



Roberts, V., Main, B., Timpson, N. J., & Haworth, S. (Accepted/In press). Genome wide association study identifies genetic associations with perceived age. *Journal of Investigative Dermatology*.
<https://doi.org/10.1016/j.jid.2020.03.970>

Peer reviewed version

Link to published version (if available):
[10.1016/j.jid.2020.03.970](https://doi.org/10.1016/j.jid.2020.03.970)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Elsevier at <https://www.sciencedirect.com/science/article/abs/pii/S0022202X20314019#!>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Informative Title: Genome wide association study identifies genetic associations with perceived age.

Authors:

Victoria Roberts¹, ORCID ID: <https://orcid.org/0000-0001-7540-1579>

Barry Main^{1,2}, ORCID ID: <https://orcid.org/0000-0003-0622-805X>

Nicholas J. Timpson³, ORCID ID: <https://orcid.org/0000-0002-7141-9189>,

Simon Haworth^{2,3} ORCID ID: <https://orcid.org/0000-0001-7793-7326>,

1. Bristol Medical School, University of Bristol, Bristol, UK
2. Bristol Dental School, University of Bristol, Bristol, UK
3. MRC Integrative Epidemiology Unit, Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

Location: Bristol, United Kingdom

Corresponding author: Simon Haworth, University of Bristol Dental School, Lower Maudlin Street, Bristol, BS12LY, UK, simon.haworth@bristol.ac.uk

Short title (45 characters): Genetic associations with perceived age

ABSTRACT

Failure of dermal protection or repair mechanisms might lead to visibly aged skin. The study aimed to identify genetic associations with perceived age. A genome-wide association study was undertaken in 423,992 adult participants of UK Biobank, using questionnaire data on perceived age and genetic data imputed to the Haplotype Reference Consortium imputation panel. Seventy-four to our knowledge previously unreported and independent associated genetic loci were identified ($P < 5 \times 10^{-8}$), which were enriched for cell signaling pathways including the *NEK6* and *SMAD2* subnetworks. Common genetic variation was estimated to account for 14% of variation in perceived age and the heritability of perceived age was partially shared with that of 75 other traits including multiple traits representing adiposity, suggesting that perceived age may be a useful proxy trait in genetic association studies.

INTRODUCTION

Skin is the interface between the internal and external environment and has functions to prevent and repair damage from exogenous factors such as ultraviolet (UV) light and bacteria (Lee et al., 2006). Studies which characterize the biological mechanisms underlying normal skin barrier and repair functions may provide insight into diseases where these mechanisms fail (Williams, 2005) such as atopy, proxy phenotypes for skin response to exogenous factors and improve understanding of these mechanisms.

One challenge is the ability of skin to respond to UV light, which has been explored using genome-wide association studies (GWAS) for proxy phenotypes such as self-reported tanning ability (Nan et al., 2009, Visconti et al., 2018), identifying a range of associated loci.

Objective or subjective measures of skin age might represent the ability of skin to respond to a wider range of environmental challenges. Studies have investigated genetic factors

associated with skin age, rationalizing that failure of photoprotective or other dermal integrity mechanisms will lead to visibly aged skin (Law M. H. et al., 2017, Liu et al., 2016). The results of these studies and GWAS investigating self-reported tanning identified association at the melanocortin 1 receptor (*MC1R*), which regulates pigmentation. Other genes identified have been linked to skin and hair pigmentation, poor tanning ability, increased freckling and skin cancers (Duffy et al., 2004, Duffy et al., 2010, Han et al., 2008, Han et al., 2011, Kita and Fraser, 2016, López et al., 2014, Sulem et al., 2007, Zhang M. et al., 2013). This apparent similarity in findings for objective and subjective measures of skin function is mirrored in studies where multiple measures of skin appearance and function are available in the same participants (Oyetakin-White et al., 2015).

We reason that subjective measures of perceived age may act as a proxy for underlying dermal integrity and photoprotective mechanisms. This study aims to characterize genetic associations with perceived age focusing on understanding heritability, identify genetic loci and understand whether genetic mechanisms regulating dermal integrity are shared with other risk factors or diseases.

RESULTS

Participants

A GWAS was performed for perceived age. After final exclusions, analysis included 423,992 adult participants. Of these, 8,630 reported looking older than their biological age, 103,300 reported looking about their age and 312,062 reported looking younger than their biological age. There were trends with both age and sex, where females were more likely to report looking young for their age than males, and older participants were more likely to report looking young for their age than young participants (S1 Table). As there was an imbalance of

responses in these three groups the effective statistical power of the experiment was smaller than the total sample size (see S1 Text for an estimate of effective sample size).

Total heritable contribution

After final quality control, approximately 9.6 million single nucleotide variants (SNVs) with MAF of 0.1% or greater were tested for association. There was evidence for inflation in test statistics (genomic control factor [λ_{GC}]=1.49), which is typical for large studies of complex, polygenic traits and lower than that reported in recent studies of height and body mass index (BMI) (Yengo et al., 2018). LDSR analysis estimated that 14 % (SE 0.6%) of variation in perceived age was due to effects of common genetic variants, and that polygenic heritability rather than inflationary bias was responsible for most of the inflation in λ_{GC} (LDSR ratio=0.09).

Single variant findings

There was evidence for association at 5,395 SNVs representing 81 independent signals of association ($p < 5 \times 10^{-8}$). 74 represented to our knowledge previously unreported discoveries, while 7 were in the region of loci previously reported for a skin-appearance or facial-age-related trait (Figure 1). A subset of lead variants are presented in Table 1, and full results are provided in S2 Table.

Of these genetic loci, the strongest statistical evidence was seen at *C9orf66-DOCK8* with a lead signal carried by rs520015_C (EAF 0.51, OR 1.07, $P = 1.0 \times 10^{-58}$). SNPs were mapped to genes using positional mapping tools in FUMA (Watanabe et al., 2017). Rs520015 was annotated as an intergenic variant in LD with a missense variant within *C9orf66* (rs481905, $r^2 = 0.81$) and multiple intronic variants within *DOCK8* (e.g. rs2484966, $r^2 = 0.88$).

Of the genetic loci discovered, the largest effect size of a minor allele was estimated for rs139356332_G (EAF 0.98, OR 0.90, $P = 8.1 \times 10^{-13}$), an uncommon intronic variant within

MFAP4. This gene encodes an extracellular glycoprotein which is thought to contribute to organization of elastin fiber components in the extracellular matrix (Pilecki et al., 2016).

The study replicated evidence for association at 7 previously-studied loci, with the strongest evidence at rs12203592_C, a common intronic variant within *IRF4* with large effects on odds of appearing youthful (EAF 0.78, OR 1.22, $P=1.2 \times 10^{-327}$). This locus has been reported for skin-age related traits, including pigmented spot severity, perceived skin ageing, tanning ability, risk of sunburn and tanning response to sun (Jacobs et al., 2015, Law Matthew H. et al., 2017, Visconti et al., 2018, Zhang Mingfeng et al., 2013).

In-silico transcriptome-wide association analysis and fine mapping

Predicted expression of 25,812 gene transcripts was tested for association with perceived facial age. 175 gene transcripts passed a Bonferroni-corrected p value threshold ($p < 1.9 \times 10^{-6}$). There was inflation in these results (S1 Figure), which might represent polygenic association signal but could also be related to correlation in predicted expression of adjacent genes seen using the S-PrediXcan method (Wainberg et al., 2019). There was high concordance with the results of single variant analysis, and nearly all transcripts were in genomic loci already highlighted by the single variant results. The strongest evidence for association was *IRF4*, where higher transcription was predicted to associate with lower odds of looking youthful ($p=1.6 \times 10^{-77}$) (S3 Table). The FOCUS method (Mancuso et al., 2019) for probabilistic fine mapping was applied to help resolve correlation in adjacent predictions and nominate a credible set of biologically causal genes underlying this association signal. This was able to resolve some association signals with a high degree of confidence, for example an association signal on chromosome 2 represented by rs1438898 in single variant results which mapped to a uncharacterized transcript *AC074093.1* using positional mapping was mapped to *ZEB2* with a high probability using fine-mapping. Conversely, the fine mapping approach was unable to

produce stable models for associated loci on chromosome 6, and results from this chromosome are not presented. The single gene with the highest posterior probability estimate for each locus is reported in S4 Table and full results are included in S5 Table.

Enrichment in gene sets and tissue expression.

Enrichment analysis was performed using DEPICT (Pers et al., 2015) implemented in GCT-VL (Cuéllar-Partida et al., 2019). Analysis identified enrichment ($FDR \leq 0.05$) in 23 predefined sets of functionally related genes, with the strongest enrichment signal seen for the *NEK6* subnetwork ($P=3.2 \times 10^{-7}$). (S6 Table). Enrichment in tissue-specific transcription in 209 tissues was tested but didn't identify any enrichment beyond chance ($P_{FDR} > 0.05$ for all tissues) (S7 Table).

Shared heritability and genetic causality proportions

Genetic correlation with 1362 traits was estimated, of which 75 traits had evidence for shared heritability with perceived age after a Benjamini-Hochberg correction for multiple testing ($P_{FDR} < 0.05$). All genetic correlation estimates lay within the range -0.5 to +0.5, suggesting that the genetic determinants of perceived age aren't fully captured by other traits in the GCT-VL (Cuéllar-Partida et al., 2019) catalogue.

The strongest evidence for genetic correlation was seen with obesity-related traits, where genotypes which associated with greater adiposity overlapped with genotypes associated with reduced odds of appearing youthful (for example BMI), $R_g = -0.25$, $P_{FDR} = 2.9 \times 10^{-12}$ and waist circumference, $R_g = -0.23$, $P_{FDR} = 2.3 \times 10^{-13}$) (S8 Table).

These correlations might be due to shared genetic influences on perceived age but may be due to vertically pleiotropic pathway effects where one trait exerts a causal effect on the other. To investigate this further, genetic causality proportions (O'Connor and Price, 2018) were estimated for all pairs of traits with a significant genetic correlation using an online platform

(Haworth et al., 2019a). The single trait with the strongest evidence for causal effect was BMI. Effects of greater BMI on reduced odds of appearing youthful were modelled to explain part of the genetic correlation with BMI (genetic causality proportion estimate=-0.64), however neither this finding ($P_{FDR}=0.07$) nor any other causality proportions were considered significant after correction for multiple testing (S9 Table).

DISCUSSION

This study used genome wide analysis to investigate genetic contributions to perceived age using the rationale that failure of dermal repair mechanisms would lead to visibly aged skin. There was evidence for a polygenic heritable contribution to youthful appearance and single variant analysis identified 74 to our knowledge previously unreported loci. These loci were enriched for gene sets encoding a range of regulatory networks, supporting the idea that a range of different biological processes are implicated in maintaining a youthful appearance. The *NEK6* subnetwork, identified in gene set enrichment analysis, helps govern the initiation of mitosis, progression through the cell cycle and prevents cell senescence (Jee et al., 2010). We hypothesize that natural variation in genes encoding the *NEK6* subnetwork leads to variation in ability to compensate for age-related decline in tissues, resulting in variation in signs of ageing. This is supported by enrichment in other regulatory networks with roles in growth signaling including the *SMAD2*, *SMAD4* and *SMAD9* subnetworks. This mechanism is likely one of several diverse pathways which contribute to maintaining a youthful appearance, for example we also observed enrichment in lipid-mediated signaling.

While perceived age may capture variation in skin biology and response to environmental challenges, it is likely to also capture other biological factors for example nutritional status. To help investigate the degree of overlap with other traits and biological specificity of the

phenotype, we estimated genetic correlations with traits and diseases in a hypothesis-free manner and identified genetic correlations with 75 other traits or diseases. These correlations were modest in magnitude, suggesting that perceived age provides a proxy for an underlying phenotype which hasn't been extensively explored yet by other GWAS studies. We followed up these genetic correlations to explore whether there are causal relationships (in either direction) between perceived age and genetically correlated traits but didn't find strong evidence supporting this. This may reflect the limited statistical power of this follow-up analysis (the genetic correlations were of modest magnitude), or that the genetic overlap between perceived age is predominantly due to biological processes which have underlying relevance for many traits rather than causal pathways between these phenotypes.

Aside from capturing the response to environmental stressors, the analysis may also capture genetic associations with stressors. Factors such as smoking and UV exposure related to geographical location were traditionally considered to be purely environmental and therefore uncorrelated with genotype. However, an increasing body of evidence now points to the heritability of the home environment (Kong et al., 2018) and evidence for correlation between genetic data and both socio-economic conditions (Tyrrell et al., 2017) and latitude (Haworth et al., 2019b) in UK Biobank. Despite reasons for caution, there was little evidence for inflationary bias in the primary results. The lead single variants show good concordance with previously-published findings and appear relevant to dermal protection functions, suggesting that the results of the study primarily capture host susceptibility and response to pro-ageing stimuli, rather than host liability to be exposed to those stimuli.

The existence of genetic predictors of skin function and likely ageing trajectory provides opportunities for research and clinical applications. The results may help prioritize relevant biology for detailed molecular study of photoprotective mechanisms and nominate proteins

whose function, if modulated by cosmetic or pharmacological products, might enhance photoprotection. Statistical power for investigations into longitudinal mechanics of skin ageing might be boosted by recruiting based on participants' genotype who are at greatest risk of accelerated ageing. In the longer term, integration of insights from population-level and individual genetic information may pave the way to precision skincare.

Aside from the conceptual limitations of complexity using perceived age as a proxy for skin traits, one practical limitation of this investigation is the use of categorical data, which is cruder than previous approaches such as using a panel of volunteers to guess the age of a participant and compare that to actual age to generate a continuous measure (Liu et al., 2016). As the phenotype used here was self-reported and subjective there will be some degree of misclassification. We modelled the likely impact of this on statistical power and false discovery rate using simulations. These showed misclassification in this study likely affected the statistical power and led to some degree of under-estimation of effect size at truly associated SNVs but wouldn't lead to false positive associations (S2 Text, S2 Figure). Despite these limitations, the study identified association at previously-reported positive controls such as variants within *IRF4*, *MC1R* and *BNC2* with high levels of statistical evidence. This indicates that the large sample size was sufficient to overcome regression dilution bias introduced by misclassification. Although we believe that the properties of the phenotypic assessment in this study would lead to under-reporting of association signals, we haven't undertaken replication in an independent sample which is a limitation of the study.

In conclusion, apparent age is a partially heritable trait. The polygenic association signal and results of gene set analysis suggest that diverse mechanisms act to preserve a youthful appearance. Biological and functional characterization of the single variant association signals identified in this study may be a useful way to gain improved understanding of skin biology or as a step towards interventions which moderate the rate of age-related skin changes.

MATERIALS AND METHODS

Participants and phenotypes

This study used data from UK Biobank, a project which recruited approximately 500,000 participants aged between 40-69 between 2006 and 2010 (Fry et al., 2017). Eligible participants were identified from health records in the UK National Health Service (NHS), and invited to participate in one of 22 assessment centers, which were in densely-populated regions of Great Britain. Participants took part in a baseline assessment including completion of questionnaires, physical measurement, donation of biological samples and consent for subsequent follow-up via linkage to NHS records. In the questionnaire, participants were asked to respond to the question ‘Do people say that you look’. The possible answers were ‘Younger than you are’, ‘Older than you are’, ‘About your age’, ‘Do not know’ or ‘Prefer not to answer’. For this analysis, participants were coded ‘1’ if they reported they reported looking younger, ‘0’ if they reported they looked older and ‘0.5’ if they reported they looked their age. Participants who didn’t know or preferred not to answer the question were excluded from analysis.

Genotypes

Genotype data was generated using one of two genotyping arrays - the UK BiLEVE Axiom array and the UK Biobank Axiom array. Quality control and imputation were undertaken centrally by UK Biobank as described previously (Bycroft et al., 2017). Following imputation, in-house quality control was undertaken to remove participants with poor quality data and to restrict analysis to participants of European ancestry, following a published protocol (Mitchell et al., 2017). Genotype data was filtered to a high-confidence set of SNVs by removing monomorphic or rare variants with MAF <0.1%, removing structural variation

such as insertion-deletions, removing sites with poor imputation quality using a graded filter (minimum INFO score >0.3 for MAF $>3\%$, INFO >0.6 for MAF in the range 1-3%, INFO >0.8 for MAF in the range 0.5-1% and INFO >0.9 for MAF in the range 0.1-0.5%) and removing sites not in the Haplotype Reference Consortium imputation panel (Mitchell et al., 2017).

Genome-wide association analysis

Genome-wide analysis was performed using a linear mixed model approach implemented in BOLT-LMM (Loh et al., 2018). This tests the relationship between genotype and phenotype while accounting for covariates (age, sex and study participation center) and relatedness, following a published protocol (Elsworth et al., 2017). This approach was chosen because the linear mixed model approach is reported to achieve good control for potential confounding due to population stratification in the UK Biobank sample (Loh et al., 2018). Genome-wide summary statistics on a linear scale were transformed into log odds ratios using a Taylor expansion series. Odds ratios greater than 1 indicate greater odds of looking youthful.

SNV selection procedure and conditional analysis

SNVs with $p < 5 \times 10^{-8}$ were considered associated with perceived age, chosen as a threshold for genome-wide significance. Nearby SNVs are typically correlated through linkage disequilibrium (LD), meaning that genetic effect sizes and p-values of nearby SNVs are also correlated. Lead SNVs were defined after reducing the association signals down to a subset of approximately-independent signals of association within single-variant results using a stepwise model selection procedure implemented in GCTA(v1.91.4) (Yang et al., 2012, Yang et al., 2011), which takes into account LD between different SNVs to select independently associated SNVs (--cojo-slc function). This subset of approximately independent signals was tested against previously reported association signals in an approximate conditional analysis

(--cojo-cond function in GCTA) in order to identify which signals capture previously-reported associations, using a list of SNVs reported for perceived facial age or related traits (S9 Table). SNVs which were conditionally independent of previously-reported signals of association with $P < 5 \times 10^{-8}$ in conditional analysis were defined as lead signals, and are reported in S9 Table.

Enrichment analysis

To test for enrichment in predefined gene sets or gene pathways, enrichment analysis was performed using the DEPICT approach (Pers et al., 2015) implemented in GCT-VL (GCT-VL). Analysis used full genome-wide results, and associated loci were defined internally by DEPICT using a reference panel for LD estimation.

Estimation of heritability

To estimate variation in perceived facial age attributable to common genetic variants, heritability was estimated using univariate linkage-disequilibrium score regression (LDSR) (Bulik-Sullivan Brendan K. et al., 2015), implemented in LD-Hub, an automated online resource (Zheng et al., 2017). Summary statistics of GWAS were uploaded and results processed through a standardized procedure. This uses a subset of approximately 1 million common variants and reference LD data to estimate heritability attributable to common genetic variants (h^2_{LDSR}) and assess for inflationary bias in GWAS results.

Estimation of genetic correlation and partial genetic causality

Genetic correlation (R_g) was assessed against 1362 traits in the Complex Trait Genomics Virtual Lab (GCT-VL) catalogue (Cuéllar-Partida et al., 2019) using bivariate LDSR (Bulik-Sullivan Brendan et al., 2015). Genetic correlation summarizes the similarity in the heritable contribution to a pair of traits assessed across the whole genome; values near 1 or -1 indicate two traits have substantial shared genetic associations, which have consistent and

proportionate effects on both traits. Values near 0 indicate largely independent genetic determinants with little overlapping heritability between the two traits. Adjustment for multiple testing used a Benjamini-Hochberg procedure and correlations with false discovery rate (FDR) <0.05 were reported.

For each trait with a detectable non-zero genetic correlation, latent causal variable models (O'Connor and Price, 2018) were fitted to help distinguish between genetic correlations resulting from horizontally pleiotropic genetic effects and genetic correlations resulting from causal relationships. Models were fitted using an automated online pipeline (Haworth et al., 2019) implemented in the GCT-VL platform (<https://genoma.io>).

Imputed transcriptome wide association study and transcriptome-informed fine mapping

To test the consequences of a range of gene transcripts on perceived facial age, tests for association with predicted gene expression were performed using S-PrediXcan (Barbeira et al., 2017). This assesses the mediating effects of expression levels on phenotypes by imputing transcriptome levels, using pre-trained models derived in datasets with measured gene expression. Analysis was performed using pre-fitted elastic net prediction models of gene expression levels in the 48 Genotype-Tissue Expression tissues (Lonsdale et al., 2013), which are available online (URLs: <http://predictdb.org/>). Summary results from GWAS were uploaded to the S-PrediXcan (Barbeira et al., 2017) web pipeline (URL: <https://cloud.hakyimlab.org/>). Results from 48 tissues were combined using the TissueXcan method (Barbeira et al., 2019) which prioritizes the most relevant tissue transcripts overall, taking into account evidence from multiple tissue-specific predictions while accounting for correlation in gene transcription between different tissues and multiple testing. In parallel, analysis using the FOCUS method (Mancuso et al., 2019) was performed using the standalone

python software provided by the authors of the method at (<https://github.com/bogdanlab/focus/blob/master/README.md>). It references LD data from the 1000 genomes project samples (European ancestry) and reference transcription data from the pre-compiled database including data from multiple sources described at (<https://github.com/bogdanlab/focus/wiki>).

DATA AVAILABILITY

Full results of analysis are provided as supplementary datasets. Underlying source data are available through UK Biobank, an open-access resource for healthcare research. Data access procedures are described at <http://www.ukbiobank.ac.uk/wp-content/uploads/2012/09/Access-Procedures-2011-1.pdf>. Genome-wide summary statistics have been uploaded to the University of Bristol data repository, data.bris, and are publicly available at <https://doi.org/10.5523/bris.21crwsnj4xwjm2g4qi8chathha>.

CONFLICT OF INTEREST

The authors state no conflict of interest.

ACKNOWLEDGEMENTS

This research has been conducted using the UK Biobank Resource under Application Number 40644. B.M is funded by a National Institute for Health Research (NIHR) Clinical Lectureship. N.J.T. is a Wellcome Trust Investigator (202802/Z/16/Z), is the PI of the Avon Longitudinal Study of Parents and Children (MRC & WT 102215/2/13/2), is supported by the University of Bristol NIHR Biomedical Research Centre (BRC-1215-20011), the MRC

Integrative Epidemiology Unit (MC_UU_12013/3) and works within the CRUK Integrative Cancer Epidemiology Programme (C18281/A19169). S.H is funded by a National Institute for Health Research (NIHR) Academic Clinical Fellowship.

REFERENCES

- Barbeira AN, Dickinson SP, Torres JM, Bonazzola R, Zheng J, Torstenson ES, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *bioRxiv* 2017.
- Barbeira AN, Pividori M, Zheng J, Wheeler HE, Nicolae DL, Im HK. Integrating predicted transcriptome from multiple tissues improves association detection. *PLOS Genetics* 2019;15(1):e1007889.
- Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic correlations across human diseases and traits. *Nature Genetics* 2015;47(11):1236-+.
- Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* 2015;47(3):291-+.
- Bycroft C, Freeman C, Petkova D, Band VOPG, Elliott LT, Sharp K, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv*2017.
- Cuéllar-Partida G, Lundberg M, Kho PF, D'Urso S, Gutiérrez-Mondragón LF, Ngo TT, et al. Complex-Traits Genetics Virtual Lab: A community-driven web platform for post-GWAS analyses. *bioRxiv* 2019:518027.
- Duffy DL, Box NF, Chen W, Palmer JS, Montgomery GW, James MR, et al. Interactive effects of MC1R and OCA2 on melanoma risk phenotypes. *Hum Mol Genet* 2004;13(4):447-61.
- Duffy DL, Iles MM, Glass D, Zhu G, Barrett JH, Höiom V, et al. IRF4 variants have age-specific effects on nevus count and predispose to melanoma. *Am J Hum Genet* 2010;87(1):6-16.

- Elsworth BL, Mitchell R, Raistrick CA, Paternoster L, Hemani G, Gaunt TR. MRC IEU UK Biobank GWAS pipeline version 1. 2017.
- Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *American Journal of Epidemiology* 2017;186(9):1026-34.
- Han J, Kraft P, Nan H, Guo Q, Chen C, Qureshi A, et al. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet* 2008;4(5):e1000074.
- Han J, Qureshi AA, Nan H, Zhang J, Song Y, Guo Q, et al. A germline variant in the interferon regulatory factor 4 gene as a novel skin cancer risk locus. *Cancer Res* 2011;71(5):1533-9.
- Haworth S, Kho PF, Holgerson PL, Hwang L-D, Timpson NJ, Rentería ME, et al. Inference and visualization of phenome-wide causal relationships using genetic data: an application to dental caries and periodontitis. *bioRxiv* 2019a:865956.
- Haworth S, Mitchell R, Corbin L, Wade KH, Dudding T, Budu-Aggrey A, et al. Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nature Communications* 2019b;10(1):333.
- Jacobs LC, Hamer MA, Gunn DA, Deelen J, Lall JS, van Heemst D, et al. A Genome-Wide Association Study Identifies the Skin Color Genes IRF4, MC1R, ASIP, and BNC2 Influencing Facial Pigmented Spots. *Journal of Investigative Dermatology* 2015;135(7):1735-42.
- Jee HJ, Kim AJ, Song N, Kim H-J, Kim M, Koh H, et al. Nek6 overexpression antagonizes p53-induced senescence in human cancer cells. *Cell Cycle* 2010;9(23):4703-10.

- Kita R, Fraser HB. Local Adaptation of Sun-Exposure-Dependent Gene Expression Regulation in Human Skin. *PLoS Genet* 2016;12(10):e1006382.
- Kong A, Thorleifsson G, Frigge ML, Vilhjalmsón BJ, Young AI, Thorgeirsson TE, et al. The nature of nurture: Effects of parental genotypes. *Science* 2018;359(6374):424-8.
- Law MH, Medland SE, Zhu G, Yazar S, Viñuela A, Wallace L, et al. Genome-Wide Association Shows that Pigmentation Genes Play a Role in Skin Aging. *Journal of Investigative Dermatology* 2017;137(9):1887-94.
- Lee SH, Jeong SK, Ahn SK. An update of the defensive barrier function of skin. *Yonsei Med J* 2006;47(3):293-306.
- Liu F, Hamer MA, Deelen J, Lall JS, Jacobs L, van Heemst D, et al. The MC1R Gene and Youthful Looks. *Curr Biol* 2016;26(9):1213-20.
- Loh P-R, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. *Nature Genetics* 2018;50(7):906-8.
- Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. *Nature Genetics* 2013;45:580.
- López S, García O, Yurrebaso I, Flores C, Acosta-Herrera M, Chen H, et al. The interplay between natural selection and susceptibility to melanoma on allele 374F of SLC45A2 gene in a South European population. *PLoS One* 2014;9(8):e104367.
- Mancuso N, Freund MK, Johnson R, Shi H, Kichaev G, Gusev A, et al. Probabilistic fine-mapping of transcriptome-wide association studies. *Nature Genetics* 2019;51(4):675-82.
- Mitchell R, Hemani G, Dudding T, Paternoster L. UK Biobank Genetic Data: MRC-IEU Quality Control, Version 1. 2017.

- Nan H, Kraft P, Qureshi AA, Guo Q, Chen C, Hankinson SE, et al. Genome-Wide Association Study of Tanning Phenotype in a Population of European Ancestry. *Journal of Investigative Dermatology* 2009;129(9):2250-7.
- O'Connor LJ, Price AL. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nature Genetics* 2018;50(12):1728-34.
- Oyetaikin-White P, Suggs A, Koo B, Matsui MS, Yarosh D, Cooper KD, et al. Does poor sleep quality affect skin ageing? *2015;40(1):17-22.*
- Pers TH, Karjalainen JM, Chan Y, Westra HJ, Wood AR, Yang J, et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nature Communications* 2015;6:9.
- Pilecki B, Holm AT, Schlosser A, Moeller JB, Wohl AP, Zuk AV, et al. Characterization of Microfibrillar-associated Protein 4 (MFAP4) as a Tropoelastin- and Fibrillin-binding Protein Involved in Elastic Fiber Formation. *The Journal of Biological Chemistry* 2016;291(3):1103-14.
- 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;39(12):1443-52.
- Tyrrell J, Wood AR, Ames RM, Yaghootkar H, Beaumont RN, Jones SE, et al. Gene-obesogenic environment interactions in the UK Biobank study. *International journal of epidemiology* 2017;46(2):559-75.
- Visconti A, Duffy DL, Liu F, Zhu G, Wu W, Chen Y, et al. Genome-wide association study in 176,678 Europeans reveals genetic loci for tanning response to sun exposure. *Nature Communications* 2018;9:1684.

- Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, et al. Opportunities and challenges for transcriptome-wide association studies. *Nature Genetics* 2019;51(4):592-9.
- Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* 2017;8(1):1826.
- Williams HC. Atopic Dermatitis. *New England Journal of Medicine* 2005;352(22):2314-24.
- Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ATC, Consortium DIGRAM-a, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics* 2012;44:369.
- Yang JA, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis. *American Journal of Human Genetics* 2011;88(1):76-82.
- Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Human Molecular Genetics* 2018;27(20):3641-9.
- Zhang M, Song F, Liang L, Nan H, Zhang J, Liu H, et al. Genome-wide association studies identify several new loci associated with pigmentation traits and skin cancer risk in European Americans. *Human Molecular Genetics* 2013;22(14):2948-59.
- Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* 2017;33(2):272-9.

TABLES

Table 1: Top 10 independently-associated lead variants in GWAS

Chr	SNP	Position	Effect allele	EAF	Beta	se	Odds ratio	Locus	p (conditional)
								<i>C9orf66</i>	
9	rs520015	211762	C	0.51	0.070	0.0043	1.07	- <i>DOCK8</i>	1.2x10 ⁻⁵⁸
8	rs10956486	130699140	T	0.68	-0.073	0.0046	0.93	<i>GSDMC</i>	2.6x10 ⁻⁵⁵
3	rs61263161	126691104	G	0.83	-0.087	0.0058	0.92	<i>CHCHD6</i>	1.4x10 ⁻⁵¹
2	rs1438898	145714354	A	0.75	0.059	0.0050	1.06	<i>AC074093.1</i>	3.5x10 ⁻³²
2	rs76032374	56058356	A	0.87	0.068	0.0064	1.07	<i>EFEMP1</i>	7.9x10 ⁻²⁷
20	rs28897169	22100542	T	0.39	0.043	0.0044	1.04	<i>LOC100270679</i>	7.0x10 ⁻²²
6	rs4869723	151579432	C	0.56	-0.040	0.0044	0.96	<i>AKAP12</i>	1.7x10 ⁻²⁰
2	rs116254882	223025055	G	0.96	0.095	0.0107	1.10	<i>PAX3</i>	5.2x10 ⁻¹⁹
15	rs1550436	74221157	C	0.53	-0.039	0.0043	0.96	<i>LOXLI</i>	5.5x10 ⁻¹⁹
2	rs7590866	223087329	G	0.86	-0.054	0.0062	0.95	<i>PAX3</i>	2.3x10 ⁻¹⁸

Each row contains a lead SNV representing an independent signal of association following a stepwise selection procedure with $p < 5.0e^{-08}$ after conditioning on previously-reported signals of association. The position column contains genomic position based on build 37 (GRCh37.p13) of the human genome. The beta coefficient and accompanying standard error are on a log-odds scale and have been exponentiated to provide an odds ratio for reference. Odds ratios greater than 1 reflect increased odds of appearing young. The locus column includes the name of the gene nearest the lead SNV. Results for all independent lead variants are provided in S2 table with both conditional and unconditional P values. Full results for all variants are provided as a link in the data access statement.

FIGURE LEGENDS

Figure 1: Manhattan plot of GWAS.

Genomic regions independently meeting genome-wide significance are highlighted in magenta (to our knowledge previously unreported loci) or navy blue (positive controls). The red line at $p=5 \times 10^{-8}$ indicates the conventional threshold for genome-wide significance. The y-axis scale is split and truncated at $p=1 \times 10^{-250}$.

SUPPLEMENTARY MATERIAL

2 Supplementary Text

9 Supplementary Tables

2 Supplementary Figures

CRedit:

Conceptualizations: VR, BM, NT, SH

Data Curation:

Formal Analysis: VR, SH

Funding Acquisition:

Investigation:

Methodology:

Project Administration:

Resources:

Software:

Supervision: BM, NT, SH

Visualization: VR, SH

Writing – Original Draft Preparation: VR

Writing – Review and Editing: VR, BM, NT, SH