bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Novel genetic loci affecting facial shape variation in humans

Ziyi Xiong[1,2 †], Gabriela Dankova[2, †], Laurence J. Howe[3], Myoung Keun Lee[4], Pirro G. Hysi[5], Markus A. de Jong[2,6,7,$], Gu Zhu[8], Kaustubh Adhikari[9], Dan Li[10], Yi Li[1], Bo Pan[11], Eleanor Feingold[4], Mary L. Marazita[4,12], John R. Shaffer[4,12], Kerrie McAloney[8], Shuhua Xu[10,13,14], Li Jin[10,15], Sijia Wang[10,14], Femke M. de Vrij[16], Bas Lendemeijer[16], Stephen Richmond[17], Alexei Zhurov[17], Sarah Lewis[3], Gemma Sharp[3,18], Lavinia Paternoster[3], Holly Thompson[3], Rolando Gonzalez-Jose[19], Maria Catira Bortolini[20], Samuel Canizales-Quinteros[21], Carla Gallo[22], Giovanni Poletti[22], Gabriel Bedoya[23], Francisco Rothhammer[24], André G. Uitterlinden[25,26], M. Arfan Ikram[26], Eppo B. Wolvius[6], Steven A. Kushner[16], Tamar Nijsten[27], Robert-Jan Palstra[28], Stefan Boehringer[7], Sarah E Medland[8], Kun Tang[10], Andrés Ruiz-Linares[15,29], Nicholas G. Martin[8], Timothy D. Spector[5, ‡], Evie Stergiakouli[3,18, ‡], Seth M. Weinberg[4,12,30, ‡], Fan Liu[1,2, ‡,*], Manfred Kayser[2, ‡,*] on behalf of the International Visible Trait Genetics (VisiGen) Consortium

1. CAS Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing, China.

2. Department of Genetic Identification, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands.

3. Medical Research Council Integrative Epidemiology Unit, Population Health Sciences, University of Bristol, Bristol, United Kingdom.

4. Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, U.S.A.

5. Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom.

6. Department of Oral & Maxillofacial Surgery, Special Dental Care, and Orthodontics, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands.

7. Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, the Netherlands.

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

8. QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia.

9. Department of Genetics, Evolution, and Environment, University College London, London, United Kingdom.

10. Chinese Academy of Sciences (CAS) Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology (PICB), Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, CAS, Shanghai, China.

11. Department of Auricular Reconstruction, Plastic Surgery Hospital, Beijing, China.

12. Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA, U.S.A.

13. School of Life Science and Technology, ShanghaiTech University, Shanghai, China.

14. Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China.

15. State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China.

16. Department of Psychiatry, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands.

17. Applied Clinical Research and Public Health, University Dental School, Cardiff University, Heath Park, Cardiff, United Kingdom.

18. School of Oral and Dental Sciences, University of Bristol, Bristol, United Kingdom.

19. Instituto Patagonico de Ciencias Sociales y Humanas, CENPAT-CONICET, Puerto Madryn, Argentina.

20. Departamento de Genetica, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil.

21. Unidad de Genomica de Poblaciones Aplicada a la Salud, Facultad de Quımica, UNAM-Instituto Nacional de Medicina Genomica, Mexico City, Mexico.

22. Laboratorios de Investigacion y Desarrollo, Facultad de Ciencias y Filosofıa, Universidad Peruana Cayetano Heredia, Lima, Peru.

23. GENMOL (Genetica Molecular), Universidad de Antioquia, Medellın, Colombia.

24. Instituto de Alta Investigacion, Universidad de Tarapaca, Arica, Chile.

25. Department of Internal Medicine, Erasmus MC University Medical Center Rotterdam, Rotterdam,

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the Netherlands.

26. Department of Epidemiology, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands.

27. Department of Dermatology, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands.

28. Department of Biochemistry, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands.

29. Laboratory of Biocultural Anthropology, Law, Ethics, and Health (Centre National de la Recherche Scientifique and Etablissement Français du Sang), Aix-Marseille University, Marseille, France.

30. Department of Anthropology, University of Pittsburgh, Pittsburgh, PA, U.S.A.

† These authors contributed equally to this work.

‡ These authors jointly supervised this work.

$ current affiliation: Department of Computer Science, VU University Amsterdam, Amsterdam, the Netherlands

*To whom correspondence should be addressed at:

CAS Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beichen West Road 1-104, Chaoyang, Beijing, 100101, P.R. China. Tel: +86-010-84097876; Fax: +86-010-84097720; Email: liufan@big.ac.cn (F.L.)

Or

Department of Genetic Identification, Erasmus MC University Medical Center Rotterdam, Wytemaweg 80, 3015 CN, Rotterdam, the Netherlands. Tel: +31 10 7038073; Email: m.kayser@erasmusmc.nl (M.K.).

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

The human face represents a combined set of highly heritable phenotypes, but knowledge on its genetic architecture remains limited despite the relevance for various fields of science and application. A series of genome-wide association studies on 78 facial shape phenotypes quantified from 3-dimensional facial images of 10,115 Europeans identified 24 genetic loci reaching genome-wide significant association, among which 17 were previously unreported. A multi-ethnic study in additional 7,917 individuals confirmed 13 loci including 8 unreported ones. A global map of polygenic face scores assembled facial features in major continental groups consistent with anthropological knowledge. Analyses of epigenomic datasets from cranial neural crest cells revealed abundant *cis*-regulatory activities at the face-associated genetic loci. Luciferase reporter assays in neural crest progenitor cells highlighted enhancer activities of several face-associated DNA variants. These results substantially advance our understanding of the genetic basis underlying human facial variation and provide candidates for future *in-vivo* functional studies.

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

The human face represents a multi-dimensional set of correlated, mostly symmetric, complex phenotypes with a highly heritable components, as illustrated by a large degree of facial similarity between monozygotic twins and, albeit less so, between other relatives (*Djordjevic, Zhurov, Richmond, & Visigen, 2016*), stable facial features within major human populations, and the enormous diversity among unrelated persons almost at the level of human individualization. Understanding the genetic basis of human facial variation has important implications for a variety of disciplines including developmental biology, evolutionary biology, medicine, and forensics. However, current knowledge on the specific genes influencing facial phenotypes in humans and the underlying molecular mechanisms forming facial morphology is limited. Over recent years, 9 separate genome-wide association studies (GWASs) have each highlighted several but largely non-overlapping genetic loci associated with facial shape phenotypes (*Adhikari et al., 2016; Cha et al., 2018; Claes et al., 2018; Cole et al., 2016; M. K. Lee et al., 2017; Liu et al., 2012; Paternoster et al., 2012; Pickrell et al., 2016; Shaffer et al., 2016*). The overlapping loci from independent facial GWAS, thus currently representing the most established findings, include DNA variants in or close to *CACNA2D3* (*Paternoster et al., 2012; Pickrell et al., 2016*), *DCHS2* (*Adhikari et al., 2016; Claes et al., 2018*), *EPHB3* (*Claes et al., 2018; Pickrell et al., 2016*), *HOXD* cluster (*Claes et al., 2018; Pickrell et al., 2016*), *PAX1* (*Adhikari et al., 2016; Shaffer et al., 2016*), *PAX3* (*Adhikari et al., 2016; Claes et al., 2018; Liu et al., 2012; Paternoster et al., 2012; Pickrell et al., 2016*), *PKDCC* (*Claes et al., 2018; Pickrell et al., 2016*), *SOX9* (*Cha et al., 2018; Claes et al., 2018; Pickrell et al., 2016*), *SUPT3H* (*Adhikari et al., 2016; Claes et al., 2018; Pickrell et al., 2016*) and *TBX15* (*Claes et al., 2018; Pickrell et al., 2016*), which were identified mostly in Europeans.

Indicated by the small effects of the previously identified genes together with the high heritability of facial appearance characteristics, the true number of genes that influence polygenetic facial traits is likely to be much larger. Despite the previous work, our understanding of the complex genetic architecture of human facial morphology remains largely incomplete, emphasizing the need for well-powered and multi-ethnic studies that can be achieved through large collaborative efforts. A most recent study (*Claes et al., 2018*) suggested that variants at genetic loci implicated in human facial shape are involved in the epigenetic regulation of human neural crest cells, suggesting that at

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

least some of the functional variants may reside within regulatory elements such as enhancers. However, understanding the functional basis of DNA variants for which facial phenotype associations have been established is largely missing thus far.

Led by the International Visible Trait Genetics (VisiGen) Consortium, the current study represents a collaborative effort to identify novel genetic variants involved in human facial variation via GWASs and replication studies in the largest set of multi-ethnic samples available thus far. Moreover, this study demonstrates the functional consequences of face-associated DNA variants using in-silico and *in-vitro* cell line studies.

Result

Facial phenotypes

The current study included seven cohorts, totaling 18,032 individuals. Demographic characteristics and phenotyping details for all cohorts are provided in **Methods** and **Supplementary Table 1**. A variety of different phenotyping methods are available for 3D facial image data, where some approaches (*Claes et al., 2018*) require access to the underlying raw image datasets, which often cannot be shared between different research groups. In order to maximize sample size via a collaborative study, we focused on linear distance measures of facial features that can be accurately derived from facial image datasets. In our cohorts with 3D facial surface images, we focused on distances between a common set of thirteen well-defined facial landmarks (**Fig. 1**). In the Rotterdam Study (RS) and the TwinsUK study (TwinsUK), we used the same ensemble method for automated 3D face landmarking that was specifically designed for large cohort studies with the flexibility of landmark choice and high robustness in dealing with various image qualities (*de Jong et al., 2018; de Jong et al., 2016*). Facial landmarking in Avon Longitudinal Study of Parents and their Children (ALSPAC) and Pittsburgh 3D Facial Norms study (PITT) was done manually. Generalized Procrustes Analysis (GPA) was used to remove affine variations due to shifting, rotation and scaling (**Supplementary Fig. 1**). A total of 78 Euclidean distances involving all combinatorial pairs of the 13 facial landmarks were considered as facial phenotypes (**Supplementary Table 2**).

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Detailed phenotype analyses were conducted in RS, where a highly significant sex effect (min p = 2.1×10^{-151}) and aging effect (min p = 8.0×10^{-44}) were noted, together explained up to 21% of the variance per facial phenotype (e.g. between Alares and Ls; **Supplementary Fig. 2A-C**, **Supplementary Table 3**). Intra-organ facial phenotypes (e.g. within nose or within eyes) showed on average higher correlations than inter-organ ones (e.g. between nose and eyes), and symmetric facial phenotypes showed higher correlations than non-symmetric ones. Genetic correlations (*Zhou & Stephens, 2014*) estimated using genome-wide data were similar to correlations obtained directly from phenotype data, with on average higher absolute values seen based on genotypes (**Supplementary Fig. 3**, **Supplementary Tables 4** and **5**). Unsupervised hierarchical clustering analysis identified four distinct clusters of facial shape phenotypes, e.g. many distances related to pronasale were clustered together (**Supplementary Fig. 2D**). Twin heritability of all facial phenotypes was estimated in TwinsUK as h^2 mean(sd) = 0.52(0.15), max = 0.86 and in Queensland Institute of Medical Research study (QIMR) as h^2 mean(sd) = 0.55(0.19), max = 0.91 (**Supplementary Table 6**). The most heritable features were linked to the central face (i.e., the outer- and inter- orbital distances, the nose breadth, and the distance between the subnasion and the inner eye corners; **Supplementary Fig. 2E** and **2F**). These twin-derived heritability estimates were generally higher than those obtained from genome-wide SNPs in the same cohorts (**Supplementary Table 6**). These results were overall consistent with expectations and suggest that the phenotype correlation within the human face may be explained in part by sets of shared genetic components.

GWASs and replications

We conducted a series of GWASs and meta-analyses to test the genetic association of 7,029,494 autosomal SNPs with 78 face phenotypes in RS, TwinsUK, ALSPAC, and PITT, totaling 10,115 subjects of European descent used as the discovery cohort. Inflation factors were in an acceptable range in all meta-analyses (average λ=1.015, sd=0.006; **Supplementary Fig. 4**) and were not further considered. A power analysis showed that our discovery data had over 90% power to detect an allelic effect explaining 0.43% of variance per phenotype (**Supplementary Fig. 5**). In the discovery GWASs we applied the traditionally used threshold of genome-wide significance at $P<5\times10^{-8}$ to select SNPs for replication. A total of 494 SNPs from 24 distinct genetic loci passed this significance threshold

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(**Supplementary Table 7**), of which 17 were novel and 7 have been reported in previous studies (*Adhikari et al., 2016; Cha et al., 2018; Claes et al., 2018; Liu et al., 2012; Paternoster et al., 2012; Pickrell et al., 2016; Shaffer et al., 2016*). These signals involved multiple facial phenotypes clustered to the central region of the face above the upper lip (**Fig. 1**), and closely resembled the twins-based face heritability map (**Supplementary Fig. 2E** and **2F**). All seven previously reported loci showed largely consistent allele effects on the same or similar sets of facial phenotypes as described in previous GWASs (**Supplementary Table 8**). Three of them have also been noted in the GWAS catalog for face morphology or related phenotypes i.e., nose size, chin dimples, ear morphology, monobrow thickness, and male pattern baldness (**Supplementary Table 9**). Six genomic loci showed a high degree of diversity among the major continental groups in terms of both allele frequency differences (>0.2) and Fst values (empirical $p < 0.05$), including *CASC17*, *INTU*, *PAX3*, *RPE65*, *RPGRIP1L* and *SALL1* regions (**Supplementary Tables 10** and **11**). In addition, four genomic loci showed signals of positive selection in terms of the integrated haplotype scores (iHS, empirical $p<0.05$), including *CASC17*, *RPE65*, *RPGRIP1L* and *TBX15* (**Supplementary Table 11**).

We selected the 24 top-associated SNPs (**Table 1**), one SNP per each of the identified 24 genetic regions, for replicating the findings from the discovery GWAS meta-analysis in the replication dataset comprising three additional cohorts i.e., the Consortium for the Analysis of the Diversity and Evolution of Latin America (CANDELA) study involving Latin Americans known to be of admixed Native American and European ancestry, the Xinjiang Uyghur (UYG) study from China known to be of admixed East Asian and European ancestry, and the Queensland Institute of Medical Research (QIMR) study from Australia involving Europeans, totaling 7,917 multi-ethnic subjects. Exact replication was conducted in UYG (N=858), where 3D facial images were available. In CANDELA (N=5,958) and QIMR (N=1,101), where 2D frontal photographs were available, a 'combined test of dependent tests' (*Kost & McDermott, 2002*) was performed, which tests for H0: no association for all facial phenotypes vs. H1: associated with at least one facial phenotypes. Evidence of replication was obtained for 13 of the 24 genetic loci, among which 8 were novel loci: 1p36.22 *CASZ1/PEX14*, 1p31.2 *RPE65/DEPDC1*, 4q28.1 *INTU*, 6p22.3 *RNF144B*, 6p21.2 *KIF6*, 12q24.21 *TBX3*, 14q32.2 *C14orf64*, and 16q12.2 *RPGRIP1L/FTO*, and 5 have been described in previous GWASs: 2q36.1 *PAX3*, 4q31.3 *SFRP2/DCHS2*, 9q22.31 *ROR2*, 17q24.3 *CASC17*, and 20p11.22 *PAX1* (*Adhikari et al., 2016; Claes et al., 2018;*

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Paternoster et al., 2012; Pickrell et al., 2016; Shaffer et al., 2016) (**Fig. 2**). The allelic effects of the replicated SNPs were highly consistent across all participating cohorts, while the individual cohorts were underpowered (**Supplementary Figs. 6** to **29**), further emphasizing on the merit of our collaborative study. The replicated locus-associated phenotypes demonstrated a clear pattern of facial symmetry i.e., the same association was observed on both sides of the face, while such a symmetric pattern was less clear for most phenotypes involved in the non-replicated loci (**Supplementary Figs. 6** to **29**). Several replicated loci showed effects on the similar sets of facial shape phenotypes; for example, 2q36.1 *PAX3* and 6p21.2 *KIF6* (**Fig. 2E**) both affected the distances between nasion and medial/lateral canthi whereas 1p36.22 *CASZ1/PEX14*, 4q28.1 *INTU* (**Fig. 2A** and **2C**) and 20p11.22 *PAX1* all affected the distances between pronasion and alares. Moreover, several replicated loci affected phenotypes in multiple facial regions e.g. 4q32.1 *C14orf64* and 12q24.1 *TBX3* (**Fig. 2F**) both affected many distances among eyes, alares and mouth. Among the 11 non-replicated loci, 9 were novel and 2 (*SOX9* and *TBX15*) were reported previously in two independent facial GWASs (*Claes et al., 2018; Pickrell et al., 2016*). It should be noted however, that the discovery dataset is of European ancestry while the replication dataset is multi-ethnic; hence, the replication results presented here are rather conservative, i.e., the possibility of false negatives in this replication study cannot be excluded (see Discussion section).

Integration with previous literature

In our discovery GWAS dataset, we also investigated 122 face-associated SNPs reported in previous facial GWASs (*Adhikari et al., 2016; Cha et al., 2018; Claes et al., 2018; Cole et al., 2016; M. K. Lee et al., 2017; Liu et al., 2012; Paternoster et al., 2012; Pickrell et al., 2016; Shaffer et al., 2016*). Out of those, a total of 44 (36%) SNPs showed nominally significant association on either the same set of facial phenotypes or at least one facial phenotypes in a combined test (**Supplementary Table 8**). Notably, none of the 5 SNPs reported in a previous African GWAS (*Cole et al., 2016*) was replicated in our European dataset. This might be explained by population heterogeneity in LD structure, and/or different genetic basis of the tested facial features in Europeans and Africans, and/or different phenotypes, i.e. in the previous study these loci were primarily associated with facial size, which was not tested on our study. We also investigated the potential links between facial phenotypes and 60

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SNPs in 38 loci previously implicated in non-syndromic cleft lip and palate (NSCL/P) phenotype, which is linked with normal variation of facial morphology as suggested previously (*Beaty et al., 2010; Beaty et al., 2011; Birnbaum et al., 2009; Grant et al., 2009; Leslie et al., 2017; Ludwig et al., 2017; Ludwig et al., 2012; Mangold et al., 2010; Sun et al., 2015; Y. Yu et al., 2017*). Among these, 17 SNPs in 13 loci showed significant face-associations in the discovery cohorts after multiple testing correction, and many (11 SNPs in 8 loci) involving cleft related facial landmarks as expected, e.g. Subnasale, Labliale superius, and Labiale inferius (**Supplementary Table 12**).

Multivariable model fitting and polygenic face scores

In the RS cohort, a series of step-wise multivariable model fitting analyses of the 494 face-associated SNPs at 24 loci highlighted 33 SNPs showing independent effects on sex- and age-adjusted face phenotypes (**Supplementary Table 13**). These 33 SNPs individually explain less than 1%, and together explain up to 4.42%, of the variance per facial shape phenotype (**Fig. 3A** and **3B**), with the top-explained phenotypes clustered to the central region of the face, i.e. the distances between Nasion, Subnasale, Pronasale, Left Endocanthion and Right Endocanthion. We then used the fitted models to derive a set of polygenetic face scores for EUR, AFR and EAS subjects (total n = 1,668) from the 1000-Genomes Project. The polygenic face scores largely reversely distributed between EUR and AFR, with EAS situated in between (**Fig. 3C**). EUR showed the longest distances among the vertical phenotypes, while AFR had the widest of the horizontal phenotypes. EAS was consistently in-between with the most differential phenotype scores observed for nose wing breadth (AFR > EAS > EUR) and the nose length (EUR > EAS > AFR). An exception was for mouth phenotypes, where AFR had the largest values in both mouth width and lip thickness (**Fig. 3D**). The nose-related score distribution, e.g. nose breadth and length, thus largely assembled the facial features in major continental groups, consistent with the hypotheses regarding adaptive evolution of human facial morphology (*Noback, Harvati, & Spoor, 2011; Weiner, 1954*), e.g. nose width and nose length have been found to correlate with temperature and humidity. Notably, the polygenic scores of the NSCL/P-associated SNPs explained up to 1.2% of the age- and sex- adjusted phenotypic variance for phenotypes mostly between nose and mouth, e.g. Right Alare – Right Cheilion and Pronasale – Right Cheilion (**Supplementary Fig. 30**).

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Gene ontology of face-associated genetic loci

A gene ontology enrichment analysis for the 24 face-associated genetic loci in **Table 1** highlighted a total of 67 biological process terms and seven molecular function terms significantly enriched with different genes group (FDR < 0.01 or q value < 0.01, **Supplementary Table 14**), with the top significant terms being 'embryonic digit morphogenesis' and 'embryonic appendage morphogenesis' (FDR<1×10^{-8}, **Supplementary Fig. 31**). Biological process terms were categorized in four broader categories, 'morphogenesis', 'differentiation', 'development' and 'regulation'. These results underlined the important role of embryonic developmental and regulatory processes in shaping human facial morphology. Furthermore, *cis*-eQTL effects of the 494 face-associated SNPs at the 24 loci (**Supplement Table 7**) were investigated using the GTEx data. Significant multi-tissue *cis*-eQTL effects were found around two genes, *ARHGEF19* and *RPGRIP1L* (**Supplementary Table 15**). No significant tissue-specific eQTL effects were observed, likely due to the lack of cell types indexed within the GTEx database that are directly involved in craniofacial morphology.

Preferential expression of face-associated genetic loci in embryonic cranial neural crest cells

Embryonic cranial neural crest cells (CNCCs) arise during weeks 3–6 of human gestation from the dorsal part of the neural tube ectoderm and migrate into the branchial arches. They later form the embryonic face, consequently establishing the central plan of facial morphology and determining species-specific and individual facial variation (*Bronner & LeDouarin, 2012; Cordero et al., 2011*). Recently, CNCCs have been successfully derived *in-vitro* from human embryonic stem cells or induced pluripotent stem cells (*Prescott et al., 2015*). Taking advantage of the available CNCC RNA-seq data (*Prescott et al., 2015*) and other public RNA-seq datasets compiled from GTEx (*G. T. Consortium, 2013*) and ENCODE (*E. P. Consortium, 2012*), we compared 24 genes, which were nearest to the top-associated SNPs at the 24 face-associated genetic loci identified in our discovery GWAS meta-analysis, with the background genes over the genome for their expression in CNCCs and 49 other cell types. Except in aorta, smooth muscle cells and endothelial cells, these 24 genes showed significantly higher expression in all cell types ($p<0.05$) compared to the background genes, and the

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

most significant difference was observed in CNCCs ($p=3.8\times10^{-6}$, **Fig. 4A**). Furthermore, these 24 genes also showed significant ($p<0.05$) preferential expression in CNCCs compared to their expression in other cell types, i.e., in 90.0% of 20 different tissue cells, in 90.0% of 20 primary cells, and in one third of 9 embryonic stem cells (**Fig. 4A**). Specifically, 17 (70.8%) of these 24 genes showed significant ($p<0.05$) preferential expression in CNCCs compared to 49 other cell types. The top-ranked three preferentially expressed genes were *PAX3* ($p=1.3\times10^{-29}$), *INTU* ($p=8.3\times10^{-25}$), and *ROR2* ($p=4.5\times10^{-23}$, **Fig. 4B**). Furthermore, some mid-preferentially expressed genes like *DCAF4L2* and non-preferentially expressed genes like *RPE65* became top-ranked preferentially expressed genes when testing sub-groups of cell types (**Fig. 4B**). Although not necessarily is the nearest gene the causative gene and not necessarily is there only one causative gene per locus, the observation of significant preferential expression of the nearby genes considered in this analysis does support the hypothesis that variation of facial shape may indeed originate during early embryogenesis (*Claes et al., 2018*), and provide a priority list in different stages of cell differentiation with genes likely involved in embryo ectoderm derived phenotypes such as facial variation for future in-vivo functional studies.

Cis-regulatory signals in face-associated genetic loci

It is becoming increasingly clear that c*is*-regulatory changes play a central role in determining human complex phenotypes (*Carroll, 2008; Wray, 2007*). In our study, apart from those at *KIF6* and *ROR2*, all face-associated DNA variants we identified were non-coding. With preferential expression pattern validated in CNCCs, we explored the potential *cis*-regulatory activity of the face-associated DNA variants using the CNCC epigenomic mapping datasets (*Prescott et al., 2015*). This dataset includes tracks on transcription factors (*TFAP2A* and *NR2F1*), general coactivator (p300), or histone modifications associated with active regulatory elements (H3K4me1, H3K4me3, and H3K27ac), and ATAC-seq evaluating chromatin accessibility. Abundant entries of *cis*-regulatory signals were detected surrounding the face-associated loci. In particular, 12 (50%) of the 24 face-associated loci, i.e., 1p12 *TBX15*, 1p36.13 *ARHGEF19,* 1p36.22 *CASZ1*, 2q36.1 *PAX3*, 3q12.1 *CMSS1,* 4q28.1 *INTU*, 6p22.3 *RNF144B*, 10q22.1 *SUPV3L1*, 16q12.1 *SALL1*, 16q12.2 *RPGRIP1L*/*FTO,* 17q24.3 *CASC17*, and 17q24.3 *SOX9* showed significant enrichment of *cis*-regulatory entries (top 1% of the genome) in at least one epigenomic track (Fisher's Exact Test $P<0.05$, **Supplementary Fig. 32**). Among these, five

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

loci contained SNPs with the highest *cis*-regulatory signals belonging to the top 0.1% of the genome, i.e. *ARHGEF19*, *CMSS1*, *INTU*, *RPGRIP1L* and *PAX3* (**Supplementary Fig. 32**). For example, in the novel locus 4q28.1, where the top-associated SNP rs12504954 is located at 606 kbp upstream of *INTU*, we identified three tagging SNPs (rs17210317, LD r^2=.92, top <0.1% in H3K27ac and H3K4me1; rs12510422, LD r^2=0.32, top <0.1% in p300, TFAP2A and ATAC and rs6828035, LD r^2=.76, top <1% in all six regulation factors) as the promising candidates for regulatory activity. Both regulatory activity and the pronounced H3K27ac and H3K4me1 signals indicating the existence of an active enhancer in this LD region (**Figs. 5A** and **5B**). In another novel locus 16q12.2, where the top-associated SNP rs7404301 is an intronic variant of *RPGRIP1L* (also ~2.5kb upstream *FTO*, **Fig. 5A**), two tagging SNPs (rs9673356, r=.57, top <0.1% in p300, TFAP2A and ATAC; rs4783818, r=.53, top <0.1% in H3K4me3) were identified as candidate variants for regulatory activity (**Fig. 5B**). In the well-established locus 2q36.1, where the top-associated SNP rs34032897 is located about 36kbp downstream to *PAX3*, we highlighted two tagging SNP (rs2855266, LD r^2=.36, top <0.1% in H3K27ac, p300, TFAP2A and ATAC and rs13410020, LD r^2=.30, top <1% in TFAP2A) among many other potential candidates for regulatory activity (**Fig. 5B**). These 5 candidate regulatory SNPs were further studied via functional experiments as described in the following section.

Experimental confirmation of cis-regulatory enhancer effects in neural crest progenitor cells

We performed *in-vitro* luciferase reporter assays to quantify the potential regulatory effect of the 5 face-associated SNPs highlighted by the analyses outlined above to be regulatory candidates i.e., rs13410020 and rs2855266 from the 2q36.1 *PAX3* region, rs17210317 and rs6828035 from the 4q28.1 *INTU* region, and rs4783818 from the 16q12.2 *RPGRIP1L* region. These selected SNPs are all located in putative enhancers defined as regions enriched in epigenetic marks associated with enhancer activity in CNCCs (*Prescott et al., 2015*) and penis foreskin melanocyte primary cells (*Chadwick, 2012; E. P. Consortium, 2012*). A luciferase reporter assay in CD271+ neural crest progenitor cells demonstrated that both SNPs in 2q36.1 and 4q28.1, respectively, showed statistically significant allele-specific effects on enhancer activity (**Fig. 5C**). In the 4q28.1 region, the derived SNP alleles were

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

associated with decreased luciferase expression of *INTU* (33.5×~10.4×, **Supplementary Table 16**) as well as a reduced length of some nose features (-0.087<beta<-0.078, **Supplementary Table 7**). In the 2q36.1 region, the derived SNP alleles were associated with increased luciferase expression of *PAX3* (1.5×~3.7×, **Supplementary Table 16**) as well as an increased length of some nasion-related features (0.131<beta<0.149, **Supplementary Table 7**). For rs4783818 from the 16q12.2 region, we obtained allele-specific effects on enhancer activity, although they were not statistically significant (**Fig. 5C**). Additional luciferase reporter assays using G361 melanoma cells provided similar results indicating that the regulatory activity associated with the variants is conserved across the neural crest cell lineage (**Supplementary Table 16**). These results provide evidence that the tested face-associated SNPs impact the activity of enhancer elements and thus have regulatory function.

Discussion

This study represents the largest genome scan to date examining facial shape phenotypes quantified from 3D facial images. We identified 24 genetic loci showing genome-wide significant association with normal-range facial shape phenotypes in Europeans, including 17 novel ones, of which 13 were successfully replicated in additional multi-ethnic population cohorts including 8 novel loci. Significant gene enrichment in morphogenesis related pathways and cis-regulation signals with preferential CNCC expression activities were observed at these genetic loci, underlining their involvement in facial morphology. Our findings were additionally supported by a strong integration with previous genetic association and functional studies. Moreover, via *in-vitro* cell culture experiments, we functionally exemplified the regulatory activity of selected face-associated SNPs at two newly proposed genes (*INTU* and *RPGRIP1L*) and one established face shape gene (*PAX3*).

The eight replicated novel loci included: 1p36.22 *CASZ1/PEX14*, 1p31.2 *RPE65/DEPDC1*, 4q28.1 *INTU*, 6p22.3 *RNF144B*, 6p21.2 *KIF6*, 12q24.21 *TBX3*, 14q32.2 *C14orf64*, and 16q12.2 *RPGRIP1L/FTO*. For 4q28.1 *INTU* and 16q12.2 *RPGRIP1L/FTO*, we demonstrated an impact of selected DNA variants on enhancer activity. *INTU*, also known as planar cell polarity (PCP) effector, has been reported to compromise proteolytic processing of *Gli3* by regulating Hh signal transduction (*Zeng, Hoover, & Liu,*

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2010). Notably, SNPs in the *GLI3* region have been previously associated with human nose wing breadth (*Adhikari et al., 2016*), supporting a link between *INTU* and facial morphogenesis. In addition, rare frameshift mutations in *INTU* have been found to cause human Oral-Facial-Digital (OFD) syndromes as characterized by abnormalities of the face and oral cavity (*Bruel et al., 2017*). A previous study found that *Rpgrip1l* mutant mouse embryos displayed facial abnormalities in cleft upper lips and hypoplastic lower jaws (*Delous et al., 2007*). *FTO* demethylase activity has been found essential for normal bone growth and bone mineralization in mice (*Sachse et al., 2018*). *TBX3*, a member of the *T-box* gene family, causes Ulnar-mammary syndrome which is a rare disorder characterized by distinct chin and nose shape together with many other abnormal morphological changes (*Bamshad et al., 1997; Joss et al., 2011*). Intriguingly, *TBX3* and *TBX15* associated facial phenotypes showed an asymmetric pattern, particularly for nose phenotypes (*TBX3*) and eye-mouth distances (*TBX15*). These two genes are thus good candidates for further studying the genetic basis of facial asymmetry and related disorders. A synonymous variant (rs7738892) in the transcript coding region of *KIF6* was associated with facial traits. It has been shown that the homologous *Kif6* transcript sequence was necessary for spine development in zebrafish, likely due to its role in cilia, which is known to utilize neural crest cells for regulating ventral forebrain morphogenesis (*Buchan et al., 2014*).

The five replicated previously reported face-associated loci included 2q36.1 *PAX3*, 4q31.3 *SFRP2/DCHS2*, 9q22.31 *ROR2*, 17q24.3 *CASC17*, and 20p11.22 *PAX1.* Their effects observed in the current study were all consistent with the previous findings, e.g. *PAX3* was associated with position of the nasion (*Liu et al., 2012; Paternoster et al., 2012*), *SFRP2/DCHS2* with distance between inner eye corners and alares (*Adhikari et al., 2016; Claes et al., 2018*), *PAX1* with nose wing breadth (*Adhikari et al., 2016; Shaffer et al., 2016*), and *ROR2* with nose size (*Pickrell et al., 2016*). The face-associated rs2230578 at 9q22.31 is in the 3'-UTR of *ROR2*. A *Ror2* knockout mouse has been used as a model for the developmental pathology of autosomal recessive Robinow syndrome, a human dwarfism syndrome characterized by limb shortening, vertebral and craniofacial malformations, and small external genitalia (*Schwabe et al., 2004*). A recent study confirmed a long-range mode of regulation of the *SOX9* gene through multiple association signals in the *SOX9* locus including a SNP in the *CASC17* region (rs9915190) (*Cha et al., 2018*).

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Among the nine non-replicated novel loci, two are worth mentioning: *KIAA1715* maps to about 100kbp upstream of the *HOXD* cluster (homeobox D cluster), which has been highlighted by two independent facial variation GWAS (*Claes et al., 2018; Pickrell et al., 2016*). *DCAF4L2* gene variants were previously associated with NSCL/P (*Leslie et al., 2017; Ludwig et al., 2012; Y. Yu et al., 2017*) and we revealed a significant effect on the distance between alares and the upper-lip. However, it is likely that true positive findings exist among the nine non-replicated novel loci. This is indicated by our finding that among the total of 11 non-replicated loci, 2 (18%) i.e., *TBX15* (*Claes et al., 2018; Pickrell et al., 2016*) and *SOX9* (*Cha et al., 2018; Claes et al., 2018; Pickrell et al., 2016*) correspond to previously identified face-associated loci. As such, their re-identification in our discovery GWAS meta-analysis represents a confirmation per se, even without additional confirmation in our replication analysis. Moreover, pleiotropic effects of these two genes on skeleton, e.g. stature, were previously demonstrated by a large-sized GWAS (*Wood et al., 2014*). Loss of function of *TBX15* in mice has been shown to cause abnormal skeletal development (*Singh et al., 2005*). Thus the fact that these two previously reported face-associated loci were identified in our discovery GWAS, but were not confirmed in our replication analysis, likely is explained by relatively low statistical power, especially for the European replication dataset (N=1,101), which is most relevant here as these two genes were highlighted in Europeans in previous and in our study. Moreover, for the European QIMR samples in the replication set, no direct replication analysis was possible due to the availability of 2D images but not 3D images as in the discovery dataset. Therefore, future studies shall further investigate these nine face-associated loci.

Of the previously identified face-associated genetic loci, several candidate genes e.g. *Pax3* (*Zalc, Rattenbach, Aurade, Cadot, & Relaix, 2015*), *Prdm16 (Bjork, Turbe-Doan, Prysak, Herron, & Beier, 2010)*, *Tp63 (Thomason, Dixon, & Dixon, 2008)*, *Sfrp2* (*Satoh, Matsuyama, Takemura, Aizawa, & Shimono, 2008*), *Gli3* (*Hui & Joyner, 1993*) have been shown to play important roles in cranial development of embryonic mice. Moreover, in adult mice only *Edar* has been investigated for its effect on normal facial variation (*Adhikari et al., 2016*), but the respective face-associated missense *EDARV370A* is nearly absent in Europeans (*Kamberov et al., 2013*). Here we showed via direct experimental evidence in neural crest cells that face-associated SNPs in *INTU*, *PAX3,* and *RPGRIP1L* are likely involved in enhancer functions and regulate gene activities. Future studies shall

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

demonstrate functional evidence for more SNPs and genes identified here or previously to be significantly association with facial shape phenotypes. As demonstrated here, this shall go beyond evidence on the bioinformatics level (*Claes et al., 2018*) and shall involve direct functional testing of the associated DNA variants in suitable cell lines.

Our results indirectly support the hypothesis that the facial differentiation observed among human populations worldwide is a consequence of positive selection. For some facial features, it is reasonable to speculate a primary selective pressure on the morphological effects of the associated variants, although many of these variants may have pleiotropic effects on other fitness-related phenotypes. For example, nasal shape, which is essential for humidifying and warming the air before it reaches the sensitive lungs, has been found to significantly correlate with climatic variables such as temperature and humidity in human populations (*Noback et al., 2011*). Notably, many of the SNPs highlighted in our study are involved in nasal phenotypes. The overall effect of these SNPs as measured by polygenic scores largely assembled the nasal shape differences among major continental groups, which implies that human facial variation in part is caused by genetic components that are shared between major human populations, although there are existing examples of population-specific effects. For example, *EDAR* clearly demonstrates an Asian specific effect, which impacts on various ectoderm-derived appearance traits such as hair morphology, size of the breast, facial shape in Asians (*Adhikari et al., 2015; Kamberov et al., 2013; Kimura et al., 2009; Tan et al., 2013*). The respective face-associated missense DNA variant is rare European or African populations, demonstrating regional developments in Asia. Nevertheless, the polygenic face scores only explained small proportions of the facial phenotype variance with up to 4.4%. This implies that many more face-predictive genetic information remains to be discovered in future studies, which - in case enough - may provide the prerequisite for practical applications of predicting human facial information from genomic data such as in forensics or anthropology. On the other hand, the increasing capacity of revealing personal information from genomic data may also have far-reaching ethical, societal and legal implications, which shall be broadly discussed by the various stakeholders alongside the genomic and technological progress to be made in the future.

In summary, this study identified multiple novel and confirmed several previously reported genetic loci influencing facial shape variation in humans. Our findings extend knowledge on natural

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

389 selection having shaped the genetic differentiation underlying human facial variation. Moreover, we
390 demonstrated the functional effects of several face-associated genetic loci as well as face-associated
391 DNA variants with in silico bioinformatics analyses and *in-vitro* cell line experiments. Overall, our
392 study strongly enhances genetic knowledge on human facial variation and provides candidate genes
393 for future *in-vivo* functional studies.
394
395

396 Acknowledgements

397 We are extremely grateful to all participants and their families who took part in the RS, TwinsUK,
398 ALSPAC, PITT, UYG, CANDELA, and QIMR studies as well as the many individuals involved in the
399 running of the studies and recruitment, which include midwives, interviewers, computer and
400 laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and
401 nurses. M.K. additionally thanks Maren von Köckritz-Blickwede for earlier discussions on PAD
402 knockout mice. **QIMR** would like to thank Natalie Garden and Jessica Adsett for data collection and
403 Scott Gordon for data management.
404

405 Funding

406 This work was financially supported in part by the European Union's Horizon 2020 Research and
407 Innovation Programme under grant agreement No 740580 (VISAGE). F.L. is supported by Erasmus
408 University Fellowship, Chinese National Thousand Young Talents Award, the National Natural Science
409 Foundation of China (NSFC) (91651507), the National Key R & D program of China
410 (2017YFC0803501), and the Strategic Priority Research Program of Chinese Academy of Sciences
411 (XDC01000000). Z.X. is supported by the China Scholarship Council (CSC).
412 **The Rotterdam Study** is supported by the Netherlands Organization of Scientific Research NWO
413 Investments (nr. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for
414 Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands
415 Organization for Scientific Research (NWO) project nr. 050-060-810. The Rotterdam Study is
416 supported by the Erasmus MC and Erasmus University Rotterdam; the Netherlands Organization for

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Scientific Research (NWO); the Netherlands Organization for Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly (RIDE) the Netherlands Genomics Initiative (NGI); the Ministry of Education, Culture and Science; the Ministry of Health Welfare and Sport; the European Commission (DG XII); and the Municipality of Rotterdam. The generation and management of GWAS genotype data for the Rotterdam Study were executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC.

The TwinsUK study is funded by the Wellcome Trust, Medical Research Council, European Union (FP7/2007-2013), the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London.

The ALSPAC was supported by The UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the University of Bristol. A comprehensive list of grants funding for ALSPAC is available online (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf). L.H., G.S., S.L. and E.S. work in a unit that receives funding from the University of Bristol and the UK Medical Research Council (MC_UU_00011/1). This publication is the work of the authors and L.H. and E.S. will serve as guarantors for the contents of this paper. GWAS data was generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe.

The PITT study was funded by the National Institute for Dental and Craniofacial Research (http://www.nidcr.nih.gov/) through the following grants: U01- DE020078, R01-DE027023, R01-DE016148, and X01-HG007821. The Centers for Disease Control (https://www.cdc.gov/) provided funding through the following grant: R01- DD000295. Funding for initial genomic data cleaning by the University of Washington was provided by contract #HHSN268201200008I from the National Institute for Dental and Craniofacial Research (http://www.nidcr.nih.gov/) awarded to the Center for Inherited Disease Research (CIDR).

The CANDELA study was funded by grants from: The Leverhulme Trust (F/07 134/DF to ARL), BBSRC (BB/I021213/1 to ARL); Universidad de Antioquia, Colombia (CPT-1602, Estrategia para sostenibilidad 2016-2018 grupo GENMOL to GB). The following institutions kindly provided facilities for the assessment of volunteers: Escuela Nacional de Antropología e Historia and Universidad Nacional

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Autónoma de México (México); Pontificia Universidad Católica del Perú, Universidad de Lima and Universidad Nacional Mayor de San Marcos (Perú); Universidade Federal do Rio Grande do Sul (Brazil); 13° Companhia de Comunicações Mecanizada do Exército Brasileiro (Brazil).

The UYG studies were supported by the National Natural Science Foundation of China (NSFC) (grant numbers 91631307 to S.W., 91731303, 31771388, 31525014 and 31711530221 to S.X., 30890034 and 31271338 to L.J.), the Science and Technology Commission of Shanghai Municipality (grant number 16JC1400504 to L.J. and S.W.), the Chinese Academy of Sciences (grant number XDB13041000 to S.W. and S.X., QYZDJ-SSW-SYS009 to S.X.), the National Basic Research Program (2015FY111700 to L.J. and S.W.), the Program of Shanghai Academic Research Leader (16XD1404700), the National Key Research and Development Program (2016YFC0906403 to S.X. and 2017YFC0803501-Z04 to K.T.), the National Program for Top-notch Young Innovative Talents of The "Wanren Jihua" Project to S.X., the National Thousand Young Talents Award and the Max Planck-CAS Paul Gerson Unna Independent Research Group Leadership Award both to S.W.

QIMR funding was in part provided by Australian NHMRC grants to N.G.M and an Australian Research Council Linkage Grant (LP 110100121: P.I. Dennis McNevin, University of Canberra). S.E.M. is supported by an NHMRC fellowship APP1103623.

Author contributions

M.K., S.B., N.G.M. and T.D.S. conceived this study as part of the International Visible Trait Genetics Consortium activities. Z.X. performed the main statistical analysis, G.D. performed functional experiments. L.J.H., M.K.L., P.G.H., M.A. de J., G.Z., K.A., D.L., Y.L. each conducted part of the phenotype ascertainment analyses and/or statistical analyses described in this work. E.F., M.L.M., J.R.S., K.M., S.X., L.J., S.W., F.M. de V., B.L., A.G.U., M.A.I., E.B.W., S.A.K., T.N., S.R., A.Z., S.L., G.S., L.P., H.T., S.C.Q., G.B., C.G., G.P., F.R., M.C.B., R.G.J., K.T., S.E.M. contributed samples or data. R.J.P. helped supervising functional experiments. T.D.S., E.S., S.M.W., F.L. and M.K. jointly supervised the work. Z.X., F.L. and M.K. wrote the manuscript with contributions from G.D., L.J.H., P.G.H., B.P., J.R.S., A.R.L., S.E.M., N.G.M., E.S., S.M.W. All authors approved the final manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Conflict of Interest Statement

The authors declared that they have no conflicts of interest to this work.

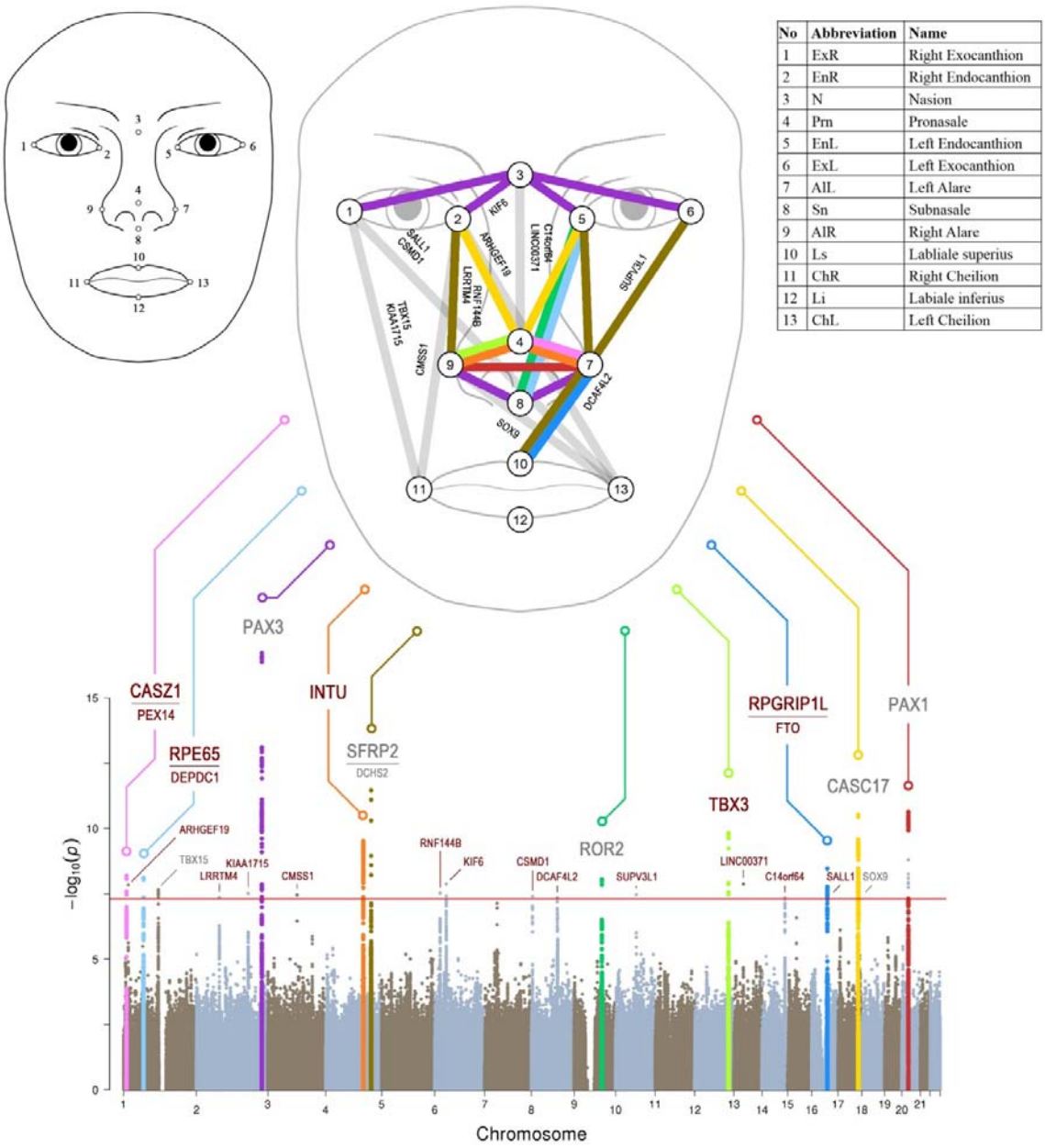
bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figures

Fig. 1. Manhattan plot from meta-analysis discovery GWAS (N = 10,115 Europeans) for 78 Euclidean distances between 13 facial landmarks.

Results from meta-analysis of four GWASs in Europeans (N=10115), which were separately conducted in RS, PITT, ALSPEC, and TwinsUK, for 78 facial shape phenotypes that are displayed in a signal 'composite' Manhattan plot at the bottom of the figure. All loci consisting of SNPs reaching significance (p<5e-8, red line) were nominated according to the nearby candidate genes in the associated loci, and the top 10 loci were highlighted in colors. Novel and previously established face-associated genetic loci were differentiated using colored and grey text, respectively. Annotation of the 13 facial landmarks used to derive the 78 facial phenotypes and the associated facial phenotypes are illustrated in the upper part of the figure.

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



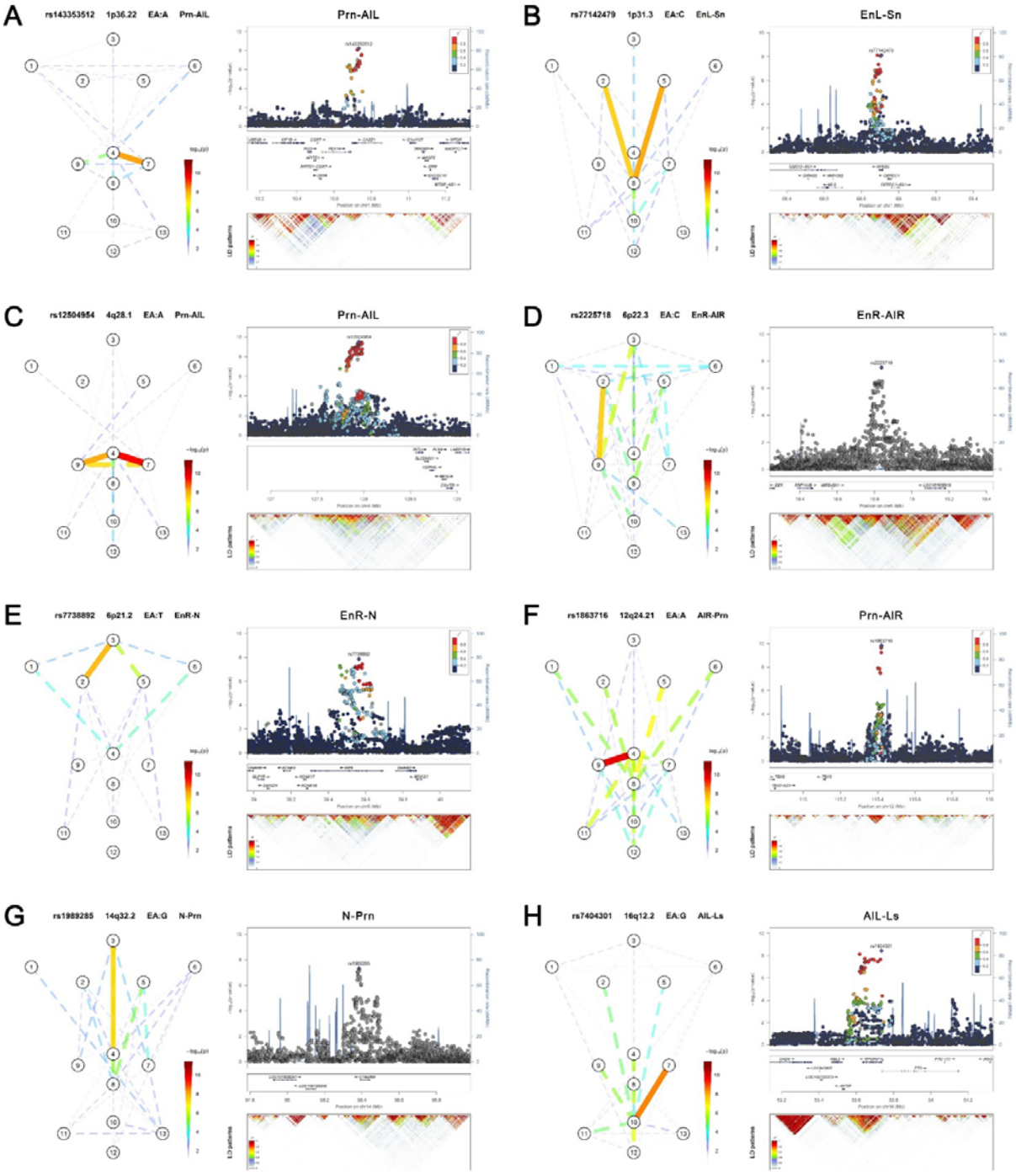
No	Abbreviation	Name
1	ExR	Right Exocanthion
2	EnR	Right Endocanthion
3	N	Nasion
4	Prn	Pronasale
5	EnL	Left Endocanthion
6	ExL	Left Exocanthion
7	AlL	Left Alare
8	Sn	Subnasale
9	AlR	Right Alare
10	Ls	Labliale superius
11	ChR	Right Cheilion
12	Li	Labiale inferius
13	ChL	Left Cheilion

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 2. Eight novel genetic loci associated with facial shape phenotypes.

The figure is composed by 8 parts, (**A**) 1p36.22 *CASZ1* region, (**B**) 1p31.2 *RPE65* region, (**C**) 4q28.1 *INTU* region, (**D**) 6p22.3 *RNF144B* region, (**E**) 6p21.2 *KIF6* region, (**F**) 12q24.21 *TBX3* region, (**G**) 14q32.2 *C14orf64* region, and (**H**) 16q12.2 *RPGRIP1L* region. Each part includes two figures, Face map (left) denoting all of the top-SNP-associated phenotypes weighted by the statistical significance ($-\log_{10}P$). LocusZoom (right) shows regional Manhattan plots for the top-associated facial phenotypes with candidate genes aligned below according to the chromosomal positions (GRCh37.p13) followed by linkage disequilibrium (LD) patterns (r^2) of EUR in the corresponding regions. Similar figures for all 24 loci are provided for all face associated loci in **Figures S6 to S29**.

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

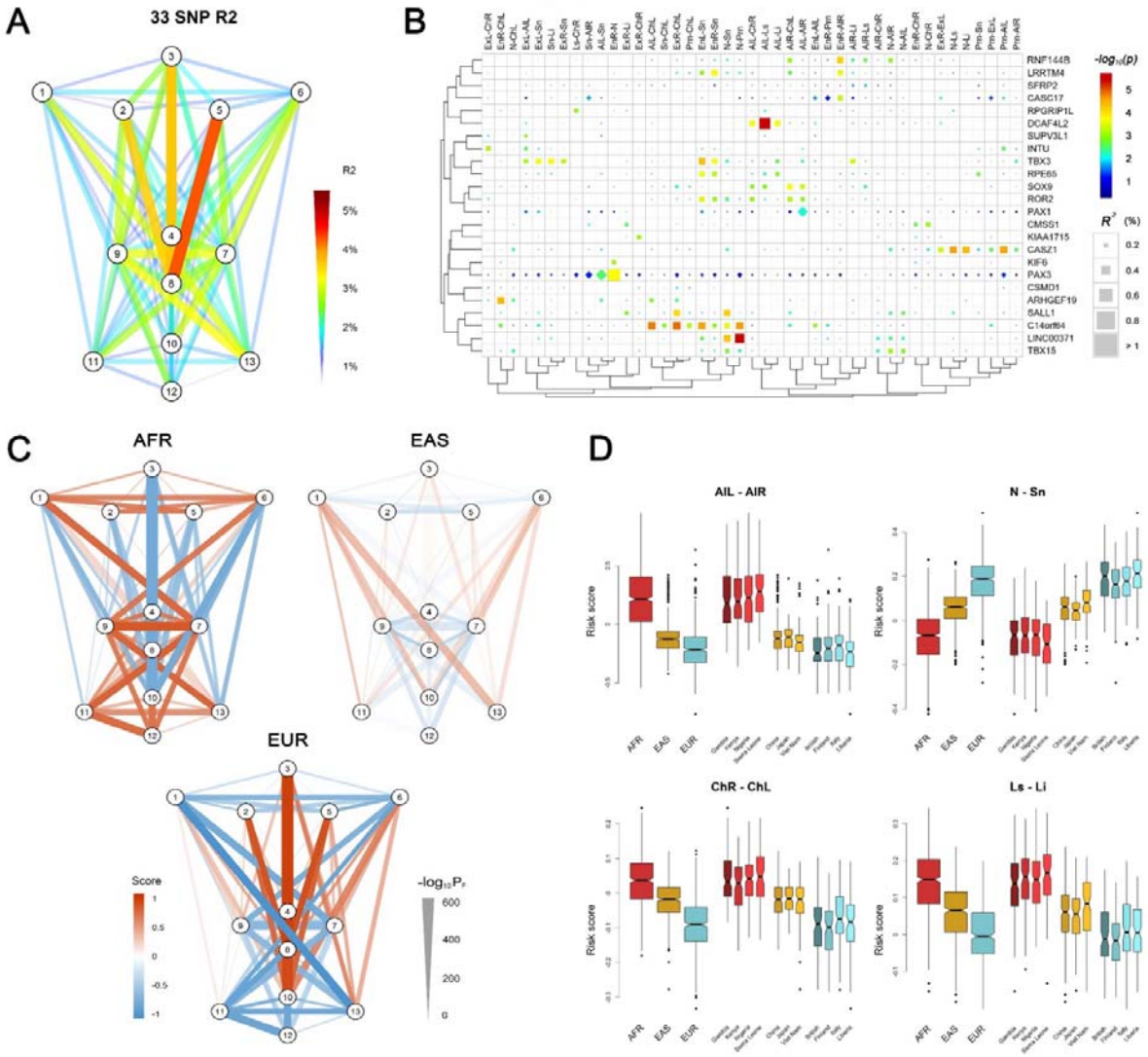


bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 3. Polygenic scores of 78 facial phenotypes.

A, Facial phenotype variance is explained by a multivariable model consisting of 33 face-associated SNPs. **B**, Unsupervised clustering reveals the details on the explained phenotype variance, where the SNPs were attributed to the candidate genes in the associated regions. All phenotypes with the total explained variance > 2% are shown. SNPs were colored according to their raw p-values and those passing FDR < 0.05 were indicated in squares. **C**, Mean standardized polygenic scores calculated by applying the multivariable model trained in the RS discovery sample to Africans (AFR), East Asians (EAS) and Europeans (EUR) in the 1000 Genomes Project data. **D**, Examples of the distributions of the face scores in samples of different ancestry origin/country.

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

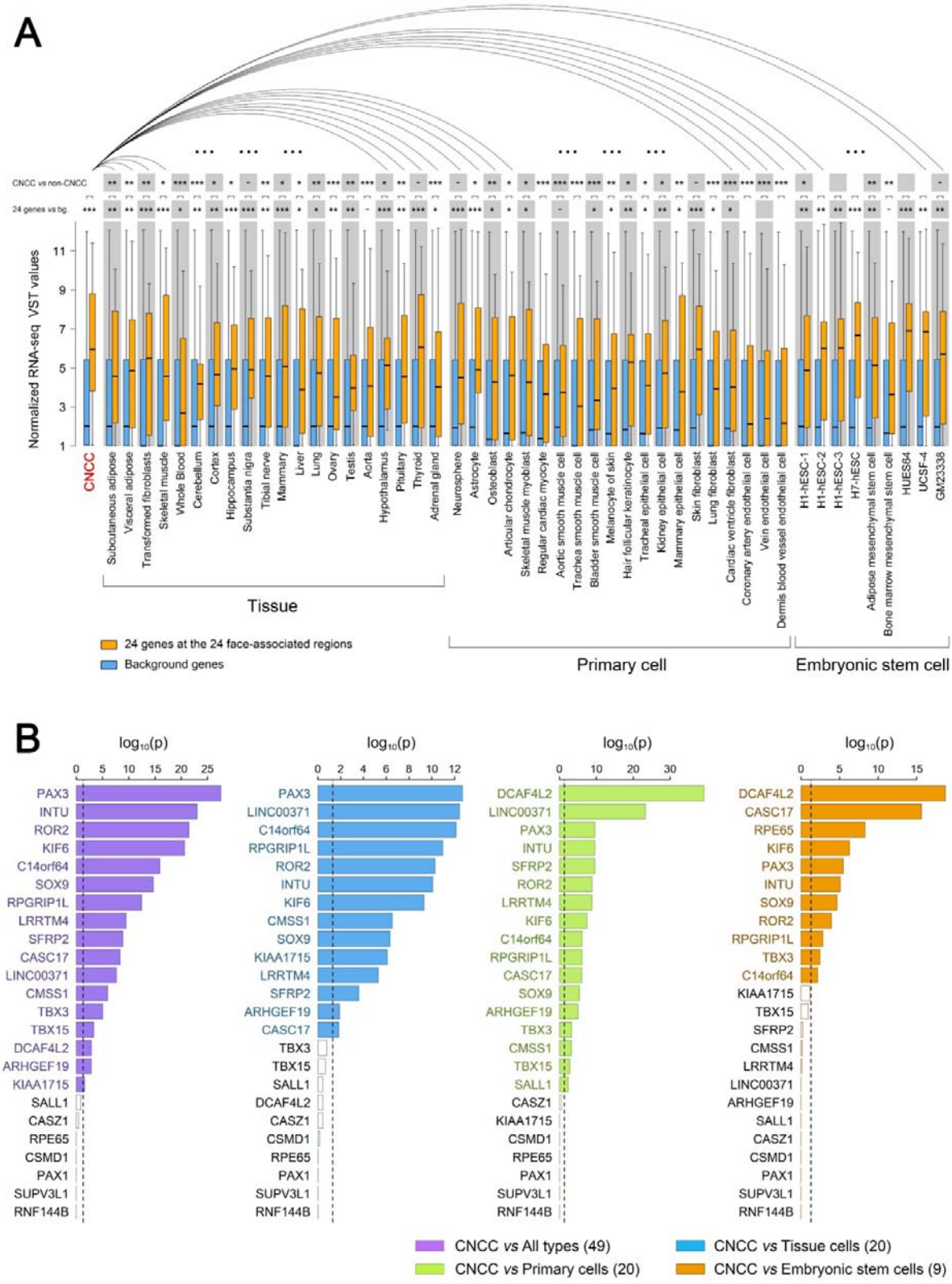


bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 4. Expression of 24 genes located at the identified 24 face-associated genetic loci in 50 cell types.

Expression levels of the genes nearest to the 24 top-associated SNPs (see **Table 1**) in 50 cell types were displayed in boxplots based on published data from the study of Prescott et al.(*Prescott et al., 2015*), the GTEx database(*G. T. Consortium, 2013*) and the ENCODE database(*E. P. Consortium, 2012*). The cell types included CNCC, 20 other tissue cells, 20 primary cells and 9 embryonic stem cells. **A**, Boxplots of normalized RNA-seq VST values for the 24 genes (in orange) and all genome background genes (~50,000 genes, in blue). Expression difference between the genome background genes and the 24 genes was tested in each cell type using the unpaired Wilcoxon rank-sum test. The expression of the 24 genes in CNCCs was also iteratively compared with that in other cell types using paired Wilcoxon rank-sum test. Statistical significance was indicated: ¯p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001. **B**, Difference significance of normalized RNA-seq VST values in each of 24 genes were displayed using barplots between CNCCs and all 49 types of cells (purple), CNCCs and 20 tissue cells (blue), CNCCs and 20 primary cells (green), CNCCs and 9 embryonic stem cells (orange). The difference between the value of CNCCs and other cell type groups was tested using the one-sample Student's t test and $-\log_{10}$(P) values after 24 genes FDR correction (y-axis). Dotted line represents significant threshold (0.05 for corrected p values). Significant gene labels were in color compared to non-significant gene labels in black.

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 5. Three examples of genetic loci associated with facial shape phenotypes.

Three examples of facial shape associated loci we discovered are depicted in details, including two novel loci (*INTU*; *RPGRIP1L*) and one (*PAX3*) that overlaps with previous face GWAS findings. The figure is composed by 3 layers (**A-C**) organized from the top to the bottom. (**A**) Denotes the linkage disequilibrium (LD) patterns (r^2) of EUR in the corresponding regions and epigenetic annotation of CNCC's regulatory elements in corresponding regions using WashU Epigenome Browser. Chip-seq profiles display histone modifications associated with active enhancers (H3K27ac, H3K4me1) or promoters (H3K4me3), and binding of general coactivator p300 and transcription factor TFAP2A. The ATAC-seq track shows chromatin accessibility. The PhyloP track indicates cross-species conservation. (**B**) SNPs in LD ($r^2 > 0.25$) with the top-associated SNP and all base-pairs in vicinity (within 20kb, $p < 0.05$) are displayed according to their log scaled quantile rank ($-\log_{10}p$, axis scale) in the circular figure, where the rank is calculated using the Chip-seq values in the whole genome. In addition, SNPs associated with facial phenotypes in our meta-analysis of GWASs were highlighted according to their association p values (P_{Meta}, color scale) and their LD (r^2) with the top-associated SNP in the region (points size). (**C**) Shows results of luciferase reporter assays in CD271+ neural crest progenitor cells by which we tested allele-specific enhancer activity of putative enhancers surrounding the selected SNPs using the Student´s t-test. Data are presented as relative luciferase activity (firefly/renilla activity ratio) fold change compared to the construct containing less active allele. Similar figures for layers **A** and **B** are provided for all face associated loci in **Figures S6 to S29.**

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

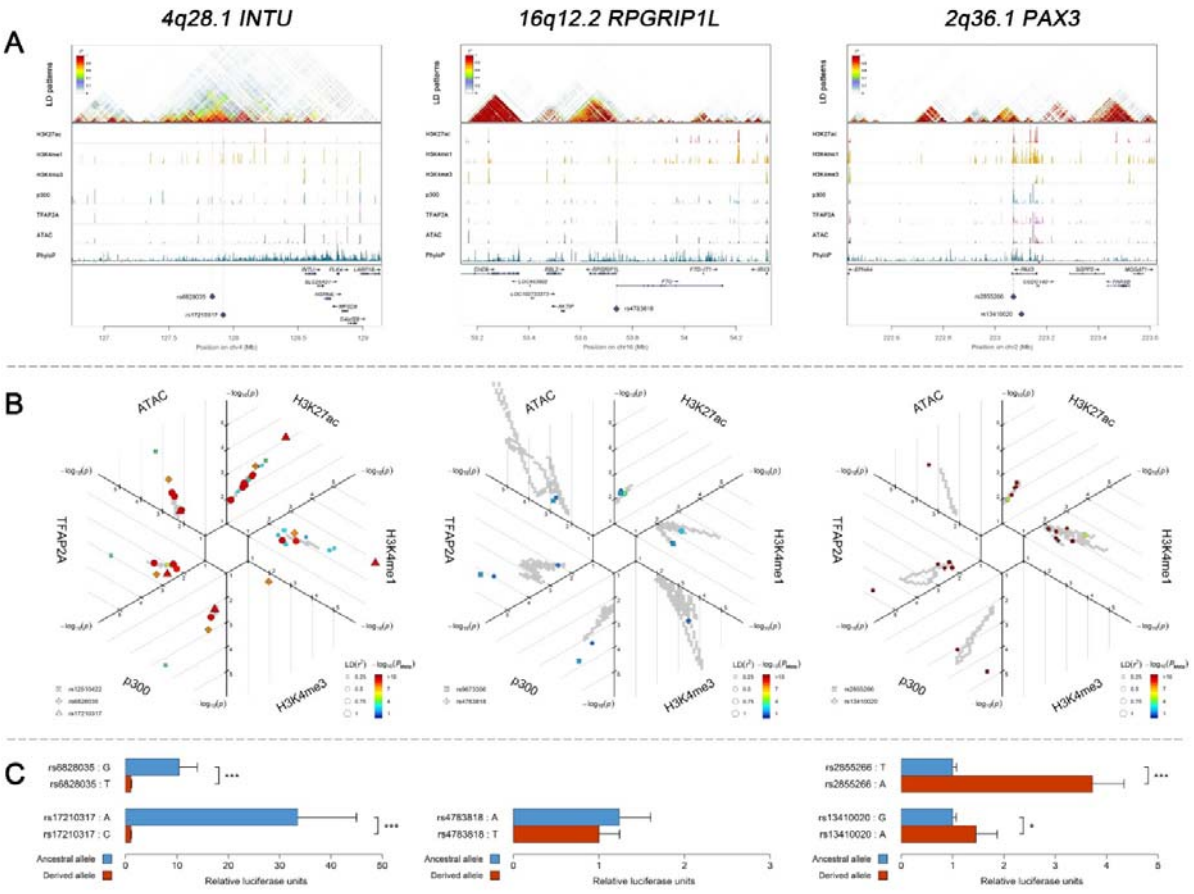


Table 1. SNPs significantly associated with facial shape phenotypes from meta-analysis of European discovery GWASs in RS, TwinksUK, ALSPAC, PITT, and their replication in UYG, CANDELA and QIMR.

						Meta-analysis (N=10,115)		UYG (N=858)			CANDELA (N=5,958)			QIMR (N=1,101)		
Nearest Genes	Region	SNP	Alleles		Trait	Beta	P	Beta	P	% (78)	Min P	Com. P	% (14)	Min P	Com. P	% (78)
CASZ1/PEX14	1p36.22	rs143353512	A	G	Prn-AlL	-0.29	6.44×10^{-9}	-0.05	0.7047	14.10%	**0.0017**	**0.0218**	14.29%	0.0089	0.8103	2.56%
ARHGEF19	1p36.13	rs200243292	I	T	EnR-ChL	-0.11	1.46×10^{-8}	-	-	-	0.0151	0.6085	14.29%	-	-	-
RPE65/DEPDC1	1p31.2	rs77142479	C	A	EnL-Sn	-0.08	7.65×10^{-9}	-0.04	0.3821	0.00%	0.0048	**0.0376**	14.29%	-	-	-
TBX15	1p12	rs1229119	T	C	ExR-ChR	0.08	2.25×10^{-8}	0.02	0.5520	5.13%	0.1556	0.9455	0.00%	0.0457	0.4747	2.56%
LRRTM4	2p12	rs10202675	T	C	EnR-AlR	-0.26	4.43×10^{-8}	0.21	0.2125	5.13%	0.0039	0.7379	7.14%	0.0050	0.2334	3.85%
KIAA1715	2q31.1	rs2884836	T	C	ExR-ChR	-0.09	3.04×10^{-8}	-0.01	0.8488	3.85%	0.0045	0.1099	14.29%	0.1628	0.9602	0.00%
PAX3	2q36.1	rs34032897	G	A	EnR-N	0.14	1.96×10^{-17}	0.16	**0.0121**	6.41%	0.0164	**0.0410**	14.29%	0.0168	0.5091	2.56%
CMSS1	3q12.1	rs113663609	A	G	EnR-ChR	-0.12	3.46×10^{-8}	-0.12	0.0690	5.13%	0.0563	0.0974	0.00%	0.0285	0.5299	5.13%
INTU	4q28.1	rs12504954	A	G	Prn-AlL	-0.09	3.03×10^{-10}	-0.07	0.0611	3.85%	**0.0022**	**0.0026**	28.57%	0.0069	0.7939	5.13%
SFRP2	4q31.3	rs6535972	C	G	EnL-AlL	0.12	3.51×10^{-12}	0.08	0.0713	28.21%	$\mathbf{2.04\times10^{-5}}$	$\mathbf{4.00\times10^{-7}}$	35.71%	0.0196	0.5677	1.28%
RNF144B	6p22.3	rs2225718	C	T	EnR-AlR	-0.09	2.97×10^{-8}	0.01	0.7764	2.56%	0.0228	**0.0279**	7.14%	0.0124	0.1821	10.26%
KIF6	6p21.2	rs7738892	T	C	EnR-N	0.11	1.34×10^{-8}	0.06	0.2988	0.00%	0.0338	0.1880	7.14%	$\mathbf{7.23\times10^{-5}}$	**0.0262**	12.82%
CSMD1	8p23.2	rs1700048	C	A	ExR-ChL	-0.22	3.94×10^{-8}	-0.19	0.2329	8.97%	0.0608	0.5228	0.00%	0.1414	0.9627	0.00%
DCAF4L2	8q21.3	rs9642796	A	G	AlL-Ls	-0.11	4.52×10^{-8}	0.06	0.1537	2.56%	0.0479	0.2676	7.14%	0.0084	0.3457	8.97%
ROR2	9q22.31	rs2230578	C	T	EnL-Sn	0.09	9.02×10^{-9}	0.09	**0.0352**	29.49%	**0.0032**	0.0564	14.29%	0.0180	0.1995	6.41%
SUPV3L1	10q22.1	rs201719697	A	D	ExL-AlL	-0.28	3.42×10^{-8}	-	-	-	-	-	-	-	-	-
TBX3	12q24.21	rs1863716	A	C	Prn-AlR	-0.09	1.47×10^{-10}	-0.01	0.7221	6.41%	**0.0010**	**0.0067**	21.43%	0.0974	0.9044	0.00%
LINC00371	13q14.3	rs7325564	T	C	N-Prn	0.09	1.34×10^{-8}	-0.01	0.8444	0.00%	0.1989	0.8095	0.00%	0.0058	0.3272	7.69%
C14orf64	14q32.2	rs1989285	G	C	N-Prn	-0.10	4.54×10^{-8}	0.03	0.4844	0.00%	**0.0001**	**0.0350**	14.29%	0.0498	0.6735	1.28%
SALL1	16q12.1	rs16949899	T	G	ExR-ChL	-0.13	3.35×10^{-8}	0.01	0.7847	0.00%	0.1543	0.9589	0.00%	0.0465	0.5553	1.28%
RPGRIP1L/FTO	16q12.2	rs7404301	G	A	AlL-Ls	-0.09	3.49×10^{-9}	-0.08	**0.0319**	17.95%	$\mathbf{1.88\times10^{-5}}$	$\mathbf{5.36\times10^{-5}}$	21.43%	0.0737	0.5233	0.00%
CASC17	17q24.3	rs8077906	A	G	Prn-EnL	-0.09	3.00×10^{-11}	-0.01	0.8076	0.00%	0.0183	**0.0446**	7.14%	0.0081	0.4746	5.13%
SOX9	17q24.3	rs35473710	G	A	AlR-ChL	0.19	3.95×10^{-8}	0.02	0.8665	0.00%	0.0648	0.2872	0.00%	0.0117	0.3604	5.13%
PAX1(LINC01427)	20p11.22	rs4813454	T	C	AlL-AlR	-0.10	2.32×10^{-11}	-0.12	**0.0060**	5.13%	**0.0003**	**0.0081**	14.29%	0.0460	0.6772	1.28%

rs2230578 is in the 3'-UTR of ROR2. rs7738892 in KIF6 is a synonymous variant. Other SNPs are intronic or intergenic variants. Gene symbols in bold indicate novel loci,. For alleles column, the left one is effective allele and the right one is other allele. P column in bold at p< 0.05 for UYG, Min P column in bold at p<0.0036 (0.05/14) for CANDELA and p<0.0014 (0.05/36) for QIMR. Com. P: p-value of a combined test of dependent p-values in bold at p< 0.05. % (78): the percentage of P values < 0.05 for 78 facial trait tests.

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Materials and methods

Cohort details

The four discovery cohorts, totaling 10,115 individuals, were Rotterdam Study (RS, N=3,193, North-Western Europeans from the Netherlands), TwinsUK (N=1,020, North-Western Europeans from the UK), ALSPAC (N=3,707, North-Western Europeans from the UK), and Pittsburgh 3D Facial Norms study (PITT, N=2,195, European ancestry from the United States). The three replication cohorts, totaling 7,917 individuals, were CANDELA Study (N=5,958, Latin Americans with ancestry admixture estimated at 48% European, 46% Native American and 6% African), Xinjiang Uyghur Study (UYG, N=858, Uyghurs from China with ancestry admixture estimated at 50% East Asian and 50% European), and Queensland Institute of Medical Research Study (QIMR, N = 1,101, North-Western European ancestry from Australia). Three-dimensional facial surface images were acquired in all discovery cohorts and UYG, while frontal 2-dimensional photographs were used in CANDELA and QIMR.

Rotterdam Study (RS). The RS is a population based cohort study of 14,926 participants aged 45 years and older, living in the same suburb of Rotterdam, the Netherlands (*Hofman et al., 2013*). The present study includes 3,193 participants of Dutch European ancestry, for whom high-resolution *3dMDface* digital photographs were taken. Genotyping was carried out using the Infinium II HumanHap 550K Genotyping BeadChip version 3 (Illumina, San Diego, California USA). Collection and purification of DNA have been described previously (*Kayser et al., 2008*). All SNPs were imputed using MACH software (www.sph.umich.edu/csg/abecasis/MaCH/) based on the 1000-Genomes Project reference population information (*Genomes Project et al., 2012*). Genotype and individual quality controls have been described in detail previously (*Lango Allen et al., 2010*). After all quality controls, the current study included a total of 6,886,439 autosomal SNPs (MAF > 0.01, imputation R2 > 0.8, SNP call rate > 0.97, HWE > 1e-4). The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC (registration number MEC 02.1015) and by the Dutch Ministry of Health, Welfare and Sport (Population Screening Act WBO, license number 1071272-159521-PG). The Rotterdam Study has been entered into the Netherlands National Trial Register (NTR; R; www.trialregister.nl) an) and into the WHO International Clinical Trials Registry Platform (ICTRP; P; www.who.int/ictrp/network/primary/en/) und under shared catalogue number NTR6831. All

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

participants provided written informed consent to participate in the study and to have their information obtained from treating physicians.

TwinsUK study. The TwinsUK study included 1,020 phenotyped participants (all female and all of Caucasian ancestry) within the TwinsUK adult twin registry based at St. Thomas' Hospital in London. All participants provided fully informed consent under a protocol reviewed by the St. Thomas' Hosptal Local Research Ethics Committee. Genotyping of the TwinsUK cohort was done with a combination of Illumina HumanHap300 and HumanHap610Q chips. Intensity data for each of the arrays were pooled separately and genotypes were called with the Illuminus32 calling algorithm, thresholding on a maximum posterior probability of 0.95 as previously described (*Small et al., 2011*). Imputation was performed using the IMPUTE 2.0 software package using haplotype information from the 1000 Genomes Project (Phase 1, integrated variant set across 1092 individuals, v2, March 2012). After all quality controls, the current study included a total of 4,699,858 autosomal SNPs (MAF > 0.01, imputation R2 > 0.8, SNP call rate > 0.97, HWE > 1e-4) and 1,020 individuals.

ALSPAC study. The Avon Longitudinal Study of Parents and Children (ALSPAC) is a longitudinal study that recruited pregnant women living in the former county of Avon in the United Kingdom, with expected delivery dates between 1st April 1991 and 31st December 1992. The initial number of enrolled pregnancies was 14,541 resulting in 14,062 live births and 13,988 children alive at the age of 1. When the oldest children were approximately 7 years of age, additional eligible cases who had failed to join the study originally were added to the study. Full details of study enrolment have been described in detail previously (*Boyd et al., 2013; Fraser et al., 2013; Golding, Pembrey, Jones, & Team, 2001*). A total of 9,912 ALSPAC children were genotyped using the Illumina HumanHap550 quad genome-wide SNP genotyping platform. Individuals were excluded from further analysis based on incorrect sex assignments; heterozygosity (0.345 for the Sanger data and 0.330 for the LabCorp data); individual missingness (>3%); cryptic relatedness (>10% IBD) and non-European ancestry (detected by a multidimensional scaling analysis seeded with HapMap 2 individuals). The resulting post-quality control dataset contained 8,237 individuals. The post-quality control ALSPAC children were combined with the ALSPAC mothers cohort (*Fraser et al., 2013*) and imputed together using a subset of markers common to both the mothers and the children. The combined sample was pre-phased using ShapeIT (v2.r644) (*Delaneau, Marchini, & Zagury, 2011*) and imputed to the 1000 Genomes reference panel (Phase 1, Version3) (***Genomes Project et al., 2015***) using IMPUTE3 V2.2.2 (*Howie,*

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Donnelly, & Marchini, 2009). The present study sample included 3,707 individuals with 3D facial images, genotype and covariate data. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Please note that the study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool: http://www.bristol.ac.uk/alspac/researchers/our-data/.

Pittsburgh 3D Facial Norms (PITT) sample, United States. The current study included 2,195 individuals from the 3D Facial Norms dataset (*Weinberg et al., 2016*). These individuals were recruited at four US sites (Pittsburgh, Seattle, Houston, and Iowa City), were of self-identified European ancestry, and ranged in age from 3 to 49 years. Exclusion criteria included any individuals with a history of significant facial trauma, congenital defects of the face, surgical alteration of the face, or any medical conditions affecting the facial posture. DNA was extracted from saliva samples and genotyped along with 72 HapMap control samples for 964,193 SNPs on the Illumina (San Diego, CA) HumanOmniEx- press+Exome v1.2 array plus 4,322 SNPs of custom content by the Center for Inherited Disease Research (CIDR). Samples were interrogated for genetic sex, chromosomal aberrations, relatedness, genotype call rate, and batch effects. SNPs were interrogated for call rate, discordance among duplicate samples, Mendelian errors among HapMap controls (parent-offspring trios), deviations from Hardy-Weinberg equilibrium, and sex differences in allele frequencies and heterozygosity. To assess population structure, we performed principal component analysis (PCA) using subsets of uncorrelated SNPs. Based on the scatterplots of the principal components (PCs) and scree plots of the eigenvalues, we determined that population structure was captured in four PCs of ancestry. Imputation of unobserved variants was performed using haplotypes from the 1000 Genomes Project Phase 3 as the reference. Imputation was performed using IMPUTE2. We used an info score of >0.5 at the SNP level and a genotype probability of >0.9 at the SNP-per-participant level as filters for imputed SNPs. Masked variant analysis, in which genotyped SNPs were imputed in order to assess imputation quality, indicated high accuracy of imputation. Institutional Review Board (IRB) approval was obtained at each recruitment center and all subjects gave written informed consent prior to participation: University of Pittsburgh IRB #PRO09060553 and IRB0405013; Seattle Children's IRB #12107; University of Iowa Human Subjects Office/IRB #200912764 and #200710721; UT Health Committee for the Protection of Human Subjects #HSC-DB-09-0508 and #HSC-MS-03-090.

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Xinjiang Uyghur (UYG) study. The UYG samples were collected at Xinjiang Medical University in 2013-2014. In total, 858 individuals (including 333 males and 525 females, with an age range of 17-25) were enrolled. The research was conducted with the official approval from the Ethics Committee of the Shanghai Institutes for Biological Sciences, Shanghai, China. All participants had provided written consent. All samples were genotyped using the Illumina HumanOmniZhongHua-8 chips, which interrogates 894,517 SNPs. Individuals with more than 5% missing data, related individuals, and the ones that failed the X-chromosome sex concordance check or had ethnic information incompatible with their genetic information were excluded. SNPs with more than 2% missing data, with a minor allele frequency smaller than 1%, and the ones that failed the Hardy–Weinberg deviation test (P < 1e-5) were also excluded. After applying these filters, we obtained a dataset of 709 samples with 810,648 SNPs for the Uyghurs. The chip genotype data were firstly phased using SHAPEIT (*O'Connell et al., 2014*). IMPUTE2 (*Howie et al., 2009*) was then used to impute genotypes at un-genotyped SNPs using the 1000 Genomes Phase 3 data as reference. Finally, for the Uyghur sample, a total of 6,414,304 imputed SNPs passed quality control and were combined with 810,648 genotyped SNPs for further analyses.

CANDELA study. CANDELA samples (*Ruiz-Linares et al., 2014*) consist of 6,630 volunteers recruited in five Latin American countries (Brazil, Colombia, Chile, México and Perú). Ethics approval was obtained from: The Universidad Nacional Autónoma de México (México), the Universidad de Antioquia (Colombia), the Universidad Peruana Cayetano Heredia (Perú), the Universidad de Tarapacá (Chile), the Universidade Federal do Rio Grande do Sul (Brazil) and the University College London (UK). All participants provided written informed consent. Individuals with dysmorphologies, a history of facial surgery or trauma, or with BMI over 33 were excluded (due to the effect of obesity on facial features). These individuals were genotyped on Illumina's Omni Express BeadChip. After applying quality control filters 669,462 SNPs and 6,357 individuals were retained for further analyses (2,922 males, 3,435 females). Average admixture proportions for this sample were estimated as: 48% European, 46% Native American and 6% African, but with substantial inter-individual variation. After all genomic and phenotypic quality controls this study included 5,958 individuals. The genetic PCs were obtained from the LD-pruned dataset of 93,328 SNPs using PLINK 1.9. These PCs were selected by inspecting the proportion of variance explained and checking scatter and scree plots. The final imputed dataset used in the GWAS analyses included genotypes for 9,143,600 SNPs using the 1000

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Genomes Phase I reference panel. Association analysis on the imputed dataset were performed using the best-guess imputed genotypes in PLINK 1.9 using linear regression with an additive genetic model incorporating age, sex and 5 genetic PCs as covariates.

Queensland Institute of Medical Research (QIMR) study. After phenotype and genotype quality control, data were available for 1,101 participants who were mainly twins aged 16 years. All participants, and where appropriate their parent or guardian, gave informed consent, and all studies were approved by the QIMR Berghofer Human Research Ethics Committee. Participants were genotyped on the Illumina Human610-Quad and Core+Exome SNP chips. Genotype data were screened for genotyping quality (GenCall < 0.7), SNP and individual call rates (< 0.95), HWE failure (P < 1e-6) and MAF (< 0.01). Data were checked for pedigree, sex and Mendelian errors and for non-European ancestry. Imputation was performed on the Michigan Imputation Server using the SHAPEIT/minimac Pipeline according to the standard protocols.

Facial shape phenotyping

We focused on 13 facial landmarks available in all cohorts with *3dMDface* digital photographs. These included right (ExR) and left (ExL) Exocanthion: points of bilateral outer canthus; right (EnR) and left (EnL) Endocanthion: points of bilateral inner canthus; nasion (Nsn): the point where the bridge of the nose meets the forehead; pronasale (Prn): the point of the nose tip; subnasale (Sbn): the point where the base of the nasal septum meets the philtrum; right (AlR) and left (AlL) alare: points of bilateral nose wing; Labliale superius (Ls): the point of labial superius; Labiale inferius (Li): the point of labial inferius; right (ChR) and left (ChL) Cheilion: points of bilateral angulus oris. Slightly different image processing algorithms were applied to 3dMD images to capture facial landmarks in the participating cohorts. In **RS** and **TwinsUK**, the raw 3D facial images were acquired using a 3D photographic scanning system manufactured by 3dMD (http://www.3dmd.com/). Participants were asked to keep their mouths closed and adopt a neutral expression during the acquisition of the 3D scans. Ensemble methods for automated 3D face landmarking were then used to derived 21 landmarks (*de Jong et al., 2018; de Jong et al., 2016*). In short, 3D surface model of participants' face obtained with commercial photogrammetry systems for faces called 3dMDface were input to the algorithm. Then a map projection of these face surfaces transformed the 3D data set into 3D relief maps and 2D images

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

were generated from these 3D relief maps. These images were subjected to a 2D landmarking method like Elastic Bunch Graph Matching. Finally, registered 2D landmarks are mapped back into 3D, inverting the projection. 13 shared facial landmarks coordinate data were processed using GPA to remove variations due to scaling, shifting and rotation. Euclidean distances were calculated between all landmarks. Outliers (>3sd) were removed before the subsequent analysis. At 15 years of age, the **ALSPAC** children were recalled (5,253 attended) and 3D facial scans where taken using two high-resolution Konica Minolta VI-900 laser scanners (lenses with a focal length 14.5 mm). The set of left and right facial scans of each individual were processed, registered, and merged using a locally developed algorithm implemented as a macro in Rapidform 2006® (*Kau, Zhurov, Scheer, Bouwman, & Richmond, 2004; Paternoster et al., 2012; Toma, Zhurov, Playle, Ong, & Richmond, 2009*). 22 landmarks were recorded manually. GPA was used to remove affine variations due to shifting, rotation and scaling. 4,747 individuals had usable images (506 individuals did not complete the assessment or facial scans were removed due to poor quality). The coordinates of 22 facial landmarks were derived from the scans, of which 13 were shared across all other studies. 3D Euclidean distances were calculated between the facial landmarks. 3D facial models from **PITT** were obtained with 3dMD digital stereo-photogrammetry systems. Facial landmarks were collected on the facial surfaces by trained raters and assessed for common placement errors during subsequent quality control. A set of 78 linear distance variables was then calculated from a subset of 13 of these landmarks, chosen because they were shared among all of the discovery cohorts in this study. These distances were screened for outliers. In **UYG**, the 3DMDface system (www.3dmd.com/3dMDface) was used to collect high-resolution 3D facial images from participants. And then a dense registration is applied to align all the 3D images (*Guo, Mei, & Tang, 2013*). Briefly, 15 salient facial landmarks were first automatically annotated based on the PCA projection of texture and shape information. Afterwards, a facial image of high quality and smooth shape surface was chosen as the reference, and its mesh was re-sampled to achieve an even density of one vertex in each 1mm×1mm grid. The reference face was then warped to register every sample face by matching all the 15 landmarks, via a non-rigid thin-plate spline (TPS) transformation. The mesh points of the reference face were then projected to the sample surface to find their one-to-one correspondents. The resulting points of projection were used to define the mesh of the sample facial surface. After the alignment, each sample face was represented by a set of 32,251 3D points and 15 target points were ready to be analyzed. A set of 78 linear

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

179 distance variables was calculated from a subset of 13 of these 15 landmarks. In **CANDELA**, ordinal
180 phenotyping was described in prior GWAS (*Adhikari et al., 2016*). In short, right side and frontal
181 photographs of participants were used to score 14 facial traits, Intra-class correlation coefficients
182 (ICCs) were calculated to indicate a moderate-to-high intra-rater reliability of the trait scores. All
183 photographs were scored by the same rater. In **QIMR**, two independent researchers manually
184 evaluated all 2D portrait photos and located the coordinates for 36 facial landmarks. Satisfactory
185 inter-rater concordance was obtained (Pearson's correlation coefficient, mean r = 0.76). The average
186 landmark coordinates between the two researchers were used for Euclidian distance calculation.
187

188 # Phenotype characteristics

189 Phenotype characteristics were explored for gender and aging effect, phenotypic and genetic
190 correlations, clustering, and twins-heritability. Gender and aging effects were estimated using beta, p,
191 and R^2 values from linear models. Phenotypic correlations were estimated using Pearson's correlation
192 coefficients. Genetic correlations were estimated using genome-wide SNPs and the multivariate
193 linear mixed model algorithm implemented in the GEMMA software package (*Zhou & Stephens,*
194 *2014*). Manual clustering was according to the involvement of intra- and inter- organ landmarks.
195 Unsupervised hierarchical clustering was conducted using the 1-abs(correlation) as the dissimilarity
196 matrix and each iteration was updated using the Lance - Williams formula (*Ward, 1963*). we
197 calculated twins heritability for TwinsUK and QIMR twins using the phenotypic correlation derived
198 from monozygotic twins (MZ) and dizygotic twins (DZ), i.e., $h^2 = 2[cor(MZ) - cor(DZ)]$. SNP-based
199 heritability was estimated for twins cohorts using the restricted estimated maximum likelihood
200 method implemented in the GCTA software package (*Yang, Lee, Goddard, & Visscher, 2011*).
201

202 # GWAS, meta-analysis, and replication studies

203 GWASs for face phenotypes were independently carried out in RS, TwinsUK, ALSPAC, PITT, CANDELA,
204 UYG, and QIMR based on linear or logistic models assuming an additive allele effect adjusted for
205 potential confounders in each cohorts, such as family relationships, sex, age, BMI and genetic PCs
206 using software packages including PLINK (*Purcell et al., 2007*), GEMMA (*Zhou & Stephens, 2012*), and
207 Rare Metal Worker (*Feng, Liu, Zhan, Wing, & Abecasis, 2014*). For **RS**, Association tests were run

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

between the 78 3D facial distance measurements and 6,886,439 genotyped and imputed SNPs with MAF > 0.01, imputation R2>0.8, SNP call rate > 0.97 and HWE > 0.0001. GWAS was performed using linear regression under the additive genetic model while adjusting for sex, age, and four PCs of ancestry as implemented in GCTA (*Yang et al., 2011*). For **TwinsUK**, Quality control was performed on SNPs (MAF > 0.01, imputation R2>0.8, SNP call rate > 0.99). One of each pair monozygotic twins (MZ) were except from association tests to reduce relatedness. The final sample included 1,020 individuals and 4,699,859 autosomal genetic variants. GWAS was performed between the 78 3D facial distance measurements and genotyped SNPs using linear regression under the additive genetic model while adjusting for age, and four PCs of ancestry as implemented in GCTA. For **ALSPAC**, 3,941 ALSPAC study participants had both genotype and 3D facial phenotype data. GCTA was used to remove individuals with excessive relatedness (GRM > 0.05) and quality control was performed on SNPs (MAF > 0.01, imputation R2>0.8, SNP call rate > 0.99). The final sample included 3,707 individuals and 4,627,882 autosomal genetic variants. PLINK 1.9 (*Purcell et al., 2007*) was used for genome-wide analysis of the 78 distances on both scaled and unscaled measures using a linear regression model. Sex, age, height and the first 4 principal components were included as covariates in the model. For **PITT**, Association tests were run between the 78 3D facial distance measurements and 9,478,231 genotyped and imputed SNPs with MAF > 1% and HWE > 0.0001. GWAS was performed using linear regression under the additive genetic model while adjusting for sex, age, age^2, height, weight and four PCs of ancestry as implemented in PLINK 1.9. For the analysis of the X-chromosome, genotypes were coded 0, 1, and 2 as per the additive genetic model for females, and coded 0, 2 for males in order to maintain the same scale between sexes. For **CANDELA**, PLINK 1.9 was used to perform the primary genome-wide association tests for each phenotype using multiple linear regression with an additive genetic model incorporating age, sex, BMI and 5 genetic PCs as covariates. Association analyses were performed on the imputed dataset with two approaches: using the best-guess imputed genotypes in PLINK, and using the IMPUTE2 genotype probabilities in SNPTEST v2.5. Both were consistent with each other and with the results from the chip genotype data. For **UYG**, GWAS were carried out using PLINK 1.9 and a linear regression model with additive SNP effects was used. Details was described in prior GWAS (*Qiao et al., 2018*). For **QIMR**, The GWAS was conducted using Rare Metal Worker with linear regression adjusted for sex, age, BMI and five PCs of ancestry as implemented in GCTA.

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Inverse variance fixed-effect meta-analysis were carried out using PLINK to combine GWAS results for samples of European origin and having 3dMDface data, i.e., RS, TwinsUK, ALSPAC, and PITT, which included 7,029,494 autosomal SNPs overlapping between the 4 discovery cohorts with physical positions aligned according to NCBI human genome sequence build GRCh37.p13. Meta-analysis results were visualized using superposed Manhattan plots and Q-Q plots. All genomic inflation factors were close to 1 and not further considered. We considered the threshold of 5×10^{-8} as the threshold for replication analysis in CANDELA, UYG, and QIMR. Since the 3dMDface data were available in UYG, exact replications were conducted, i.e., the genetic association was tested for the corresponding phenotypes in the meta-analysis. Since only 2D photos were available in CANDELA and QIMR, trait-wise replications were conducted to test H0: no association for all face phenotypes vs. H1: associated with at least one face phenotypes. We used a combined test of dependent tests to combine genetic association tests involving highly correlated face phenotypes (*Kost & McDermott, 2002*). In short, the test statistic of the combined test follows a scaled chi-squared distribution when tests are dependent. The mean of the $\boldsymbol{\chi^2}$ values is:

$\boldsymbol{E(\chi^2) = 2k}$, and the variance can be estimated as

$\boldsymbol{var(\chi^2) = 4k + 2\sum_{1\le i<j\le k} cov(-2logP_i, -2logP_j)}$, where

$\boldsymbol{k}$ is the number of phenotypes, $\boldsymbol{P_i}$ is the $\boldsymbol{p}$ value of the $\boldsymbol{i}$th test. The covariance term $\boldsymbol{cov(-2logP_i, -2logP_j)}$ can be approximated by the following formula using the phenotype correlation matrix:

$$\boldsymbol{cov(-2logP_i, -2logP_j) \approx 3.263r_{ij} + 0.71r_{ij}^2 + 0.027r_{ij}^3}$$

This approximation has been shown nearly identical to the exact value with difference less than 0.0002 for -0.98 < r_{ij} < 0.98 (*Kost & McDermott, 2002*). Regional linkage disequilibrium r^2 values were calculated in PLINK and visualized using R scripting. Regional Manhattan plots were produced using LocusZoom (*Pruim et al., 2010*).

Genetic diversity and positive selection signal analysis

To test for signals of allele frequency diversity and positive selection in the face associated genomic regions, we derived two commonly used statistics in population genetics, i.e., the fixation index (FST)

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and the integrated haplotype score (iHS) for all SNPs in the associated regions. The FST is a parameter in population genetics for quantifying the differentiation due to genetic structure between a pair of groups by measuring the proportion of genetic diversity due to allele frequency differences among populations potentially as a result of divergent natural selection (*Holsinger & Weir, 2009*). We performed a genome-wide FST analysis for three pairs of populations (AFR-EAS, EAS-EUR, and EUR-AFR) in the samples from the 1000 Genome Project using the PLINK software. The empirical significance cutoff for each of the three population pairs was set to the value corresponding to the top 1% FST values in the genome. The iHS is the integrate of the extended haplotype homozygosity (EHH) statistic to highlight recent positive selected variants that have not yet reached fixation, and large values suggest the adaption to the selective pressures in the most recent stage of evolution (*Sabeti et al., 2002; Voight, Kudaravalli, Wen, & Pritchard, 2006*). The EHH statistic was calculated for all SNPs in AFR, EAS and EUR samples. Then the iHS statistic, integral of the decay of EHH away from each SNP until EHH reaches 0.05, was calculated for all SNPs, in which an allele frequency bin of 0.05 was used to standardize iHS scores against other SNPs of the same frequency class. Finally, we calculated empirical significance threshold over the genome assuming a Gaussian distribution of iHS scores under the neutral model. The cutoff was based on the top 1% absolute iHS scores calculated in three samples respectively as the absolute score indicate the selection intensity.

Multivariable and polygenic score (PS) analysis

Multivariable and polygenic score analyses were conducted in the RS cohort. For all SNPs showing genome-wide significant association with face phenotypes ($p<5x10^{-8}$), we performed a tagging SNP analysis excluding SNPs in strong LD (pair-wise $r^2>0.5$), which resulted in 33 SNPs in 24 genomic regions for the multivariable and PS analyses. A forward step-wise regression was conducted to access the independent effects of the selected SNPs on sex- and age- adjusted face phenotypes, where the R^2 contribution of each SNP was estimated by comparing the current model fitting R^2 with the R^2 in a previous iteration. Next, we calculated PS for AFR, EUR, EAS individuals of the 1000 Genomes Project using the selected SNPs and the effect sizes estimated from the multivariable analysis, i.e., $\boldsymbol{PS} = \sum_{i=1}^{n} \boldsymbol{\beta_i A_i}$. The mean values and distributions of the standardized PS values for all face phenotypes in each continental group was visualized using a face map and multiple boxplots, in

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

294 which the differences between groups were highlighted according to the p-values from the ANOVA
295 test, which tests H0: PS values are the same in all continental groups and H1: PS values are different
296 in at least one of the groups.
297

298 Gene ontology analysis, expression analysis and SNP functional
299 annotation

300 To further study the involved functions and regulation relationships of the genome-wide significantly
301 associated SNP markers and candidate genes nearby, the genes located within 100kb up- and
302 down-stream of the SNPs were selected to perform biological process enrichment analysis based on
303 the Gene Ontology (GO) database. The enrichment analysis was performed using the R package
304 'clusterProfiler' (*G. Yu, Wang, Han, & He, 2012*). RNA-seq data from the study of Prescott et al.
305 (*Prescott et al., 2015*), GTEx (*G. T. Consortium, 2013*) and ENCODE (*E. P. Consortium, 2012*) were used
306 to compare the differences in gene expression levels between our face-associated genes and all
307 genes over the genome in CNCCs. We examined entries in GWAS catalog (*Welter et al., 2014*) to
308 explore the potential pleotropic effects in our face associated regions. The GTEx (*G. T. Consortium,*
309 *2013*) dataset was used to annotate our face associated SNPs for significant single-tissue *cis*-eQTL
310 effects in all 48 available tissues. We used the CNCC data from Prescott et al. and UCSC (*Aken et al.,*
311 *2016*) to perform a variety of functional annotations for the face associated SNPs in non-coding
312 regions using the WashU Epigenome Browser, including histone modifications (H3K27ac, H3K4me1
313 and H3K4me3 tracks), chromatin modifier (p300 tracks), transcription factors (TFAP2A tracks), chromatin
314 accessibility (ATACseq tracks) and conserved regions (PhyloP (*Siepel et al., 2005*) tracks) from UCSC. To
315 investigate the pattern of regulation within each association region in details, we calculated the log
316 scaled quantile rank and variation of all LD SNP ($r^2 > 0.25$) in 3 replicated measurements from 6 types
317 of Chip-seq data, and for all significant base-pair in vicinity (within 20kb) of all top-associated SNPs.
318 All analyses were conducted using R Environment for Statistical Computing (version 3.3.1) unless
319 otherwise specified.
320

321 Cell culture

322 Neural progenitors were derived from human induced pluripotent stem cells as described previously

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(male, age 57 years, derived from primary skin fibroblasts) (*Gunhanlar et al., 2018*). Written informed consent was obtained in accordance with the Medical Ethical Committee of the Erasmus University Medical Center. Neural progenitors contain a subpopulation of neural crest progenitor cells that were enriched by fluorescence-activated cell sorting for CD271 (BD Bioscience 560927, BD FACSAria III) (*Dupin & Coelho-Aguiar, 2013; G. Lee et al., 2007*). CD271 (p75) is a surface antigen marking neural crest progenitor cells which give rise to craniofacial tissues as well as melanocyte primary cells (*Morrison, White, Zock, & Anderson, 1999*). CD271+ cells were grown in DMEM:F12, 1% N2 supplement, 2% B27 supplement, 1% P/S, 20 ng/ml FGF. G361 cells were cultured in DMEM/10% FCS at 37°C/5% CO2.

In vitro functional experiments: Luciferase reporter assay

We explored the effect of the 5 selected SNPs on gene expression in CD271+ and G361 melanoma cells by a luciferase reporter assay. Custom 500bp gBlocks Gene fragments (IDT) containing putative enhancers surrounding each of the 5 selected SNPs were cloned into a modified pGL3 firefly luciferase reporter vector (SV40 promoter and 3´UTR replaced by HSP promoter and 3´UTR). For each SNP, constructs containing both ancestral and derived alleles were tested. Equimolar amount of constructs (measured by qPCR detecting the firefly luciferase gene) were co-transfected with renilla luciferase control vector into CD271+ and G361 cells using Lipofectamine LTX (Invitrogen). Luciferase activity was measured 48h after transfection using the Dual-Glo Luciferase Assay System (Promega). Data are presented as relative luciferase activity (firefly/renilla activity ratio) fold change compared to the construct containing the less active allele and represent at least three independent replicates. Student´s t-test was performed to test differences in gene expression.

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Adhikari, K., Fuentes-Guajardo, M., Quinto-Sanchez, M., Mendoza-Revilla, J., Camilo Chacon-Duque, J., Acuna-Alonzo, V., . . . Ruiz-Linares, A. (2016). A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. *Nat Commun, 7*, 11616. doi: 10.1038/ncomms11616

Adhikari, K., Reales, G., Smith, A. J., Konka, E., Palmen, J., Quinto-Sanchez, M., . . . Ruiz-Linares, A. (2015). A genome-wide association study identifies multiple loci for variation in human ear morphology. *Nat Commun, 6*, 7500. doi: 10.1038/ncomms8500

Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., . . . Searle, S. M. (2016). The Ensembl gene annotation system. *Database (Oxford), 2016*. doi: 10.1093/database/baw093

Bamshad, M., Lin, R. C., Law, D. J., Watkins, W. C., Krakowiak, P. A., Moore, M. E., . . . Jorde, L. B. (1997). Mutations in human TBX3 alter limb, apocrine and genital development in ulnar-mammary syndrome. *Nat Genet, 16*(3), 311-315. doi: 10.1038/ng0797-311

Beaty, T. H., Murray, J. C., Marazita, M. L., Munger, R. G., Ruczinski, I., Hetmanski, J. B., . . . Scott, A. F. (2010). A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. *Nat Genet, 42*(6), 525-529. doi: 10.1038/ng.580

Beaty, T. H., Ruczinski, I., Murray, J. C., Marazita, M. L., Munger, R. G., Hetmanski, J. B., . . . Scott, A. F. (2011). Evidence for gene-environment interaction in a genome wide study of nonsyndromic cleft palate. *Genet Epidemiol, 35*(6), 469-478. doi: 10.1002/gepi.20595

Birnbaum, S., Ludwig, K. U., Reutter, H., Herms, S., Steffens, M., Rubini, M., . . . Mangold, E. (2009). Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nat Genet, 41*(4), 473-477. doi: 10.1038/ng.333

Bjork, B. C., Turbe-Doan, A., Prysak, M., Herron, B. J., & Beier, D. R. (2010). Prdm16 is required for normal palatogenesis in mice. *Hum Mol Genet, 19*(5), 774-789. doi: 10.1093/hmg/ddp543

Boyd, A., Golding, J., Macleod, J., Lawlor, D. A., Fraser, A., Henderson, J., . . . Davey Smith, G. (2013). Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol, 42*(1), 111-127. doi: 10.1093/ije/dys064

Bronner, M. E., & LeDouarin, N. M. (2012). Development and evolution of the neural crest: an overview. *Dev Biol, 366*(1), 2-9. doi: 10.1016/j.ydbio.2011.12.042

Bruel, A. L., Franco, B., Duffourd, Y., Thevenon, J., Jego, L., Lopez, E., . . . Thauvin-Robinet, C. (2017). Fifteen years of research on oral-facial-digital syndromes: from 1 to 16 causal genes. *J Med Genet, 54*(6), 371-380. doi: 10.1136/jmedgenet-2016-104436

Buchan, J. G., Gray, R. S., Gansner, J. M., Alvarado, D. M., Burgert, L., Gitlin, J. D., . . . Goldsmith, M. I. (2014). Kinesin family member 6 (kif6) is necessary for spine development in zebrafish. *Dev Dyn, 243*(12), 1646-1657. doi: 10.1002/dvdy.24208

Carroll, S. B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell, 134*(1), 25-36. doi: 10.1016/j.cell.2008.06.030

Cha, S., Lim, J. E., Park, A. Y., Do, J. H., Lee, S. W., Shin, C., . . . Oh, B. (2018). Identification of five novel genetic loci related to facial morphology by genome-wide association studies. *BMC Genomics, 19*(1), 481. doi: 10.1186/s12864-018-4865-9

Chadwick, L. H. (2012). The NIH Roadmap Epigenomics Program data resource. *Epigenomics, 4*(3), 317-324. doi: 10.2217/epi.12.18

Claes, P., Roosenboom, J., White, J. D., Swigut, T., Sero, D., Li, J., . . . Weinberg, S. M. (2018).

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Genome-wide mapping of global-to-local genetic effects on human facial shape. *Nat Genet, 50*, 414-423. doi: 10.1038/s41588-018-0057-4

Cole, J. B., Manyama, M., Kimwaga, E., Mathayo, J., Larson, J. R., Liberton, D. K., . . . Spritz, R. A. (2016). Genomewide Association Study of African Children Identifies Association of SCHIP1 and PDE8A with Facial Size and Shape. *PLoS Genet, 12*(8), e1006174. doi: 10.1371/journal.pgen.1006174

Consortium, Encode Project. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489*(7414), 57-74. doi: 10.1038/nature11247

Consortium, G. TEx. (2013). The Genotype-Tissue Expression (GTEx) project. *Nat Genet, 45*(6), 580-585. doi: 10.1038/ng.2653

Cordero, D. R., Brugmann, S., Chu, Y., Bajpai, R., Jame, M., & Helms, J. A. (2011). Cranial neural crest cells on the move: their roles in craniofacial development. *Am J Med Genet A, 155A*(2), 270-279. doi: 10.1002/ajmg.a.33702

de Jong, M. A., Hysi, P., Spector, T., Niessen, W., Koudstaal, M. J., Wolvius, E. B., . . . Bohringer, S. (2018). Ensemble landmarking of 3D facial surface scans. *Sci Rep, 8*(1), 12. doi: 10.1038/s41598-017-18294-x

de Jong, M. A., Wollstein, A., Ruff, C., Dunaway, D., Hysi, P., Spector, T., . . . Bohringer, S. (2016). An Automatic 3D Facial Landmarking Algorithm Using 2D Gabor Wavelets. *IEEE Trans Image Process, 25*(2), 580-588. doi: 10.1109/TIP.2015.2496183

Delaneau, O., Marchini, J., & Zagury, J. F. (2011). A linear complexity phasing method for thousands of genomes. *Nat Methods, 9*(2), 179-181. doi: 10.1038/nmeth.1785

Delous, M., Baala, L., Salomon, R., Laclef, C., Vierkotten, J., Tory, K., . . . Saunier, S. (2007). The ciliary gene RPGRIP1L is mutated in cerebello-oculo-renal syndrome (Joubert syndrome type B) and Meckel syndrome. *Nat Genet, 39*(7), 875-881. doi: 10.1038/ng2039

Djordjevic, J., Zhurov, A. I., Richmond, S., & Visigen, Consortium. (2016). Genetic and Environmental Contributions to Facial Morphological Variation: A 3D Population-Based Twin Study. *PLoS One, 11*(9), e0162250. doi: 10.1371/journal.pone.0162250

Dupin, E., & Coelho-Aguiar, J. M. (2013). Isolation and differentiation properties of neural crest stem cells. *Cytometry A, 83*(1), 38-47. doi: 10.1002/cyto.a.22098

Feng, S., Liu, D., Zhan, X., Wing, M. K., & Abecasis, G. R. (2014). RAREMETAL: fast and powerful meta-analysis for rare variants. *Bioinformatics, 30*(19), 2828-2829. doi: 10.1093/bioinformatics/btu367

Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., . . . Lawlor, D. A. (2013). Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol, 42*(1), 97-110. doi: 10.1093/ije/dys066

Genomes Project, Consortium, Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., . . . McVean, G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature, 491*(7422), 56-65. doi: 10.1038/nature11632

Genomes Project, Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., . . . Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature, 526*(7571), 68-74. doi: 10.1038/nature15393

Golding, J., Pembrey, M., Jones, R., & Team, Alspac Study. (2001). ALSPAC--the Avon Longitudinal Study of Parents and Children. I. Study methodology. *Paediatr Perinat Epidemiol, 15*(1), 74-87.

Grant, S. F., Wang, K., Zhang, H., Glaberson, W., Annaiah, K., Kim, C. E., . . . Hakonarson, H. (2009). A

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. *J Pediatr, 155*(6), 909-913. doi: 10.1016/j.jpeds.2009.06.020

Gunhanlar, N., Shpak, G., van der Kroeg, M., Gouty-Colomer, L. A., Munshi, S. T., Lendemeijer, B., . . . Kushner, S. A. (2018). A simplified protocol for differentiation of electrophysiologically mature neuronal networks from human induced pluripotent stem cells. *Mol Psychiatry, 23*(5), 1336-1344. doi: 10.1038/mp.2017.56

Guo, J., Mei, X., & Tang, K. (2013). Automatic landmark annotation and dense correspondence registration for 3D human facial images. *BMC Bioinformatics, 14*, 232. doi: 10.1186/1471-2105-14-232

Hofman, A., Darwish Murad, S., van Duijn, C. M., Franco, O. H., Goedegebure, A., Ikram, M. A., . . . Vernooij, M. W. (2013). The Rotterdam Study: 2014 objectives and design update. *Eur J Epidemiol, 28*(11), 889-926. doi: 10.1007/s10654-013-9866-z

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nat Rev Genet, 10*(9), 639-650. doi: 10.1038/nrg2611

Howie, B. N., Donnelly, P., & Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet, 5*(6), e1000529. doi: 10.1371/journal.pgen.1000529

Hui, C. C., & Joyner, A. L. (1993). A mouse model of greig cephalopolysyndactyly syndrome: the extra-toesJ mutation contains an intragenic deletion of the Gli3 gene. *Nat Genet, 3*(3), 241-246. doi: 10.1038/ng0393-241

Joss, S., Kini, U., Fisher, R., Mundlos, S., Prescott, K., Newbury-Ecob, R., & Tolmie, J. (2011). The face of Ulnar Mammary syndrome? *Eur J Med Genet, 54*(3), 301-305. doi: 10.1016/j.ejmg.2010.12.010

Kamberov, Y. G., Wang, S., Tan, J., Gerbault, P., Wark, A., Tan, L., . . . Sabeti, P. C. (2013). Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell, 152*(4), 691-702. doi: 10.1016/j.cell.2013.01.016

Kau, C. H., Zhurov, A., Scheer, R., Bouwman, S., & Richmond, S. (2004). The feasibility of measuring three-dimensional facial morphology in children. *Orthod Craniofac Res, 7*(4), 198-204.

Kayser, M., Liu, F., Janssens, A. C., Rivadeneira, F., Lao, O., van Duijn, K., . . . van Duijn, C. M. (2008). Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. *Am J Hum Genet, 82*(2), 411-423. doi: 10.1016/j.ajhg.2007.10.003

Kimura, R., Yamaguchi, T., Takeda, M., Kondo, O., Toma, T., Haneji, K., . . . Oota, H. (2009). A common variation in EDAR is a genetic determinant of shovel-shaped incisors. *Am J Hum Genet, 85*(4), 528-535. doi: 10.1016/j.ajhg.2009.09.006

Kost, James T., & McDermott, Michael P. (2002). Combining dependent P-values. *Statistics & Probability Letters, 60*(2), 183-190.

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., . . . Hirschhorn, J. N. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature, 467*(7317), 832-838. doi: 10.1038/nature09410

Lee, G., Kim, H., Elkabetz, Y., Al Shamy, G., Panagiotakos, G., Barberi, T., . . . Studer, L. (2007). Isolation and directed differentiation of neural crest stem cells derived from human embryonic stem cells. *Nat Biotechnol, 25*(12), 1468-1475. doi: 10.1038/nbt1365

Lee, M. K., Shaffer, J. R., Leslie, E. J., Orlova, E., Carlson, J. C., Feingold, E., . . . Weinberg, S. M. (2017). Genome-wide association study of facial morphology reveals novel associations with FREM1 and PARK2. *PLoS One, 12*(4), e0176566. doi: 10.1371/journal.pone.0176566

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Leslie, E. J., Carlson, J. C., Shaffer, J. R., Butali, A., Buxo, C. J., Castilla, E. E., . . . Marazita, M. L. (2017). Genome-wide meta-analyses of nonsyndromic orofacial clefts identify novel associations between FOXE1 and all orofacial clefts, and TP63 and cleft lip with or without cleft palate. *Hum Genet, 136*(3), 275-286. doi: 10.1007/s00439-016-1754-7

Li, J., & Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb), 95*(3), 221-227. doi: 10.1038/sj.hdy.6800717

Liu, F., van der Lijn, F., Schurmann, C., Zhu, G., Chakravarty, M. M., Hysi, P. G., . . . Kayser, M. (2012). A genome-wide association study identifies five loci influencing facial morphology in Europeans. *PLoS Genet, 8*(9), e1002932. doi: 10.1371/journal.pgen.1002932

Ludwig, K. U., Bohmer, A. C., Bowes, J., Nikolic, M., Ishorst, N., Wyatt, N., . . . Mangold, E. (2017). Imputation of orofacial clefting data identifies novel risk loci and sheds light on the genetic background of cleft lip +/- cleft palate and cleft palate only. *Hum Mol Genet, 26*(4), 829-842. doi: 10.1093/hmg/ddx012

Ludwig, K. U., Mangold, E., Herms, S., Nowak, S., Reutter, H., Paul, A., . . . Nothen, M. M. (2012). Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. *Nat Genet, 44*(9), 968-971. doi: 10.1038/ng.2360

Mangold, E., Ludwig, K. U., Birnbaum, S., Baluardo, C., Ferrian, M., Herms, S., . . . Nothen, M. M. (2010). Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. *Nat Genet, 42*(1), 24-26. doi: 10.1038/ng.506

Morrison, S. J., White, P. M., Zock, C., & Anderson, D. J. (1999). Prospective identification, isolation by flow cytometry, and in vivo self-renewal of multipotent mammalian neural crest stem cells. *Cell, 96*(5), 737-749.

Noback, M. L., Harvati, K., & Spoor, F. (2011). Climate-related variation of the human nasal cavity. *Am J Phys Anthropol, 145*(4), 599-614. doi: 10.1002/ajpa.21523

O'Connell, J., Gurdasani, D., Delaneau, O., Pirastu, N., Ulivi, S., Cocca, M., . . . Marchini, J. (2014). A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet, 10*(4), e1004234. doi: 10.1371/journal.pgen.1004234

Paternoster, L., Zhurov, A. I., Toma, A. M., Kemp, J. P., St Pourcain, B., Timpson, N. J., . . . Evans, D. M. (2012). Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasion position. *Am J Hum Genet, 90*(3), 478-485. doi: 10.1016/j.ajhg.2011.12.021

Pickrell, J. K., Berisa, T., Liu, J. Z., Segurel, L., Tung, J. Y., & Hinds, D. A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet, 48*(7), 709-717. doi: 10.1038/ng.3570

Prescott, S. L., Srinivasan, R., Marchetto, M. C., Grishina, I., Narvaiza, I., Selleri, L., . . . Wysocka, J. (2015). Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell, 163*(1), 68-83. doi: 10.1016/j.cell.2015.08.036

Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., . . . Willer, C. J. (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics, 26*(18), 2336-2337. doi: 10.1093/bioinformatics/btq419

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet, 81*(3), 559-575. doi: 10.1086/519795

Qiao, L., Yang, Y., Fu, P., Hu, S., Zhou, H., Peng, S., . . . Tang, K. (2018). Genome-wide variants of

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Eurasian facial shape differentiation and a prospective model of DNA based face prediction. *J Genet Genomics, 45*(8), 419-432. doi: 10.1016/j.jgg.2018.07.009

Ruiz-Linares, A., Adhikari, K., Acuna-Alonzo, V., Quinto-Sanchez, M., Jaramillo, C., Arias, W., . . . Gonzalez-Jose, R. (2014). Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genet, 10*(9), e1004572. doi: 10.1371/journal.pgen.1004572

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., . . . Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature, 419*(6909), 832-837. doi: 10.1038/nature01140

Sachse, G., Church, C., Stewart, M., Cater, H., Teboul, L., Cox, R. D., & Ashcroft, F. M. (2018). FTO demethylase activity is essential for normal bone growth and bone mineralization in mice. *Biochim Biophys Acta, 1864*(3), 843-850. doi: 10.1016/j.bbadis.2017.11.027

Satoh, W., Matsuyama, M., Takemura, H., Aizawa, S., & Shimono, A. (2008). Sfrp1, Sfrp2, and Sfrp5 regulate the Wnt/beta-catenin and the planar cell polarity pathways during early trunk formation in mouse. *Genesis, 46*(2), 92-103. doi: 10.1002/dvg.20369

Schwabe, G. C., Trepczik, B., Suring, K., Brieske, N., Tucker, A. S., Sharpe, P. T., . . . Mundlos, S. (2004). Ror2 knockout mouse as a model for the developmental pathology of autosomal recessive Robinow syndrome. *Dev Dyn, 229*(2), 400-410. doi: 10.1002/dvdy.10466

Shaffer, J. R., Orlova, E., Lee, M. K., Leslie, E. J., Raffensperger, Z. D., Heike, C. L., . . . Weinberg, S. M. (2016). Genome-Wide Association Study Reveals Multiple Loci Influencing Normal Human Facial Morphology. *PLoS Genet, 12*(8), e1006149. doi: 10.1371/journal.pgen.1006149

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., . . . Haussler, D. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res, 15*(8), 1034-1050. doi: 10.1101/gr.3715005

Singh, M. K., Petry, M., Haenig, B., Lescher, B., Leitges, M., & Kispert, A. (2005). The T-box transcription factor Tbx15 is required for skeletal development. *Mech Dev, 122*(2), 131-144. doi: 10.1016/j.mod.2004.10.011

Small, K. S., Hedman, A. K., Grundberg, E., Nica, A. C., Thorleifsson, G., Kong, A., . . . Mu, Ther Consortium. (2011). Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat Genet, 43*(6), 561-564. doi: 10.1038/ng.833

Sun, Y., Huang, Y., Yin, A., Pan, Y., Wang, Y., Wang, C., . . . Yang, Y. (2015). Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. *Nat Commun, 6*, 6414. doi: 10.1038/ncomms7414

Tan, J., Yang, Y., Tang, K., Sabeti, P. C., Jin, L., & Wang, S. (2013). The adaptive variant EDARV370A is associated with straight hair in East Asians. *Hum Genet, 132*(10), 1187-1191. doi: 10.1007/s00439-013-1324-1

Thomason, H. A., Dixon, M. J., & Dixon, J. (2008). Facial clefting in Tp63 deficient mice results from altered Bmp4, Fgf8 and Shh signaling. *Dev Biol, 321*(1), 273-282. doi: 10.1016/j.ydbio.2008.06.030

Toma, A. M., Zhurov, A., Playle, R., Ong, E., & Richmond, S. (2009). Reproducibility of facial soft tissue landmarks on 3D laser-scanned facial images. *Orthod Craniofac Res, 12*(1), 33-42. doi: 10.1111/j.1601-6343.2008.01435.x

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biol, 4*(3), e72. doi: 10.1371/journal.pbio.0040072

bioRxiv preprint doi: https://doi.org/10.1101/693002; this version posted July 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Ward, Joe H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association, 58*(301), 236-244. doi: 10.2307/2282967

Weinberg, S. M., Raffensperger, Z. D., Kesterke, M. J., Heike, C. L., Cunningham, M. L., Hecht, J. T., . . . Marazita, M. L. (2016). The 3D Facial Norms Database: Part 1. A Web-Based Craniofacial Anthropometric and Image Repository for the Clinical and Research Community. *Cleft Palate Craniofac J, 53*(6), e185-e197. doi: 10.1597/15-199

Weiner, J. S. (1954). Nose shape and climate. *Am J Phys Anthropol, 12*(4), 615-618.

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., . . . Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res, 42*(Database issue), D1001-1006. doi: 10.1093/nar/gkt1229

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., . . . Frayling, T. M. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet, 46*(11), 1173-1186. doi: 10.1038/ng.3097

Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet, 8*(3), 206-216. doi: 10.1038/nrg2063

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet, 88*(1), 76-82. doi: 10.1016/j.ajhg.2010.11.011

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS, 16*(5), 284-287. doi: 10.1089/omi.2011.0118

Yu, Y., Zuo, X., He, M., Gao, J., Fu, Y., Qin, C., . . . Bian, Z. (2017). Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. *Nat Commun, 8*, 14364. doi: 10.1038/ncomms14364

Zalc, A., Rattenbach, R., Aurade, F., Cadot, B., & Relaix, F. (2015). Pax3 and Pax7 play essential safeguard functions against environmental stress-induced birth defects. *Dev Cell, 33*(1), 56-66. doi: 10.1016/j.devcel.2015.02.006

Zeng, H., Hoover, A. N., & Liu, A. (2010). PCP effector gene Inturned is an important regulator of cilia formation and embryonic development in mammals. *Dev Biol, 339*(2), 418-428. doi: 10.1016/j.ydbio.2010.01.003

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat Genet, 44*(7), 821-824. doi: 10.1038/ng.2310

Zhou, X., & Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat Methods, 11*(4), 407-409. doi: 10.1038/nmeth.2848