-coding RNA, associated with red hair in our study, a trait with a low prevalence in our cohort, making necessary further validation. Some of the identified lead SNPs, mapping *IRF4*, *HERC2*, and *TYR*, were also found to be associated with melanoma or other skin diseases in the UKBiobank. Mutations in the *MC1R* gene lead to fair skin and red hair in humans, which is also seen with inactivating human *POMC* gene mutations. The powerful effects of proopiomelanocortin (POMC) and probably Corticotropin-releasing hormone (CRH) on the skin pigmentary, immune, and adnexal systems are consistent with stress-neutralizing activity addressed at maintaining skin integrity to restrict disruptions of internal homeostasis. This CRH/POMC skin system appears to generate a function analogous to the HPA axis, that in the skin is expressed as a highly localized response which neutralizes noxious stimuli and attendant immune reactions [63, 64]. Alterations of this system linked to phototype could lead to a different immune status related to some observational associations with diseases of immune component as psoriasis.

While most of the GWAS studies are restricted to SNVs and small indels, our study incorporates large structural variants in the analysis [49]. We identified 15 SVs, mostly small deletions, that were at least at the suggestive significance level for any of the pigmentary traits, mostly located in intronic or intergenic regions of genes known to influence pigmentary traits (*SLC45A2*, *TYR*, *OCA2* and *HERC2*), *CHMP1A* and *GAS8*, both close to *MC1R*, associated with red hair [17] and hair color [65] respectively, or potentially involved with new candidate genes, as *BNC2*, a transmission factor of keratinocytes, and *SPIRE2* involved in the transport of intracellular vesicles [66, 67], *SEMA6D*, a member of the semaphorin family showing a suggestive association with hair greying [68] and *CPNE7* associated with skin tanning [54]. Unfortunately, most of the observed SV do not map in any known functional regulatory element to attribute a functional impact. However, some identified regions, as the observed SV mapping at *LURAP1L-AS1*, are interestingly linked to melanin genes. *LURAP1L-AS1* is an antisense RNAs, a unique DNA transcript, small, noncoding, and diffusible molecule that complements mRNA, and are recognized as intracellular gene regulators. *LURAP1L-AS1* gene overlaps *TYRP1*, involved in eumelanin synthesis [69, 70], and suggest a role in *TYRP1* transcript regulation and putative melanin dysregulation.

Complementary to single SNP analysis, gene-based analyses provided additional power due to the aggregate effect of multiple SNPs being larger than that of individual SNPs. The functional enrichment analysis revealed an overrepresentation of defective pigmentation diseases involving altered pigment metabolic processes (albinism, lentigines), involved in melanogenesis, or tyrosine metabolism, involving monoxygenase activity or the glycolipid binding functions, probably related to melanin, and bile acids biosynthesis. Glycosphingolipid metabolism is enriched in the gene dataset, and glycosphingolipids have been reported essential for a protein-sorting step in the Golgi complex, and being suggested that the Hermansky-Pudlak syndrome, a rare disease with associated albinism have their origin in the altered lipid metabolism [71]. Indeed, a strong reduction of melanin production and reduced tyrosinase activity in melanocytes of both light and dark skin type origin have been reported in cell cultured human melanocytes with lipid metabolism inhibitors. Other enriched pathways include well-reported anti-oxidant molecules, glutathione (GSH). Cellular antioxidant systems, such as the thioredoxin (Trx) and glutathione (GSH) systems, function to reduce oxidative stress. Therefore, the antioxidant defense systems play a vital role in maintaining an optimal redox balance in melanocytes by quenching ROS and protecting against oxidative stress, excessive melanogenesis and photo-damaged skin. Furthermore, our results show an overrepresentation of genes linked to key cellular components as pigment granules, or ESCRT complex, that are required for *Typr1* transport from early endosomes to the melanosome limiting membrane [72]. In addition, it is worth noting that our identified genes are enriched in genes involved in several eye

diseases, such as photophobia, myopia, hyperopia, retinal degeneration, or fovea hypoplasia. Foveal hypoplasia is an ocular abnormality that is typically associated with other ocular findings as nystagmus, aniridia, cataract, and skin hypopigmentation.

Pleiotropy analysis confirmed shared genetics for pigimentary traits, melanoma, and non-melanoma skin cancer. The results suggest a higher risk of skin cancer in individuals with a fair phototype, consequent with a protective effect from UVR damage to the skin and the ability of melanin to absorb from 50% to 75% of UVR, but also could be related to the described pathological role of melanin, linking a defective secretion mechanism to an increase in tumor severity [35]. In addition, vitiligo, cataracts, and overweight shared genetics with pigimentary traits, being darker phototypes at higher risk of vitiligo, cataracts and overweight. In the case of vitiligo, this result supports the observed different prevalence across latitudes, having a higher prevalence in populations with darker skin color [73]. The mechanism of vitiligo is still unclear but, an indirect effect of the deficient levels of vitamin-D production, could alter immune status, thus being involved in autoimmune dysregulation. Another example of immune disease for which a deficit of vitamin-D has been largely reported is psoriasis [74], however we do not observe any phototype association in our dataset when analyzed. Regarding increased risk of cataracts in individuals having darker phototypes, the results confirm previous epidemiological observational that have shown an association between dark eyes and cataracts [75]. Plausible mechanisms could be related to the increase in the temperature of the lens when dispersing the UVR [76]. However, reversal association could not be discarded, since cataracts are caused by the degradation of proteins and fibers of the lens, it may be linked to behavioral habits, as spending too much time in the sun without sunglasses, which is more common in individuals with a darkness phototype based on a self-perception of security. In the case of the overweight the association, darker color, ease sunburn and weekly hours spent in the exterior are associated with no apparent mechanistic explanation but linked to a more complex behavior. Indeed, from the 13 included pigimentary traits, three phenotypes ("sunscreen in the exterior", "sunscreen when suntanning", "weekly hours spend in the exterior") did not show any gene association. These variables are defined as behavioral habits, more related to social and self-risk-perception than to a direct biological on skin function, and probably are prone to reversal causation in the association analysis. Obesity is responsible for changes in skin barrier function, wound healing, microcirculation, and subcutaneous fat. Moreover, obesity is implicated in a large number of dermatological conditions (see review in Yosipovitch et al. [77]). Interestingly, Sultan et al. have reported that after UV injury, the induction of the melanogenic phase activates fatty acid biosynthesis results in the formation of TAGs, suggesting that an alteration of the fatty acid metabolism being involved in a variety of cutaneous diseases manifesting hyper pigimentary phenotype [78].

Our results provide a comprehensive analysis of the genetic basis of pigimentary traits in a large South European population and offer a landscape of their genetic basis for a better understanding of human pigmentation pleiotropy based on the common genetic background to some previously observed associations. Furthermore, our findings support the importance of incorporating population information to adapt personalized care approaches to consider differential risk in general population, being of great importance in modern mixed populations. It is well accepted that natural selection shaped the skin pigmentation phenotype, resulting, in a gradient of variation, across populations, directly correlated with absolute latitude [13, 14, 79]. Understanding the genetic basis of adaptation of skin color in different populations have many implications for human evolution and medicine [13]. Among the limitations of the study, we need to consider that the cohort, even if there is a wide distribution of phototypes, there may be some underrepresented ancestries. Also, self-reported questionnaires could introduce some bias, due to self-perception of ancestry or social pressure. And lastly, the partial genetic analysis conducted, where a complete analysis including all the cohort and more detailed genetic data, such as rare variants coming from WGS or WES, would be needed to identify more spe-

cific private population effects. Phototype could be a tool to define preventing health strategies in general population, however since reversal association is possible, further detailed studies are needed to understand the role of pigmentation traits in health and disease.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Scheme S1: Pipeline of the study; Figure S1: Pigmentation traits distribution; Figure S2: Phototype associated diagnosis; Figure S3: Genome-wide association study plots of the 13 pigmentation traits in GCATcore (n = 4988) plotted using the general linear model and mixed model methods; Figure S4: Genetic Correlations among the 13 pigmentation traits; Figure S5: Functional Enrichment in significant genes from gene-based analysis; Figure S6: LocusZoom of genome-wide association analysis of structural variants (n = 15) and trait combination reaching the suggestive threshold (n = 38); Figure S7: Association of PRS of skin and sense organs diseases with pigmentation traits; Table S1: Pigmentation traits definitions and number of individuals in the GWAS; Table S2: Questions and scores used to calculate phototype score and Fitzpatrick scale; Table S3: Phototype score thresholds used to define Fitzpatrick scale phenotype; Table S4: Disease PRS used from the PGS catalog; Table S5: Results of the logistic regression of 177 ICD-9 diagnosis with Fitzpatrick scale, corrected by age and gender; Table S6: Independent significant hits, lead SNPs and genomic risk loci of each pigmentation trait GWAS; Table S7: SNP heritability of the pigmentation traits computed with LDSC; Table S8: Pairwise genetic correlations of the pigmentation traits computed with LDSC; Table S9: Results of MAGMA, the gene-based test associations; Table S10: Functional enrichment performed with WebGestalt and the genes resulting from MAGMA; Table S11: Annotation of the SV that showed at least a suggestive association ($p < 1e-5$) in any of the pigmentation traits; Table S12: Association of the lead SNPs with diseases from the UKBiobank; Table S13: Results of the association between the PRS of the diseases and the pigmentation traits.

Author Contributions: Conceptualization, R.dC.; Methodology, X.F., N.B., B.C. & R.dC.; Formal Analysis, X.F. & N.B.; Investigation, X.F., N.B., B.C., S.I-G. & R.dC. ; Data Curation, N.B.; Writing – Original Draft Preparation, X.F., N.B., B.C. & R.dC.; Writing – Review & Editing, all authors.; Visualization, B.C.; Supervision, R.dC. & S.I-G.; Project Administration, R.dC. & S.I-G; Funding Acquisition, R.dC. All authors have read and agreed to the published version of the manuscript.

Funding: This study makes use of data generated by the GCAT-Genomes for Life. Cohort study of the Genomes of Catalonia, Fundacio IGTP. IGTP is part of the CERCA Program / Generalitat de Catalunya. GCAT was funded by Acció de Dinamització del ISCIII-MINECO and the Ministry of Health of the Generalitat of Catalunya (ADE 10/00026); and have additional support by the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) (2017-SGR 529), National Grant PI18/01512 and VEIS project (001-P-001647) (co-funded by European Regional Development Fund (ERDF), "A way to build Europe"). This study was carried out using anonymized data provided by the Catalan Agency for Quality and Health Assessment, within the framework of the PADRIS Program. The authors of this study would like to acknowledge all GCAT project investigators who contributed to the generation of the GCAT data. A full list of the investigators is available from www.genomesforlife.com/. We thank the Blood and Tissue Bank from Catalonia (BST) and all the GCAT volunteers that participated in the study. Xavier Farré is supported by VEIS project (001-P-001647) (co-funded by European Regional Development Fund (ERDF), "A way to build Europe"). Beatriz Cortés is supported by national grants PI18/01512.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee for Research with Medical Products of Hospital Germans Trias I Pujol (CEIC-HGTP) (protocol code GCAT, REF. CEIC PI-13-020 and date of approval 26/04/2013).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: All data is available at EGA (Dataset ID EGAD00010001665 and EGAD00010002153).

Acknowledgments: The authors thank all the GCAT participants and all BST members for generously helping with this research. The authors of this study would like to acknowledge all GCAT project investigators who contributed to the generation of the GCAT data. A full list of the investigators is available from www.genomesforlife.com/. We thank the Blood and Tissue Bank from Catalonia (BST) and all the GCAT volunteers that participated in the study.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Liu Y, Hong L, Wakamatsu K, Ito S, Adhyaru B, Cheng CY, Bowers CR, Simon JD: Comparison of structural and chemical properties of black and red human hair melanosomes. *Photochem Photobiol* **2005**, *81*:135-144.
- Matts PJ, Dykes PJ, Marks R: The distribution of melanin in skin determined in vivo. *Br J Dermatol* **2007**, *156*:620-628.
- Sturm RA, Frudakis TN: Eye color: portals into pigmentation genes and ancestry. *Trends Genet* **2004**, *20*:327-332.
- Thong HY, Jee SH, Sun CC, Boissy RE: The patterns of melanosome distribution in keratinocytes of human skin as one determining factor of skin color. *Br J Dermatol* **2003**, *149*:498-505.
- Wielgus AR, Sarna T: Melanin in human irides of different color and age of donors. *Pigment Cell Res* **2005**, *18*:454-464.
- Graf J, Hodgson R, van Daal A: Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation. *Hum Mutat* **2005**, *25*:278-284.
- Yamaguchi Y, Brenner M, Hearing VJ: The regulation of skin pigmentation. *J Biol Chem* **2007**, *282*:27557-27561.
- Wilson PD, Kligman AM: Do freckles protect the skin from actinic damage? *Br J Dermatol* **1982**, *106*:27-32.
- Cordero RJB, Casadevall A: Melanin. *Curr Biol* **2020**, *30*:R142-R143.
- Sakai D, Takagi S, Totani K, Yamamoto M, Matsuzaki M, Yamanari M, Sugiyama S, Yokota S, Maeda A, Hiramami Y, et al: Retinal pigment epithelium melanin imaging using polarization-sensitive optical coherence tomography for patients with retinitis pigmentosa. *Sci Rep* **2022**, *12*:7115.
- Scherer D, Kumar R: Genetics of pigmentation in skin cancer--a review. *Mutat Res* **2010**, *705*:141-153.
- Hysi PG, Valdes AM, Liu F, Furlotte NA, Evans DM, Bataille V, Visconti A, Hemani G, McMahon G, Ring SM, et al: Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability. *Nat Genet* **2018**, *50*:652-656.
- Deng L, Xu S: Adaptation of human skin color in various populations. *Hereditas* **2018**, *155*:1.
- Lao O, de Gruijter JM, van Duijn K, Navarro A, Kayser M: Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. *Ann Hum Genet* **2007**, *71*:354-369.
- Makova K, Norton H: Worldwide polymorphism at the MC1R locus and normal pigmentation variation in humans. *Peptides* **2005**, *26*:1901-1908.
- Valverde P, Healy E, Jackson I, Rees JL, Thody AJ: Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. *Nat Genet* **1995**, *11*:328-330.
- Morgan MD, Pairo-Castineira E, Rawlik K, Canela-Xandri O, Rees J, Sims D, Tenesa A, Jackson IJ: Genome-wide study of hair color in UK Biobank explains most of the SNP heritability. *Nat Commun* **2018**, *9*:5271.
- White D, Rabago-Smith M: Genotype-phenotype associations and human eye color. *J Hum Genet* **2011**, *56*:5-7.
- Lamason RL, Mohideen MA, Mest JR, Wong AC, Norton HL, Aros MC, Jurynec MJ, Mao X, Humphreville VR, Humbert JE, et al: SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* **2005**, *310*:1782-1786.
- Adhikari K, Fontanil T, Cal S, Mendoza-Revilla J, Fuentes-Guajardo M, Chacon-Duque JC, Al-Saadi F, Johansson JA, Quinto-Sanchez M, Acuna-Alonzo V, et al: A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features. *Nat Commun* **2016**, *7*:10815.
- Liu F, Wen B, Kayser M: Colorful DNA polymorphisms in humans. *Semin Cell Dev Biol* **2013**, *24*:562-575.
- Meyer OS, Lunn MMB, Garcia SL, Kjaerbye AB, Morling N, Borsting C, Andersen JD: Association between brown eye color in rs12913832:GG individuals and SNPs in TYR, TYRP1, and SLC24A4. *PLoS One* **2020**, *15*:e0239131.
- Nan H, Kraft P, Hunter DJ, Han J: Genetic variants in pigmentation genes, pigmentary phenotypes, and risk of skin cancer in Caucasians. *Int J Cancer* **2009**, *125*:909-917.
- MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, et al: The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* **2017**, *45*:D896-D901.
- Gomez-Acebo I, Dierssen-Sotos T, Palazuelos C, Fernandez-Navarro P, Castano-Vinyals G, Alonso-Molero J, Urriaga C, Fernandez-Villa T, Ardanaz E, Rivas-Del-Fresno M, et al: Pigmentation phototype and prostate and breast cancer in a select Spanish population--A Mendelian randomization analysis in the MCC-Spain study. *PLoS One* **2018**, *13*:e0201750.
- Hernando B, Sanz-Page E, Pitarch G, Mahiques L, Valcuende-Cavero F, Martinez-Cadenas C: Genetic variants associated with skin photosensitivity in a southern European population from Spain. *Photodermatol Photoimmunol Photomed* **2018**, *34*:415-422.
- Rocha J: The Evolutionary History of Human Skin Pigmentation. *J Mol Evol* **2020**, *88*:77-87.
- Beleza S, Santos AM, McEvoy B, Alves I, Martinho C, Cameron E, Shriver MD, Parra EJ, Rocha J: The timing of pigmentation lightening in Europeans. *Mol Biol Evol* **2013**, *30*:24-35.
- Jablonski NG, Chaplin G: Human skin pigmentation, migration and disease susceptibility. *Philos Trans R Soc Lond B Biol Sci* **2012**, *367*:785-792.
- Amos CI, Wang LE, Lee JE, Gershenwald JE, Chen WV, Fang S, Kosoy R, Zhang M, Qureshi AA, Vattathil S, et al: Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. *Hum Mol Genet* **2011**, *20*:5012-5023.
- Duffy DL, Zhao ZZ, Sturm RA, Hayward NK, Martin NG, Montgomery GW: Multiple pigmentation gene polymorphisms account for a substantial proportion of risk of cutaneous malignant melanoma. *J Invest Dermatol* **2010**, *130*:520-528.

32. Ibarrola-Villava M, Hu HH, Guedj M, Fernandez LP, Descamps V, Basset-Seguín N, Bagot M, Bensussan A, Saiag P, Fargnoli MC, et al: MC1R, SLC45A2 and TYR genetic variants involved in melanoma susceptibility in southern European populations: results from a meta-analysis. *Eur J Cancer* **2012**, *48*:2183-2191.
33. Jorgenson E, Choquet H, Yin J, Hoffmann TJ, Banda Y, Kvale MN, Risch N, Schaefer C, Asgari MM: Genetic ancestry, skin pigmentation, and the risk of cutaneous squamous cell carcinoma in Hispanic/Latino and non-Hispanic white populations. *Commun Biol* **2020**, *3*:765.
34. Spritz RA, Andersen GH: Genetics of Vitiligo. *Dermatol Clin* **2017**, *35*:245-255.
35. Cabaco LC, Tomas A, Pojo M, Barral DC: The Dark Side of Melanin Secretion in Cutaneous Melanoma Aggressiveness. *Front Oncol* **2022**, *12*:887366.
36. Kuan V, Martineau AR, Griffiths CJ, Hypponen E, Walton R: DHCR7 mutations linked to higher vitamin D status allowed early human migration to northern latitudes. *BMC Evol Biol* **2013**, *13*:144.
37. Obon-Santacana M, Vilardell M, Carreras A, Duran X, Velasco J, Galvan-Femenia I, Alonso T, Puig L, Sumoy L, Duell EJ, et al: GCAT | Genomes for life: a prospective cohort study of the genomes of Catalonia. *BMJ Open* **2018**, *8*:e018324.
38. Fitzpatrick TB: The Validity and Practicality of Sun-Reactive Skin Types I Through VI. *Arch Dermatol* **1988**:869.
39. Galvan-Femenia I, Obon-Santacana M, Pineyro D, Guindo-Martinez M, Duran X, Carreras A, Pluvinet R, Velasco J, Ramos L, Ausso S, et al: Multitrait genome association analysis identifies new susceptibility genes for human anthropometric variation in the GCAT cohort. *J Med Genet* **2018**, *55*:765-778.
40. Valls-Margarit J, Galvan-Femenia I, Matias-Sanchez D, Blay N, Puiggros M, Carreras A, Salvoro C, Cortes B, Amela R, Farre X, et al: GCAT | Panel, a comprehensive structural variant haplotype map of the Iberian population from high-coverage whole-genome sequencing. *Nucleic Acids Res* **2022**, *50*:2464-2479.
41. Agència de Qualitat i Avaluació Sanitàries de Catalunya (AQuAS) [<https://Aguas.Gencat.Cat>]
42. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **2007**, *81*:559-575.
43. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM: A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **2012**, *6*:80-92.
44. de Leeuw CA, Mooij JM, Heskes T, Posthuma D: MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **2015**, *11*:e1004219.
45. Watanabe K, Taskesen E, van Bochoven A, Posthuma D: Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **2017**, *8*:1826.
46. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B: WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* **2019**, *47*:W199-W205.
47. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, Patterson N, Daly MJ, Price AL, Neale BM: LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **2015**, *47*:291-295.
48. Gagliano Taliun SA, VandeHaar P, Boughton AP, Welch RP, Taliun D, Schmidt EM, Zhou W, Nielsen JB, Willer CJ, Lee S, et al: Exploring and visualizing large-scale genetic associations by using PheWeb. *Nat Genet* **2020**, *52*:550-552.
49. UKBiobank TOPMed-imputed PheWeb [<https://Pheweb.Org/UKB-TOPMed/>]
50. Lambert SA, Gil L, Jupp S, Ritchie SC, Xu Y, Buniello A, McMahon A, Abraham G, Chapman M, Parkinson H, et al: The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nat Genet* **2021**, *53*:420-425.
51. Rotimi CN, Jorde LB: Ancestry and disease in the age of genomic medicine. *N Engl J Med* **2010**, *363*:1551-1558.
52. Jin Y, Andersen G, Yorgov D, Ferrara TM, Ben S, Brownson KM, Holland PJ, Birlea SA, Siebert J, Hartmann A, et al: Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. *Nat Genet* **2016**, *48*:1418-1424.
53. Liu F, Visser M, Duffy DL, Hysi PG, Jacobs LC, Lao O, Zhong K, Walsh S, Chaitanya L, Wollstein A, et al: Genetics of skin color variation in Europeans: genome-wide association studies with functional follow-up. *Hum Genet* **2015**, *134*:823-835.
54. Visconti A, Duffy DL, Liu F, Zhu G, Wu W, Chen Y, Hysi PG, Zeng C, Sanna M, Iles MM, et al: Genome-wide association study in 176,678 Europeans reveals genetic loci for tanning response to sun exposure. *Nat Commun* **2018**, *9*:1684.
55. Sinha RP, Hader DP: UV-induced DNA damage and repair: a review. *Photochem Photobiol Sci* **2002**, *1*:225-236.
56. Tanaka H, Yamashita Y, Umezawa K, Hirobe T, Ito S, Wakamatsu K: The Pro-Oxidant Activity of Pheomelanin is Significantly Enhanced by UVA Irradiation: Benzothiazole Moieties Are More Reactive than Benzothiazine Moieties. *Int J Mol Sci* **2018**, *19*.
57. Bernard JJ, Gallo RL, Krutmann J: Photoimmunology: how ultraviolet radiation affects the immune system. *Nat Rev Immunol* **2019**, *19*:688-701.
58. Vranic L, Mikolasevic I, Milic S: Vitamin D Deficiency: Consequence or Cause of Obesity? *Medicina (Kaunas)* **2019**, *55*.
59. Azevedo M, Bandeira L, Luza C, Lemos A, Bandeira F: Vitamin D Deficiency, Skin Phototype, Sun Index, and Metabolic Risk Among Patients with High Rates of Sun Exposure Living in the Tropics. *J Clin Aesthet Dermatol* **2018**, *11*:15-18.
60. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Magnusson KP, Manolescu A, Karason A, Palsson A, Thorleifsson G, et al: Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet* **2007**, *39*:1443-1452.

61. Zhang M, Song F, Liang L, Nan H, Zhang J, Liu H, Wang LE, Wei Q, Lee JE, Amos CI, et al: Genome-wide association studies identify several new loci associated with pigmentation traits and skin cancer risk in European Americans. *Hum Mol Genet* **2013**, *22*:2948-2959.
62. Prakash T, Sharma VK, Adati N, Ozawa R, Kumar N, Nishida Y, Fujikake T, Takeda T, Taylor TD: Expression of conjoined genes: another mechanism for gene regulation in eukaryotes. *PLoS One* **2010**, *5*:e13284.
63. Wintzen M, Gilchrist BA: Proopiomelanocortin, its derived peptides, and the skin. *J Invest Dermatol* **1996**, *106*:3-10.
64. Millington GW: Proopiomelanocortin (POMC): the cutaneous roles of its melanocortin products and receptors. *Clin Exp Dermatol* **2006**, *31*:407-412.
65. Gerstenblith MR, Shi J, Landi MT: Genome-wide association studies of pigmentation and skin cancer: a review and meta-analysis. *Pigment Cell Melanoma Res* **2010**, *23*:587-606.
66. Schuh M: An actin-dependent mechanism for long-range vesicle transport. *Nat Cell Biol* **2011**, *13*:1431-1436.
67. Visser M, Palstra RJ, Kayser M: Human skin color is influenced by an intergenic DNA polymorphism regulating transcription of the nearby BNC2 pigmentation gene. *Hum Mol Genet* **2014**, *23*:5750-5762.
68. Pospiech E, Kukla-Bartoszek M, Karłowska-Pik J, Zielinski P, Wozniak A, Boron M, Dabrowski M, Zubanska M, Jarosz A, Grzybowski T, et al: Exploring the possibility of predicting human head hair greying from DNA using whole-exome and targeted NGS data. *BMC Genomics* **2020**, *21*:538.
69. D'Mello SA, Finlay GJ, Baguley BC, Askarian-Amiri ME: Signaling Pathways in Melanogenesis. *Int J Mol Sci* **2016**, *17*.
70. Gautron A, Migault M, Bachelot L, Corre S, Galibert MD, Gilot D: Human TYRP1: Two functions for a single gene? *Pigment Cell Melanoma Res* **2021**, *34*:836-852.
71. Sprong H, Degroote S, Claessens T, van Drunen J, Oorschot V, Westerink BH, Hirabayashi Y, Klumperman J, van der Sluijs P, van Meer G: Glycosphingolipids are required for sorting melanosomal proteins in the Golgi complex. *J Cell Biol* **2001**, *155*:369-380.
72. Truschel ST, Simoes S, Setty SR, Harper DC, Tenza D, Thomas PC, Herman KE, Sackett SD, Cowan DC, Theos AC, et al: ESCRT-I function is required for Tyrp1 transport from early endosomes to the melanosome limiting membrane. *Traffic* **2009**, *10*:1318-1336.
73. Zhang Y, Cai Y, Shi M, Jiang S, Cui S, Wu Y, Gao XH, Chen HD: The Prevalence of Vitiligo: A Meta-Analysis. *PLoS One* **2016**, *11*:e0163806.
74. Barrea L, Savanelli MC, Di Somma C, Napolitano M, Megna M, Colao A, Savastano S: Vitamin D and its role in psoriasis: An overview of the dermatologist and nutritionist. *Rev Endocr Metab Disord* **2017**, *18*:195-205.
75. Leske MC, Chylack LT, Jr., Wu SY: The Lens Opacities Case-Control Study. Risk factors for cataract. *Arch Ophthalmol* **1991**, *109*:244-251.
76. Cumming RG, Mitchell P, Lim R: Iris color and cataract: the Blue Mountains Eye Study. *Am J Ophthalmol* **2000**, *130*:237-238.
77. Yosipovitch G, DeVore A, Dawn A: Obesity and the skin: skin physiology and skin manifestations of obesity. *J Am Acad Dermatol* **2007**, *56*:901-916; quiz 917-920.
78. Sultan F, Basu R, Murthy D, Kochar M, Attri KS, Aggarwal A, Kumari P, Dnyane P, Tanwar J, Motiani RK, et al: Temporal analysis of melanogenesis identifies fatty acid metabolism as key skin pigment regulator. *PLoS Biol* **2022**, *20*:e3001634.
79. Jablonski NG, Chaplin G: The evolution of human skin coloration. *J Hum Evol* **2000**, *39*:57-106.