CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/127273/

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Xiong, Ziyi, Dankova, Gabriela, Howe, Laurence J, Lee, Myoung Keun, Hysi, Pirro G, deJong, Markus A, Zhu, Gu, Adhikari, Kaustubh, Li, Dan, Pan, Bo, Feingold, Eleanor, Marazita, Mary L, Shaffer, John R, McAloney, Kerrie, Xu, Shuhua, Jin, Li, Wang, Sijia, de Vrij, Femke M, Lendermeijer, Bas, Richmond, Stephen , Zhurov, Alexei , Lewis, Sarah, Sharp, Gemma, Paternoster, Lavinia, Thompson, Holly, Gonzalez-Jose, Rolando, Bortolini, Maria Catira, Canizales-Quiteros, Samuel, Gallo, Carla, Poletti, Giovanni, Bedoya, Gabriel, Rothhammer, Francisco, Utterlinden, Andre G, Ikram, M Arfan, Wolvius, Eppo B, Kushner, Steven A, Nijsten, Tamar, Palstra, Robert-Jan, Boehringer, Stefan, Medland, Sarah E, Tang, Kun, Ruiz-Linares, Andres, Martin, Nicholas G, Spector, Timothy D, Stergiakouli, Evie, Weinberg, Seth M, Liu, Fan and Kayser, Manfred 2019. Novel genetic loci affecting facial shape variation in humans. eLife 8 , e49898. 10.7554/eLife.49898

Publishers page: http://dx.doi.org/10.7554/eLife.49898

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

Novel genetic loci affecting facial shape variation in humans

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

1. Department of Genetic Identification, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands.
2. Department of Epidemiology, 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. Center for Craniofacial and Dental Genetics, Department of Oral Biology, 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 Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands.
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. CAS Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing, China.
12. Department of Auricular Reconstruction, Plastic Surgery Hospital, Beijing, China.
13. Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, U.S.A.
14. School of Life Science and Technology, ShanghaiTech University, Shanghai, China.
15. Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China.
16. State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China.
17. Department of Psychiatry, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands.
18. Applied Clinical Research and Public Health, University Dental School, Cardiff University, Heath Park, Cardiff, United Kingdom.
19. School of Oral and Dental Sciences, University of Bristol, Bristol, United Kingdom.
20. Instituto Patagonico de Ciencias Sociales y Humanas, CENPAT-CONICET, Puerto Madryn, Argentina.
21. Departamento de Genetica, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil.
22. Unidad de Genomica de Poblaciones Aplicada a la Salud, Facultad de Quımica, UNAM-Instituto Nacional de Medicina Genomica, Mexico City, Mexico.
23. Laboratorios de Investigacion y Desarrollo, Facultad de Ciencias y Filosofıa, Universidad Peruana Cayetano Heredia, Lima, Peru.
24. GENMOL (Genetica Molecular), Universidad de Antioquia, Medellın, Colombia.
25. Instituto de Alta Investigacion, Universidad de Tarapaca, Arica, Chile.
26. Department of Internal Medicine, 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:
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.); Email: f.liu@erasmusmc.nl (F.L.).
Or
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; Email: liufan@big.ac.cn (F.L.)

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. 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 study-wide suggestive association ($p<5x10^{-8}$), among which 17 were previously unreported. A follow-up multi-ethnic study in additional 7,917 individuals confirmed 10 loci including 6 unreported ones ($p_{adjusted}<2.1x10^{-3}$). A global map of derived 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.

Introduction

The human face represents a multi-dimensional set of correlated, mostly symmetric, complex phenotypes with high heritability. This is illustrated by a large degree of facial similarity between monozygotic twins and, albeit less so, between other relatives (*Djordjevic et al., 2016*), stable facial features within and differences between 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 several disciplines of fundamental and applied sciences, including human genetics, developmental biology, evolutionary biology, medical genetics, and forensics. However, current knowledge on the specific genes influencing human facial appearance 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

112 independent facial GWASs, thus currently representing the most established genetic findings,
113 include DNA variants in or close to 10 genes i.e., *CACNA2D3* (*Paternoster et al., 2012; Pickrell et al.,*
114 *2016*), *DCHS2* (*Adhikari et al., 2016; Claes et al., 2018*), *EPHB3* (*Claes et al., 2018; Pickrell et al., 2016*),
115 *HOXD* cluster (*Claes et al., 2018; Pickrell et al., 2016*), *PAX1* (*Adhikari et al., 2016; Shaffer et al., 2016*),
116 *PAX3* (*Adhikari et al., 2016; Claes et al., 2018; Liu et al., 2012; Paternoster et al., 2012; Pickrell et al.,*
117 *2016*), *PKDCC* (*Claes et al., 2018; Pickrell et al., 2016*), *SOX9* (*Cha et al., 2018; Claes et al., 2018;*
118 *Pickrell et al., 2016*), *SUPT3H* (*Adhikari et al., 2016; Claes et al., 2018; Pickrell et al., 2016*), and *TBX15*
119 (*Claes et al., 2018; Pickrell et al., 2016*), which were identified mostly in Europeans.
120 Indicated by the small effect size of the previously identified face-associated DNA variants,
121 together with the high heritability of facial phenotypes, the true number of genes that influence
122 polygenetic facial traits is likely to be much larger than currently known. Despite the previous work,
123 our understanding of the complex genetic architecture of human facial morphology thus remains
124 largely incomplete. This emphasizes on the need for well-powered GWASs that can be achieved
125 through large collaborative efforts, which has been recently demonstrated for appearance traits
126 involving pigmentation variation (*Hysi et al., 2018; Visconti et al., 2018*), but is currently missing for
127 the human face.
128 A most recent study (*Claes et al., 2018*) suggested that variants at genetic loci implicated in
129 human facial shape are involved in the epigenetic regulation of human neural crest cells, suggesting
130 that at least some of the functional variants may reside within regulatory elements such as
131 enhancers. Although this is in line with evidence for other appearance phenotypes such as human
132 pigmentation traits, where experimental studies have proven enhancer effects of pigmentation-
133 associated DNA variants (*Visser et al., 2012; Visser et al., 2014*, *2015*), experimental evidence for
134 understanding the functional basis of DNA variants for which facial phenotype associations have
135 been previously established is missing as of yet.
136 Led by the International Visible Trait Genetics (VisiGen) Consortium and together with its study
137 partners, the current study represents a collaborative effort to identify novel genetic variants
138 involved in human facial variation in the largest set of multi-ethnic samples available thus far.
139 Moreover, it demonstrates the functional consequences of identified face-associated DNA variants
140 based on in-silico work and *in-vitro* experimental evidence.

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 (**Figure 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 (**Figure 1-figure supplementary 1**). A total of 78 Euclidean distances involving all combinatorial pairs of the 13 facial landmarks were considered as facial phenotypes.

Detailed phenotype analyses were conducted in RS. In RS, almost all (72 out of 78, 92%,) of the facial phenotypes followed the normal distribution as formally investigated by applying the Shapiro-Wilk normality test and obtaining non-significant results (**Supplementary Table 2**). For only 6 phenotypes, this test revealed significant results; however, their histogram looked largely normal. A highly significant sex effect (min $p = 2.1\times10^{-151}$) and aging effect (min $p = 8.0\times10^{-44}$) were noted, together explained up to 21% of the variance per facial phenotype (e.g. between alares and Ls; **Figure 1-figure supplementary 2A-2C**, **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 (**Figure 1-figure supplementary 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 (**Figure 1-figure supplementary 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; **Figure 1-figure supplementary 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 dataset. Inflation factors were in an acceptable range in all meta-analyses (average λ=1.015, sd=0.006; **Figure 1-figure supplementary 4**), and were thus not further considered. A matrix spectral decomposition analysis estimated the effective number of independent facial traits as 43, which corresponds to a Bonferroni corrected study-wide significant threshold of 1.2×10^{-9}. A power analysis showed that under this threshold, our discovery data had over 90% power to detect an allelic effect explaining 0.54% variance for individual phenotypes (**Figure 1-figure supplementary 5A**). In the discovery GWAS meta-analysis, a total of 221 SNPs from 6 distinct loci showed study-wide significant association ($P<1.2\times10^{-9}$) with facial phenotypes, among which 2 have not been previously reported: 4q28.1 *INTU* (top SNP rs1250495 $p=3.03\times10^{-10}$), and 12q24.21 *TBX3* (rs1863716 $p=1.47\times10^{-10}$). The other 4 have been described in previous GWASs of facial variation, including 2q36.1 *PAX3*, 4q31.3 *SFRP2*, 17q24.3 *CASC17*, and 20p11.22 *PAX1* (**Figure 1** and **Supplementary Table 7**) (Liu et al., 2012; Paternoster et al., 2012; Adhikari et al.,

2016; Shaffer et al., 2016; Claes et al., 2018). Considering the traditional genome-wide significant threshold of 5×10^{-8} as our study-wide suggestive significance threshold for replication studies, the power analysis showed that under this threshold, our discovery data had over 90% power to detect an allelic effect explaining 0.43% variance for individual phenotypes (**Figure 1-figure supplementary 5B**). In addition to the study-wide significant ones, 273 SNPs from 18 distinct loci passed the 5×10^{-8} threshold (**Supplementary Table 7**), of which 15 were novel and 3 (1p12 *TBX15*, 9q22.31 *ROR2* and 17q24.3 *SOX9*) have been reported in previous studies (*Cha et al., 2018; Claes et al., 2018; Pickrell et al., 2016*). Thus, in total, we identified 24 face-associated genetic loci, of which 17 were not reported in previous face GWASs, that we followed-up via a multi-ethnic replication analysis (see below). In general, no significant heterogeneity was observed for the 24 top SNPs of the study-wide significant and suggestive associated loci, i.e., only one SNP (rs3403289 at 2q36.1 *PAX3*) showed nominally significant (Cochran's Q test p=0.04, **Table 1**) heterogeneity between the discovery cohorts, but was not significant after multiple testing correction.

The association signals at the 24 genetic loci involved multiple facial phenotypes clustered to the central region of the face above the upper lip (**Figure 1**), and closely resembled the twins-based face heritability map (**Figure 1-figure supplementary 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**). Four (2q36.1 *PAX3*, 9q22.31 *ROR2*, 17q24.3 *CASC17*, and 20p11.22 *PAX1*) out of the 24 face-associated loci have been noted in the GWAS catalog for face morphology or related phenotypes i.e., nose size, chin dimples, monobrow thickness, and male pattern baldness (**Supplementary Table 9**). Five out of the 24 loci showed genome-wide significant association (p<5e-8) with non-facial phenotypes and / or diseases in the GeneATLAS database (*Canela-Xandri et al., 2018*) (**Supplementary Table 10**), suggesting pleiotropic effects. Six out of the 24 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 1p31.2 *RPE65*, 2q36.1 *PAX3*, 4q28.1 *INTU*, 16q12.1 *SALL1*, 16q12.2 *RPGRIP1L* and 17q24.3 *CASC17* (**Supplementary Tables 11** and **12**). Three out of the 24 face-associated loci showed signals of positive selection in terms of the integrated haplotype scores (iHS, empirical p<0.01), including *RPE65*, *RPGRIP1L* and *CASC17* (**Supplementary Table 12**). None of the 24 loci overlapped with the previously reported

signals of singleton density score (SDS, empirical p<0.01) in Europeans (*Field et al., 2016*), which may indicate that the observed selection signals are unlikely the consequences from the recent selective pressures detectable via SDS analysis.

We selected the 24 top-associated SNPs (**Table 1**), one SNP per for each of the identified 24 genetic loci passing the study-wide suggestive significance threshold ($P<5x10^{-8}$), for replication studies in a total of 7,917 multi-ethnic subjects from 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. We used a "combined test of dependent tests" (*Kost & McDermott, 2002*) to account for the incompatibility among facial traits from the three replication cohorts, which tests for H0: no association for all facial phenotypes in all replication cohorts vs. H1: associated with at least one facial phenotypes in at least one replication cohorts. After Bonferroni correction of multiple testing of 24 SNPs, significant evidence of replication ($p<2.1x10^{-3}$) was obtained for 10 SNPs, among which 6 were novel loci: 1p36.22 *CASZ1*, 4q28.1 *INTU*, 6p21.2 *KIF6*, 12q24.21 *TBX3*, 14q32.2 *C14orf64*, and 16q12.2 *RPGRIP1L* (**Figure 2**), and 4 have been described in previous face GWASs: 2q36.1 *PAX3*, 4q31.3 *SFRP2*, 9q22.31 *ROR2*, and 20p11.22 *PAX1* (*Adhikari et al., 2016; Claes et al., 2018; Paternoster et al., 2012; Pickrell et al., 2016; Shaffer et al., 2016*).

The allelic effects of the replicated SNPs were highly consistent across all participating cohorts, while the individual cohorts were underpowered (**Figure 2-figure supplementary 1-6**), further emphasizing on the merit of our collaborative study. The associated phenotypes of the replicated SNPs 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 (**Figure 2-figure supplementary 7-17**). Several novel replicated loci showed effects on the similar sets of facial shape phenotypes; for example, 1p36.22 *CASZ1* and 4q28.1 *INTU* (**Figure 2A** and **2C**) both affected the distances between pronasion and alares. Moreover, several novel replicated loci affected phenotypes in multiple facial regions e.g. 4q32.1 *C14orf64* and 12q24.1 *TBX3* (**Figure 2D** and **2E**) both affected many distances among eyes, alares

and mouth. It should be noted however, that the discovery dataset is of European ancestry while the replication dataset is multi-ethnic and the facial phenotypes could not be obtained in the same way as in the discovery set; hence, the replication results were rather conservative, i.e., the possibility of false negatives in this replication study cannot be excluded. This was evident from the results of the replication attempts of the 7 known loci (1p12 *TBX15*, 2q36.1 *PAX3*, 4q31.3 *SFRP2*, 9q22.31 *ROR2*, 17q24.3 *CASC17*, 17q24.3 *SOX9*, 20p11.22 *PAX1*) (**Figure 2-figure supplementary 18-24**). Out of these 7, only 4 were replicated (*PAX3*, *SFRP2*, *ROR2* and *PAX1*, **Table 1**). For this reason, we considered all of the 494 SNPs from the 24 loci passing the study-wide suggestive threshold in the subsequent polygenic score analyses and in the functional annotation analyses.

Integration with previous literature knowledge

In our discovery GWAS dataset, we also investigated 122 face-associated SNPs reported in previously published 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 significant association on either the same set of facial phenotypes or at least one facial phenotypes in a combined test, after Bonferroni correction for the number of SNPs (adjusted $p<0.05$) (**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 is unlikely explained by the allele frequency differences between Sub-Saharan African (AFR) and Europeans (EUR) (absolute frequency differences between AFR and EUR were less than 0.1 in the 1000 Genomes project) but more likely due to 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 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 Bonferroni correction for the number of SNPs and phenotypes, and many (11 SNPs in 8 loci) involving cleft related facial landmarks as

expected, e.g. subnasale, labliale superius, and labiale inferius (**Supplementary Table 13**).

Multivariable model fitting and polygenic face scores

In the RS cohort, a multiple-regression conditioning on the effects of the 24 lead SNPs identified 31 SNPs showing significant (adjusted $p<0.05$) independent effects on sex- and age-adjusted face phenotypes (**Supplementary Table 14**). These 31 SNPs individually explain less than 1%, and together explain up to 4.62%, of the variance for individual phenotypes (**Figure 3A** and **3B**). 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. The variance explained was potentially overestimated in RS cohort since the SNP selection and model building were conducted using the same dataset. However, potential overfitting is unexpected to have a drastic effect, given that some of the selected SNPs showed weaker effects in RS than in other cohorts used (**Supplementary Table 7**). We then used the fitted models to derive a set of polygenetic face scores for EUR, AFR and East Asian (EAS) subjects (total N=1,668) from the 1000-Genomes Project. The polygenic face scores based on the 31 SNPs were largely reversely distributed between EUR and AFR, with EAS situated in between (**Figure 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 (**Figure 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 et al., 2011; Weiner, 1954*), e.g. nose width and nose length have been found to correlate with temperature and humidity. It should be noted, however, that the possibility of potential bias cannot be excluded due to the fact that the discovery GWAS was conducted in only Europeans. 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 (**Figure 3-figure supplementary 1**).

Gene ontology of face-associated genetic loci

A gene ontology enrichment analysis for the 24 face-associated genetic loci (**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 15**), with the top significant terms being 'embryonic digit morphogenesis' and 'embryonic appendage morphogenesis' (FDR<1×10^{-8}, **Figure 4-figure supplementary 1**). 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 (**Supplementary Table 7**) were investigated using the GTEx data. Significant multi-tissue *cis*-eQTL effects (adjusted p<0.05) were found around three genes, *ARHGEF19*, *RPE65*, and *TBX15* (**Supplementary Table 16**). 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.

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. In 89% of 9 embryonic stem cells, 65% of 20 different tissue cells, and 15% of 20 primary cells these 24 genes showed significantly higher expression in all cell types (p<0.05)

after multiple testing correction compared to the background genes (**Figure 4A**), and the most significant difference was observed in CNCCs ($p=3.8\times10^{-6}$). Furthermore, these 24 genes also showed preferential expression in CNCCs compared to their expression in other cell types, i.e., in 65% of 20 different tissue cells, in 60% of 20 primary cells, and in 22% of 9 embryonic stem cells (**Figure 4A**). Specifically, 16 (66.7%) of these 24 genes showed 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* ($p=4.5\times10^{-23}$, **Figure 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 (**Figure 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 identified face-associated DNA variants were non-exonic. 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 includes ChIP-seq data on transcription factor (*TFAP2A* and *NR2F1*) binding, general coactivator (p300) binding, or histone modifications associated with active regulatory elements (H3K4me1, H3K4me3, and H3K27ac), as well as ATAC-seq data on chromatin accessibility. Abundant entries of *cis*-regulatory signals were detected surrounding the face-associated loci. In particular, 6 (25%) of the 24 face-associated loci, i.e., 1p36.13 *ARHGEF19,* 2q36.1 *PAX3*, *CMSS1*, 4q28.1 *INTU*, 16q12.1 *SALL1*, and 16q12.2 *RPGRIP1L*, showed significant enrichment of *cis*-regulatory entries at the top 0.1% of the genome level in at least one epigenomic tracks after Bonferroni correction (adjusted $P<0.05$,

Figure 5-figure supplementary 1). We selected 5 SNPs as candidate regulatory SNPs (**Figure 5A** and **5B, Supplementary Table 17**) as they not only overlapped with the putative enhancers defined in CNCCs (*Prescott et al., 2015*) and penis foreskin melanocyte primary cells (*Chadwick, 2012; E. P. Consortium, 2012*), but also showed high LD (r^2>0.7) with at least one of the face-associated SNPs at the study-wide suggestive significance level. These 5 candidate regulatory SNPs were further studied via functional experiments as described in the following section.

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 suggested by the *in-silico* analyses outlined above as 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 (**Supplementary Table 17**). Luciferase reporter assays 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 (**Figure 5C**). In the 4q28.1 region, the derived SNP alleles were associated with decreased luciferase expression (33.5×~10.4×, **Supplementary Table 18**) 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 (1.5×~3.7×, **Supplementary Table 18**) as well as an increased length of some nasion-related features (0.131<beta<0.149, **Supplementary Table 7**). Using a luciferase assay for rs4783818 from the 16q12.2 region in CD271+ neural crest progenitor cells, we obtained an allele-specific effect on enhancer activity, although it was not statistically significant (**Figure 5C**). Additional luciferase reporter assays using G361 melanoma cells provided similar results, indicating that the regulatory activity associated with the DNA variants is conserved across the neural crest cell lineage (**Supplementary Table 18**). 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 in humans quantified from 3D facial images. We identified 24 genetic loci showing study-wide suggestive association with normal-range facial shape phenotypes in Europeans, including 17 novel ones, of which 10 were successfully replicated in additional multi-ethnic population cohorts including 6 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 identified (*INTU* and *RPGRIP1L*) and one established (*PAX3*) face shape genes.

The six replicated novel face-associated loci included 1p36.22 *CASZ1* 4q28.1 *INTU* 6p21.2 *KIF6*, 12q24.21 *TBX3*, 14q32.2 *C14orf64*, and 16q12.2 *RPGRIP1L*. For 4q28.1 *INTU* and 16q12.2 *RPGRIP1L*, 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 et al., 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*, a proximate gene to *RPGRIP1L*, has also 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*). For the remaining two novel loci (*CASZ1* and *C14orf64*), there is current lack of existing literature evidence supporting their involvement in facial variation.

The four replicated previously reported face-associated loci included 2q36.1 *PAX3*, 4q31.3 *SFRP2*, 9q22.31 *ROR2*, 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* 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*).Among the 11 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. We therefore regard it as likely that true positive findings exist among the 11 non-replicated novel loci. This is further supported by our finding that among the total of 14 non-replicated loci, 3 loci (21%) i.e., *TBX15* (*Claes et al., 2018; Pickrell et al., 2016*), *CASZ17* (*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 significant evidence from our replication analysis. Moreover, pleiotropic effects of *TBX15* 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 three 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) being most relevant here as these three genes were highlighted in Europeans in the 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 11 novel face-associated loci.

Of the previously identified face-associated genetic loci, several candidate genes e.g. *Pax3* (*Zalc et al., 2015*), *Prdm16 (Bjork et al., 2010)*, *Tp63 (Thomason et al., 2008)*, *Sfrp2* (*Satoh et al., 2008*), and *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 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 or mouse models.

Our results indirectly support the hypothesis that the facial differentiation observed among worldwide human populations is shaped by 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. However, 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 in European and African populations, demonstrating regional developments in Asia.

Nevertheless, the polygenic face scores only explained small proportions of the facial phenotype variance with up to 4.6%. 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 made here and to be made in future studies.

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

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 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 et al., 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 et al., 2011*) and imputed to the 1000 Genomes reference panel (Phase 1, Version3) (*Genomes Project et al., 2015*) using IMPUTE3 V2.2.2 (*Howie et al., 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.

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, 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). DNA samples 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 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

In our study, we focused on 13 facial landmarks available in all cohorts with 3-dimensional 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 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. Shapiro-Wilk test (*Royston, 1995*) was used to test the normality of phenotype distribution. 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 et al., 2004; Paternoster et al., 2012; Toma et al., 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. A dense registration was applied to align all the 3D images as described elsewhere (*Guo et al., 2013*). In brief, 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 distance variables was calculated from a subset of 13 of these 15 landmarks. In **CANDELA**, ordinal phenotyping was described in prior GWAS (*Adhikari et al., 2016*). In short, right side and frontal 2D photographs of participants were used to score 14 facial traits, Intra-class correlation coefficients (ICCs) were calculated to indicate a moderate-to-high intra-rater reliability of the trait scores. All photographs were scored by the same rater. In **QIMR**, two independent researchers manually evaluated all 2D portrait photos and located the coordinates for 36 facial landmarks. Satisfactory inter-rater concordance was obtained (Pearson's correlation coefficient, mean r = 0.76). The average landmark coordinates between the two researchers were used for Euclidian distance calculation.

Phenotype characteristics

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

GWAS, meta-analysis, and replication studies

GWASs for face phenotypes were independently carried out in RS, TwinsUK, ALSPAC, PITT, CANDELA, UYG, and QIMR based on linear or logistic models assuming an additive allele effect adjusted for potential confounders in each cohorts, such as family relationships, sex, age, BMI and genetic PCs using software packages including PLINK (*Purcell et al., 2007*), GEMMA (*Zhou &*

718 *Stephens, 2012*), and Rare Metal Worker (*Feng et al., 2014*). It should be noted that we used the
719 scaled General Procrustes Analysis (GPA) to eliminate potential confounding effect of BMI although
720 in some cohorts BMI was additionally included as a covariate, which shouldn't have affected the
721 results of our meta-analysis. For **RS**, association tests were run between the 78 3D facial distance
722 measurements (after scaled GPA) and 6,886,439 genotyped and imputed SNPs with MAF > 0.01,
723 imputation R2>0.8, SNP call rate > 0.97 and HWE > 0.0001. GWAS was performed using linear
724 regression under the additive genetic model while adjusting for sex, age, and the first four genomic
725 PCs derived using GCTA (*Yang et al., 2011*). For TwinsUK, quality control was performed on SNPs
726 (MAF > 0.01, imputation R2>0.8, SNP call rate > 0.99). One of each pair of monozygotic twins (MZ)
727 were excluded from association tests to reduce relatedness. The final sample included 1,020
728 individuals and 4,699,859 autosomal genotyped and imputed variants. GWAS was performed
729 between the 78 3D facial distance measurements (after scaled GPA) and genotyped and imputed
730 SNPs using linear regression under the additive genetic model while adjusting for age and the first
731 four genomic PCs. For **ALSPAC**, 3,941 study participants had both genotype and 3D facial
732 phenotype data. GCTA was used to remove individuals with excessive relatedness (GRM > 0.05) and
733 quality control was performed on SNPs (MAF > 0.01, imputation R2>0.8, SNP call rate > 0.99). The
734 final sample included 3,707 individuals and 4,627,882 autosomal genotyped and imputed SNPs.
735 GWAS was performed between the 78 3D facial distance measurements (after scaled GPA) and
736 genotyped and imputed SNPs using linear regression under the additive genetic model while
737 adjusting for age, sex, and the first four genomic PCs using PLINK 1.9 (*Purcell et al., 2007*). For **PITT**,
738 association tests were run between the 78 3D facial distance measurements (after scaled GPA) and
739 9,478,231 genotyped and imputed SNPs with MAF >1% and HWE >0.0001. GWAS was performed
740 using linear regression under the additive genetic model while adjusting for sex, age, age^2, height,
741 weight and the first four genomic PCs using PLINK 1.9. For the analysis of the X-chromosome,
742 genotypes were coded 0, 1, and 2 as per the additive genetic model for females, and coded 0, 2 for
743 males in order to maintain the same scale between sexes. Note that in the meta-analysis of all
744 cohorts, the X-chromosome was not included. For **CANDELA**, PLINK 1.9 was used to perform the
745 primary genome-wide association tests for each phenotype using multiple linear regression with an
746 additive genetic model incorporating age, sex, BMI and 5 genetic PCs as covariates. Association

747 analyses were performed on the imputed dataset with two approaches: using the best-guess
748 imputed genotypes in PLINK, and using the IMPUTE2 genotype probabilities in SNPTEST v2.5. Both
749 were consistent with each other and with the results from the chip genotype data. GWAS results
750 with using best-guess genotypes were used in the subsequent meta-analysis. For **UYG**, GWAS were
751 carried out using PLINK 1.9 and a linear regression model with additive SNP effects was used. Details
752 was described in prior GWAS (*Qiao et al., 2018*). For **QIMR**, the GWAS was conducted using Rare
753 Metal Worker with linear regression adjusted for sex, age, BMI and the first five genomic PCs.
754 Inverse variance fixed-effect meta-analysis were carried out using PLINK to combine GWAS
755 results for samples of European origin and having 3dMDface data, i.e., RS, TwinsUK, ALSPAC, and
756 PITT, which included 7,029,494 autosomal SNPs overlapping between the 4 discovery cohorts with
757 physical positions aligned according to NCBI human genome sequence build GRCh37.p13. X
758 chromosome was not included in meta-analyses. Summary statistics of all SNPs and all facial
759 phenotypes from the GWAS meta-analysis from the four cohorts of European descent (RS, TwinsUK,
760 ALSPAC and PITT) are available via figshare https://doi.org/10.6084/m9.figshare.10298396 and will
761 be made available through the NHGRI-EBI GWAS
762 Catalog https://www.ebi.ac.uk/gwas/downloads/summary-statistics. Meta-analysis results were
763 visualized using superposed Manhattan plots and Q-Q plots. All genomic inflation factors were close
764 to 1 and not further considered. Because many of the 78 facial phenotypes were correlated, we
765 applied the matrix decomposition method described by Li and Ji (*Li & Ji, 2005*) to determine the
766 effective number of independent phenotypes as 43. We then set our study-wide significant
767 threshold at $p<1.2\times10^{-9}$ using Bonferroni correction of 43 phenotypes, considering significant
768 threshold for single phenotype GWAS as $P<5\times10^{-8}$). We considered the threshold of 5×10^{-8} as the
769 threshold for replication studies in CANDELA, UYG, and QIMR. Since the 3dMDface data were
770 available in UYG, exact replications were conducted, i.e., the genetic association was tested for the
771 corresponding phenotypes in the meta-analysis. Since only 2D photos were available in CANDELA
772 and QIMR, trait-wise replications were conducted to test H0: no association for all face phenotypes
773 vs. H1: associated with at least one face phenotypes. We used a combined test of dependent tests to
774 combine genetic association tests involving highly correlated face phenotypes (*Kost & McDermott,*
775 *2002*). In short, the test statistic of the combined test follows a scaled chi-squared distribution when

tests are dependent. The mean of the χ^2 values is:

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

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

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

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

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*). We then again used the combined test of dependent tests to combine replication evidence from all three cohorts. Multiple testing was corrected for the number of selected SNPs using the Bonferroni method. Regional linkage disequilibrium r^2 values were calculated in PLINK and visualized using R scripting. Regional association 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) 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 et al., 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 study-wide suggestively significant association with face phenotypes ($p<5x10^{-8}$), we performed a multivariable regression analysis aiming to select SNPs that still be significant after regressing out the effects of the 24 lead SNPs by including them as covariates in the model, which resulted in 31 SNPs for the subsequent modeling. 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 which the differences between groups were highlighted according to the p-values from the ANOVA test, which tests H0: PS values are the same in all continental groups and H1: PS values are different in at least one of the groups.

Gene ontology analysis, expression analysis and SNP functional annotation

To further study the involved functions and regulation relationships of the genome-wide significantly associated SNP markers and candidate genes nearby, the genes located within 100kb up- and down-stream of the SNPs were selected to perform biological process enrichment analysis based on the Gene Ontology (GO) database. The enrichment analysis was performed using the R package 'clusterProfiler' (*G. Yu et al., 2012*). RNA-seq data from the study of Prescott et al. (*Prescott*

et al., 2015), GTEx (*G. T. Consortium, 2013*) and ENCODE (*E. P. Consortium, 2012*) were used to compare the differences in gene expression levels between our face-associated genes and all genes over the genome in CNCCs. We examined entries in GWAS catalog (*Welter et al., 2014*) to explore the potential pleotropic effects in our face associated regions. The GTEx (*G. T. Consortium, 2013*) dataset was used to annotate our face associated SNPs for significant single-tissue *cis*-eQTL effects in all 48 available tissues. We used the CNCC data from Prescott et al. and UCSC (*Aken et al., 2016*) to perform a variety of functional annotations for the face associated SNPs in non-coding regions using the WashU Epigenome Browser, including histone modifications (H3K27ac, H3K4me1 and H3K4me3 tracks), chromatin modifier (p300 tracks), transcription factors (TFAP2A tracks), chromatin accessibility (ATACseq tracks) and conserved regions (PhyloP (*Siepel et al., 2005*) tracks) from UCSC. To investigate the pattern of regulation within each association region in details, we calculated the log scaled quantile rank and variation of all LD SNP ($r^2 > 0.25$) in 3 replicated measurements from 6 types of Chip-seq data, and for all significant base-pair in vicinity (within 20kb) of all top-associated SNPs. All analyses were conducted using R Environment for Statistical Computing (version 3.3.1) unless otherwise specified.

Cell culture

Neural progenitors were derived from human induced pluripotent stem cells as described previously (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 et al., 1999*). CD271+ cells were grown in DMEM:F12, 1% N2 supplement, 2% B27 supplement, 1% P/S, 20 ng/ml FGF at 37°C/5% CO2. G361 cells were cultured in DMEM/10% FCS at 37°C/5% CO2. The cell line was tested negative for mycoplasma contamination and its identity was authenticated by STR profiling using PowerPlex 16 System (Promega).

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 using GoTaq qPCR Master Mix (Promega) 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. The Key Resources Table summarizing reagents and resources used in the *in-vitro* functional experiments is included in the **Supplementary Table 19 (see Supplementary File 1).**

Acknowledgements

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

Funding

This work was financially supported in part by the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 740580 (VISAGE). Z.X. is supported by the China Scholarship Council (CSC). 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). S.E.M. is supported by an NHMRC fellowship APP1103623. F.L. is supported by an Erasmus University Fellowship, the Chinese National Thousand Young Talents Award, the National Natural Science Foundation of China (NSFC) (91651507), the National Key Research and Development Program of China (2017YFC0803501), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDC01040100), and the Shanghai Municipal Science and Technology Major Project (2017SHZDZX01).

The Rotterdam Study is supported by the Netherlands Organization of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) project nr. 050-060-810. The Rotterdam Study is supported by the Erasmus MC and Erasmus University Rotterdam; the Netherlands Organization for 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).

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 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).

Conflict of Interest Statement

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

Figures

Figure 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=10,115), which were separately conducted in RS, TwinsUK, ALSPAC, and PITT 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 study-wide suggestive significance ($p<5\times10^{-8}$, red line) were nominated according to the nearby candidate genes in the associated loci, and the top 10 loci were highlighted in colors. The black line indicates study-wide significance after multiple trait correction ($p<1.2x10^{-9}$). 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.

Figure 2. Six novel genetic loci associated with facial shape phenotypes and successfully replicated.

The figure is composed by 6 parts, (**A**) 1p36.22 *CASZ1* region, (**B**) 4q28.1 *INTU* region, (**C**) 6p21.2 *KIF6* region, (**D**) 12q24.21 *TBX3* region, (**E**) 14q32.2 *C14orf64* region, and (**F**) 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 ($-\log_{10}P$). The dashed line means that the P value was nominally significant ($p<0.01$) but did not reach study-wide suggestive significance ($p>5e-8$). LocusZoom (right) shows regional association 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 study-wide suggestively significant loci are provided in **Figure 2-figure supplementary 1-24**.

Figure 3. Polygenic scores of 78 facial phenotypes.

(**A**), Facial phenotype variance is explained by a multivariable model consisting of 31 face-associated SNPs with independent effects. (**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 Sub-Saharan 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.

Figure 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. Distribution of the expression levels in CNCC showed smaller variation than those of several embryonic stem cells which showed higher median. 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 < 0.05 after multiple correcting test of cell lines amount. (**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. Dotted line represents Bonferroni corrected significant threshold (p<0.0021). Significant gene labels were in color compared to non-significant gene labels in black.

Figure 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 **Figure 2-figure supplementary 1-24.**

Supplementary Figures

Figure 1-figure supplementary 1. An example of GPA, facial landmarks obtained from 3dMD images in 3,193 participants from the discovery cohorts RS.

A, All raw landmarks before GPA process; B, after GPA process.

Figure 1-figure supplementary 2. Phenotype characteristics of 78 facial traits.

(**A**) Sex difference (RS, N=3,193); (**B**) Aging effect (RS, N=3,193); (**C**) Variance explained by sex and age (RS, N=3,193); (**D**) Unsupervised hierarchical clustering of phenotype correlation matrix resulted in 4 clusters of phenotypes (RS, N=3,193); (**E**) Twins heritability estimates in QIMR (N=1,101) using 2D frontal photographs; (**F**) Twins heritability estimates in TwinsUK (N=1,020) using 3D images.

Figure 1-figure supplementary 3. Phenotypic and genetic correlation matrix between all facial phenotypes in cohort RS.

Positive correlations are shown as brown squares; Negative correlations are shown as blue squares. Phenotypic correlation is shown above the diagonal and SNP-based genetic correlation is shown below the diagonal. The matrix was organized according to the involvement of intra- and inter- organ landmarks.

Figure 1-figure supplementary 4. Quantile-Quantile plots and genomic inflation factors for 22 significant associated facial traits.

Figure 1-figure supplementary 5. Power analysis.

The power of association tests under a variety of sample sizes (N=3,000, 6,000, 10,000 and 15,000) with varied proportions of explained phenotypic variance (R^2), absolute effect size (|ES|) and minor allele frequency (MAF) are investigated at (**A**) our study-wide significant level (1.2×10^{-9}; and (**B**) our study-wide suggestive significance level (5×10^{-8})**.** Power and R^2 are indicated by different colors. The dashed line represents the distribution of 90% power value. For example, under the sample size of

10,000, our study has 90% power to detect suggestive significant SNP explaining at least 0.43% of the phenotypic variance (R^2).

Figure 2-figure supplementary 1-24. Overviews of association, selection, LD and regulation patterns in 24 associated regions. (1-6 show 6 novel replicated loci, 7-17 show 11 novel non-replicated loci, 18-24 show 7 previously reported loci)

(**A**) From up to down, the 1st panel shows regional association plots (LocusZoom) for the top-associated facial phenotypes with candidate genes aligned below according to the chromosomal positions (GRCh37.p13). The 2nd panel shows signals of local positive selection using diversity (F_{ST}) and conservation (|iHS|) statistics between and within EUR, AFR, and EAS population samples from the 1000 Genomes Project. The 3rd panel denotes the linkage disequilibrium (LD) patterns (r^2) in the associated region. The 4th panel shows 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**) Shows a face map denoting the significance ($-\log_{10}P$) for the associations between the top SNP and facial phenotypes. The dashed line means that the P value was nominally significant ($p<0.01$) but did not reach study-wide suggestive significance ($p>5e-8$). (**C**) Effect sizes for the derived allele at index SNPs. Blue, red, and green boxes represent effect sizes estimated in the participating discovery cohort, combined meta-analysis, and in the replication UYG cohort, respectively. Box sizes are proportional to sample size. Horizontal bars indicate a 95% confidence interval of width equal to 1.96 standard errors. (**D**) All 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 below, 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).

Figure 2-figure supplementary 1. Association, selection, LD and regulation patterns for rs143353512 located around *CASZ1.*

Figure 2-figure supplementary 2. Association, selection, LD and regulation patterns for rs12504954 located around *INTU.*

Figure 2-figure supplementary 3. Association, selection, LD and regulation patterns for rs7738892 located around *KIF6.*

Figure 2-figure supplementary 4. Association, selection, LD and regulation patterns for rs1863716 located around *TBX3.*

Figure 2-figure supplementary 5. Association, selection, LD and regulation patterns for rs1989285 located around *C14orf64.*

Figure 2-figure supplementary 6. Association, selection, LD and regulation patterns for rs7404301 located around *RPGRIP1L.*

Figure 2-figure supplementary 7. Association, selection, LD and regulation patterns for rs200243292 located around *ARHGEF19.*

Figure 2-figure supplementary 8. Association, selection, LD and regulation patterns for rs77142479 located around *RPE65.*

Figure 2-figure supplementary 9. Association, selection, LD and regulation patterns for rs10202675 located around *LRRTM4.*

Figure 2-figure supplementary 10. Association, selection, LD and regulation patterns for rs2884836 located around *KIAA1715.*

Figure 2-figure supplementary 11. Association, selection, LD and regulation patterns for rs113663609 located around *CMSS1.*

Figure 2-figure supplementary 12. Association, selection, LD and regulation patterns for rs2225718 located around *RNF144B.*

Figure 2-figure supplementary 13. Association, selection, LD and regulation patterns for rs1700048 located around *CSMD1.*

Figure 2-figure supplementary 14. Association, selection, LD and regulation patterns for rs9642796 located around *DCAF4L2.*

Figure 2-figure supplementary 15. Association, selection, LD and regulation patterns for rs201719697 located around *SUPV3L1.*

Figure 2-figure supplementary 16. Association, selection, LD and regulation patterns for rs7325564 located around *LINC00371.*

Figure 2-figure supplementary 17. Association, selection, LD and regulation patterns for rs16949899 located around *SALL1.*

Figure 2-figure supplementary 18. Association, selection, LD and regulation patterns for rs1229119 located around *TBX15.*

Figure 2-figure supplementary 19. Association, selection, LD and regulation patterns for rs34032897 located around *PAX3.*

Figure 2-figure supplementary 20. Association, selection, LD and regulation patterns for rs6535972 located around *SFRP2.*

Figure 2-figure supplementary 21. Association, selection, LD and regulation patterns for rs2230578 located around *ROR2.*

Figure 2-figure supplementary 22. Association, selection, LD and regulation patterns for rs8077906 located around *CASC17.*

Figure 2-figure supplementary 23. Association, selection, LD and regulation patterns for rs35473710 located around *SOX9.*

Figure 2-figure supplementary 24. Association, selection, LD and regulation patterns for rs4813454 located around *PAX1.*

Figure 3-figure supplementary 1. Effects of NSCL/P-associated SNPs to 78 facial phenotypes.

Facial phenotype variance is explained by polygenic scores of 47 NSCL/P-associated SNPs after LD prune (r^2>0.5).

Figure 4-figure supplementary 1. Gene ontology enrichment analysis for 24 face-associated genes.

Enrichment of the genes annotating the 24 identified face-associated genes. Enriched terms (FDR<0.01 and n>2, x-axis) are indicated by corrected P-value (FDR, different color), candidate genes contained by the terms (n, sizes of dots) and ratio of enrichment between test genes and background genes (y-axis).

Figure 5-figure supplementary 1. Overview of regulation activity of 24 face-associated loci in CNCCs.

For the top-associated SNPs at the 24 face associated loci, their base-pairs in vicinity within 20kb (red bar) and LD (r^2 > 0.25) SNPs (blue bar) were examined for *cis*-regulatory entries in six epigenomic tracks in CNCCs at three thresholds (top 1%, top 0.5%, top 0.1% of the genome). The proportions of the base-pairs or LD SNPs at each locus passing the set thresholds were displayed using bars (left y-axis), and the folds of the increase over the set thresholds were displayed using dashed lines (right y-

1178 axis). Statistical significance was derived using the Fisher's exact test (*adjust.p < 0.05).

Table

Table 1. SNPs significantly associated ($p<5x10^{-8}$) with facial shape phenotypes from European discovery GWAS meta-analysis (RS, TwinksUK, ALSPAC, and PITT), and their multi-ethnic replication (UYG, CANDELA and QIMR).

						Discovery meta-analysis (N=10,115)			Replication (N=7,917)
Region	SNP	Nearest Gene	EA	OA	Trait	Beta	P	Q	Com. P
Novel face-associated loci									
1p36.22	rs143353512	***CASZ1***	A	G	Prn-AlL	-0.29	6.44×10^{-9}	0.73	**0.0001**
1p36.13	rs200243292	*ARHGEF19*	I	T	EnR-ChL	-0.11	1.46×10^{-8}	0.54	0.0640
1p31.2	rs77142479	*RPE65*	C	A	EnL-Sn	-0.08	7.65×10^{-9}	0.90	0.0121
2p12	rs10202675	*LRRTM4*	T	C	EnR-AlR	-0.26	4.43×10^{-8}	0.40	0.0226
2q31.1	rs2884836	*KIAA1715*	T	C	ExR-ChR	-0.09	3.04×10^{-8}	0.52	0.1096
3q12.1	rs113663609	*CMSS1*	A	G	EnR-ChR	-0.12	3.46×10^{-8}	0.96	0.0782
4q28.1	rs12504954	***INTU***	A	G	Prn-AlL	-0.09	$\mathbf{3.03\times10^{-10}}$	0.52	**0.0015**
6p22.3	rs2225718	*RNF144B*	C	T	EnR-AlR	-0.09	2.97×10^{-8}	0.60	0.0071
6p21.2	rs7738892	***KIF6***	T	C	EnR-N	0.11	1.34×10^{-8}	0.09	**0.0018**
8p23.2	rs1700048	*CSMD1*	C	A	ExR-ChL	-0.22	3.94×10^{-8}	0.73	0.1131
8q21.3	rs9642796	*DCAF4L2*	A	G	AlL-Ls	-0.11	4.52×10^{-8}	0.07	0.3090
10q22.1	rs201719697	*SUPV3L1*	A	D	ExL-AlL	-0.28	3.42×10^{-8}	0.66	NA
12q24.21	rs1863716	***TBX3***	A	C	Prn-AlR	-0.09	$\mathbf{1.47\times10^{-10}}$	0.53	$\mathbf{2.45\times10^{-5}}$
13q14.3	rs7325564	*LINC00371*	T	C	N-Prn	0.09	1.34×10^{-8}	0.16	0.3519
14q32.2	rs1989285	***C14orf64***	G	C	N-Prn	-0.10	4.54×10^{-8}	0.82	**0.0019**
16q12.1	rs16949899	*SALL1*	T	G	ExR-ChL	-0.13	3.35×10^{-8}	0.76	0.8912
16q12.2	rs7404301	***RPGRIP1L***	G	A	AlL-Ls	-0.09	3.49×10^{-9}	0.43	$\mathbf{2.87\times10^{-7}}$
Previously reported face-associated loci									
1p12	rs1229119	*TBX15*	T	C	ExR-ChR	0.08	2.25×10^{-8}	0.26	0.1219
2q36.1	rs34032897	***PAX3***	G	A	EnR-N	0.14	$\mathbf{1.96\times10^{-17}}$	0.04	**0.0020**
4q31.3	rs6535972	***SFRP2***	C	G	EnL-AlL	0.12	$\mathbf{3.51\times10^{-12}}$	0.12	$\mathbf{5.89\times10^{-12}}$
9q22.31	rs2230578	***ROR2***	C	T	EnL-Sn	0.09	9.02×10^{-9}	0.66	$\mathbf{1.09\times10^{-6}}$
17q24.3	rs8077906	*CASC17*	A	G	Prn-EnL	-0.09	$\mathbf{3.00\times10^{-11}}$	0.24	0.0505
17q24.3	rs35473710	*SOX9*	G	A	AlR-ChL	0.19	3.95×10^{-8}	0.37	0.2634
20p11.22	rs4813454	***PAX1***	T	C	AlL-AlR	-0.10	$\mathbf{2.32\times10^{-11}}$	0.58	**0.0002**

Except rs2230578 (3'-UTR of *ROR2*) and rs7738892 (a synonymous variant of *KIF6*), all SNPs are intronic or intergenic. Gene symbols in bold indicate successful replication after Bonferroni correction of 24 loci ($p<2.1x10^{-3}$). Meta P in bold indicates study-wide significance in the discovery analysis after multiple trait correction ($p<1.2x10^{-9}$). Replication P in bold means significant after Bonferroni correction for multiple tests ($p<2.1x10^{-3}$). Q: p-value from the Cochran's Q test for testing heterogeneity between discovery cohorts. EA and OA, the effect allele and the another allele. Com. P: p-values from a combined test of dependent tests. NA, not available.

Supplementary File 1

Which include:

Supplementary Table 1. Characteristics of the all cohorts.

Supplementary Table 2 Characteristics of 78 facial phenotypes in RS.

Supplementary Table 3 Effects of 6 covariates to 78 facial phenotypes in RS.

Supplementary Table 4. Phenotype correlations between all facial phenotypes in RS.

Supplementary Table 5. Genetic correlation between all facial phenotypes in RS.

Supplementary Table 6. Twins heritability and SNP-based heritability of all facial phenotypes in TwinsUK and QIMR.

Supplementary Table7. 494 SNPs associated (P<5e-8) with facial shape phenotypes in meta-analysis of GWAS.

Supplementary Table 8. Replication of reported variants of facial variation.

Supplementary Table 9. GWAS Catalog entries (release on 2018-01-01), for polymorphisms that were significantly associated with facial traits in meta-analyses.

Supplementary Table 10. Significant pleiotropic effect of study-wide suggestive SNPs on GeneATLAS traits.

Supplementary Table 11. Frequency distribution of 24 top-associated SNPs.

Supplementary Table 12. Population differentiation and positive selection at the suggestive significant GWAS SNPs.

Supplementary Table 13. Replication of reported variants of NSCL/P trait.

Supplementary Table 14. Multiple linear regression for 31 top-associated SNPs fitting with SEX and AGE in RS.

Supplementary Table 15. Enrichment of the genes annotating the associated 24 regions.

Supplementary Table 16. Significant expression quantitative trait loci effects for 494 SNPs significantly associated with facial traits and the target genes.

Supplementary Table 17. Overviews of 5 selected candidate regulatory SNPs.

Supplementary Table 18. Luciferase reporter assay results overview presented as relative luciferase activity fold change compared to construct containing less active allele ± standard deviation.

Supplementary Table 19. Key Resource Table for functional experiments.

References

Adhikari, K, Fuentes-Guajardo, M, Quinto-Sanchez, M, Mendoza-Revilla, J, Camilo Chacon-Duque, J, Acuna-Alonzo, V, et al. (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, AJ, Konka, E, Palmen, J, Quinto-Sanchez, M, et al. (2015). A genome-wide association study identifies multiple loci for variation in human ear morphology. *Nat Commun, 6*, 7500. doi:10.1038/ncomms8500

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

Bamshad, M, Lin, RC, Law, DJ, Watkins, WC, Krakowiak, PA, Moore, ME, et al. (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, TH, Murray, JC, Marazita, ML, Munger, RG, Ruczinski, I, Hetmanski, JB, et al. (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, TH, Ruczinski, I, Murray, JC, Marazita, ML, Munger, RG, Hetmanski, JB, et al. (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, KU, Reutter, H, Herms, S, Steffens, M, Rubini, M, et al. (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, BC, Turbe-Doan, A, Prysak, M, Herron, BJ, & Beier, DR. (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, DA, Fraser, A, Henderson, J, et al. (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, ME, & LeDouarin, NM. (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, AL, Franco, B, Duffourd, Y, Thevenon, J, Jego, L, Lopez, E, et al. (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, JG, Gray, RS, Gansner, JM, Alvarado, DM, Burgert, L, Gitlin, JD, et al. (2014). Kinesin family member 6 (kif6) is necessary for spine development in zebrafish. *Dev Dyn, 243*(12), 1646-1657. doi:10.1002/dvdy.24208

Canela-Xandri, O, Rawlik, K, & Tenesa, A. (2018). An atlas of genetic associations in UK Biobank. *Nat Genet, 50*(11), 1593-1599. doi:10.1038/s41588-018-0248-z

Carroll, SB. (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, JE, Park, AY, Do, JH, Lee, SW, Shin, C, et al. (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, LH. (2012). The NIH Roadmap Epigenomics Program data resource. *Epigenomics, 4*(3), 317-324. doi:10.2217/epi.12.18

Claes, P, Roosenboom, J, White, JD, Swigut, T, Sero, D, Li, J, et al. (2018). Genome-wide mapping of global-to-

1263 local genetic effects on human facial shape. *Nat Genet, 50*, 414-423. doi:10.1038/s41588-018-0057-4
1264 Cole, JB, Manyama, M, Kimwaga, E, Mathayo, J, Larson, JR, Liberton, DK, et al. (2016). Genomewide Association
1265 Study of African Children Identifies Association of SCHIP1 and PDE8A with Facial Size and Shape. *PLoS*
1266 *Genet, 12*(8), e1006174. doi:10.1371/journal.pgen.1006174
1267 Consortium, EP. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489*(7414),
1268 57-74. doi:10.1038/nature11247
1269 Consortium, GT. (2013). The Genotype-Tissue Expression (GTEx) project. *Nat Genet, 45*(6), 580-585.
1270 doi:10.1038/ng.2653
1271 Cordero, DR, Brugmann, S, Chu, Y, Bajpai, R, Jame, M, & Helms, JA. (2011). Cranial neural crest cells on the move:
1272 their roles in craniofacial development. *Am J Med Genet A, 155A*(2), 270-279. doi:10.1002/ajmg.a.33702
1273 de Jong, MA, Hysi, P, Spector, T, Niessen, W, Koudstaal, MJ, Wolvius, EB, et al. (2018). Ensemble landmarking of
1274 3D facial surface scans. *Sci Rep, 8*(1), 12. doi:10.1038/s41598-017-18294-x
1275 de Jong, MA, Wollstein, A, Ruff, C, Dunaway, D, Hysi, P, Spector, T, et al. (2016). An Automatic 3D Facial
1276 Landmarking Algorithm Using 2D Gabor Wavelets. *IEEE Trans Image Process, 25*(2), 580-588.
1277 doi:10.1109/TIP.2015.2496183
1278 Delaneau, O, Marchini, J, & Zagury, JF. (2011). A linear complexity phasing method for thousands of genomes.
1279 *Nat Methods, 9*(2), 179-181. doi:10.1038/nmeth.1785
1280 Delous, M, Baala, L, Salomon, R, Laclef, C, Vierkotten, J, Tory, K, et al. (2007). The ciliary gene RPGRIP1L is mutated
1281 in cerebello-oculo-renal syndrome (Joubert syndrome type B) and Meckel syndrome. *Nat Genet, 39*(7),
1282 875-881. doi:10.1038/ng2039
1283 Djordjevic, J, Zhurov, AI, Richmond, S, & Visigen, C. (2016). Genetic and Environmental Contributions to Facial
1284 Morphological Variation: A 3D Population-Based Twin Study. *PLoS One, 11*(9), e0162250.
1285 doi:10.1371/journal.pone.0162250
1286 Dupin, E, & Coelho-Aguiar, JM. (2013). Isolation and differentiation properties of neural crest stem cells.
1287 *Cytometry A, 83*(1), 38-47. doi:10.1002/cyto.a.22098
1288 Feng, S, Liu, D, Zhan, X, Wing, MK, & Abecasis, GR. (2014). RAREMETAL: fast and powerful meta-analysis for rare
1289 variants. *Bioinformatics, 30*(19), 2828-2829. doi:10.1093/bioinformatics/btu367
1290 Field, Y, Boyle, EA, Telis, N, Gao, Z, Gaulton, KJ, Golan, D, et al. (2016). Detection of human adaptation during the
1291 past 2000 years. *Science, 354*(6313), 760-764. doi:10.1126/science.aag0776
1292 Fraser, A, Macdonald-Wallis, C, Tilling, K, Boyd, A, Golding, J, Davey Smith, G, et al. (2013). Cohort Profile: the
1293 Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol, 42*(1), 97-110.
1294 doi:10.1093/ije/dys066
1295 Genomes Project, C, Abecasis, GR, Auton, A, Brooks, LD, DePristo, MA, Durbin, RM, et al. (2012). An integrated
1296 map of genetic variation from 1,092 human genomes. *Nature, 491*(7422), 56-65.
1297 doi:10.1038/nature11632
1298 Genomes Project, C, Auton, A, Brooks, LD, Durbin, RM, Garrison, EP, Kang, HM, et al. (2015). A global reference
1299 for human genetic variation. *Nature, 526*(7571), 68-74. doi:10.1038/nature15393
1300 Golding, J, Pembrey, M, Jones, R, & Team, AS. (2001). ALSPAC--the Avon Longitudinal Study of Parents and
1301 Children. I. Study methodology. *Paediatr Perinat Epidemiol, 15*(1), 74-87.
1302 Grant, SF, Wang, K, Zhang, H, Glaberson, W, Annaiah, K, Kim, CE, et al. (2009). A genome-wide association study
1303 identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. *J Pediatr, 155*(6), 909-
1304 913. doi:10.1016/j.jpeds.2009.06.020
1305 Gunhanlar, N, Shpak, G, van der Kroeg, M, Gouty-Colomer, LA, Munshi, ST, Lendemeijer, B, et al. (2018). A
1306 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, CM, Franco, OH, Goedegebure, A, Ikram, MA, et al. (2013). The Rotterdam Study: 2014 objectives and design update. *Eur J Epidemiol, 28*(11), 889-926. doi:10.1007/s10654-013-9866-z

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

Howie, BN, 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, CC, & Joyner, AL. (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

Hysi, PG, Valdes, AM, Liu, F, Furlotte, NA, Evans, DM, Bataille, V, et al. (2018). Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability. *Nat Genet, 50*(5), 652-656. doi:10.1038/s41588-018-0100-5

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

Kamberov, YG, Wang, S, Tan, J, Gerbault, P, Wark, A, Tan, L, et al. (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, CH, 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, AC, Rivadeneira, F, Lao, O, van Duijn, K, et al. (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, et al. (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, JT, & McDermott, MP. (2002). Combining dependent P-values. *Statistics & Probability Letters, 60*(2), 183-190.

Lango Allen, H, Estrada, K, Lettre, G, Berndt, SI, Weedon, MN, Rivadeneira, F, et al. (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, et al. (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, MK, Shaffer, JR, Leslie, EJ, Orlova, E, Carlson, JC, Feingold, E, et al. (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

Leslie, EJ, Carlson, JC, Shaffer, JR, Butali, A, Buxo, CJ, Castilla, EE, et al. (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, MM, Hysi, PG, et al. (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, KU, Bohmer, AC, Bowes, J, Nikolic, M, Ishorst, N, Wyatt, N, et al. (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, KU, Mangold, E, Herms, S, Nowak, S, Reutter, H, Paul, A, et al. (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, KU, Birnbaum, S, Baluardo, C, Ferrian, M, Herms, S, et al. (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, SJ, White, PM, Zock, C, & Anderson, DJ. (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, ML, 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, et al. (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, AI, Toma, AM, Kemp, JP, St Pourcain, B, Timpson, NJ, et al. (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, JK, Berisa, T, Liu, JZ, Segurel, L, Tung, JY, & Hinds, DA. (2016). Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet, 48*(7), 709-717. doi:10.1038/ng.3570

Prescott, SL, Srinivasan, R, Marchetto, MC, Grishina, I, Narvaiza, I, Selleri, L, et al. (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, RJ, Welch, RP, Sanna, S, Teslovich, TM, Chines, PS, Gliedt, TP, et al. (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, MA, Bender, D, et al. (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, et al. (2018). Genome-wide variants of 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

Royston, P. (1995). Remark AS R94: A Remark on Algorithm AS 181: The W-test for Normality. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 44*(4), 547-551. doi:10.2307/2986146

Ruiz-Linares, A, Adhikari, K, Acuna-Alonzo, V, Quinto-Sanchez, M, Jaramillo, C, Arias, W, et al. (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, PC, Reich, DE, Higgins, JM, Levine, HZ, Richter, DJ, Schaffner, SF, et al. (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, RD, et al. (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, GC, Trepczik, B, Suring, K, Brieske, N, Tucker, AS, Sharpe, PT, et al. (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, JR, Orlova, E, Lee, MK, Leslie, EJ, Raffensperger, ZD, Heike, CL, et al. (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, JS, Hinrichs, AS, Hou, M, Rosenbloom, K, et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res, 15*(8), 1034-1050. doi:10.1101/gr.3715005

Singh, MK, 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, KS, Hedman, AK, Grundberg, E, Nica, AC, Thorleifsson, G, Kong, A, et al. (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, et al. (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, PC, 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, HA, Dixon, MJ, & 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, AM, 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

Visconti, A, Duffy, DL, Liu, F, Zhu, G, Wu, W, Chen, Y, et al. (2018). Genome-wide association study in 176,678 Europeans reveals genetic loci for tanning response to sun exposure. *Nat Commun, 9*(1), 1684. doi:10.1038/s41467-018-04086-y

Visser, M, Kayser, M, & Palstra, RJ. (2012). HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. *Genome Res, 22*(3), 446-455. doi:10.1101/gr.128652.111

Visser, M, Palstra, RJ, & Kayser, M. (2014). Human skin color is influenced by an intergenic DNA polymorphism regulating transcription of the nearby BNC2 pigmentation gene. *Hum Mol Genet, 23*(21), 5750-5762. doi:10.1093/hmg/ddu289

Visser, M, Palstra, RJ, & Kayser, M. (2015). Allele-specific transcriptional regulation of IRF4 in melanocytes is mediated by chromatin looping of the intronic rs12203592 enhancer to the IRF4 promoter. *Hum Mol Genet, 24*(9), 2649-2661. doi:10.1093/hmg/ddv029

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

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

Weinberg, SM, Raffensperger, ZD, Kesterke, MJ, Heike, CL, Cunningham, ML, Hecht, JT, et al. (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, JS. (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, et al. (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, AR, Esko, T, Yang, J, Vedantam, S, Pers, TH, Gustafsson, S, et al. (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, GA. (2007). The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet, 8*(3), 206-216. doi:10.1038/nrg2063

Yang, J, Lee, SH, Goddard, ME, & Visscher, PM. (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, LG, Han, Y, & He, QY. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS, 16*(5), 284-287. doi:10.1089/omi.2011.0118

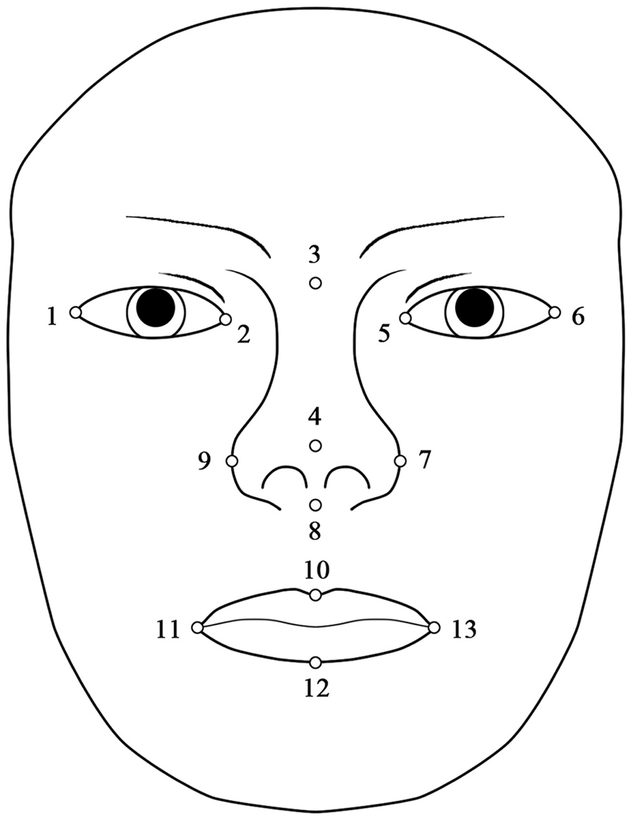
Yu, Y, Zuo, X, He, M, Gao, J, Fu, Y, Qin, C, et al. (2017). Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. *Nat Commun, 8*, 14364. doi:10.1038/ncomms14364

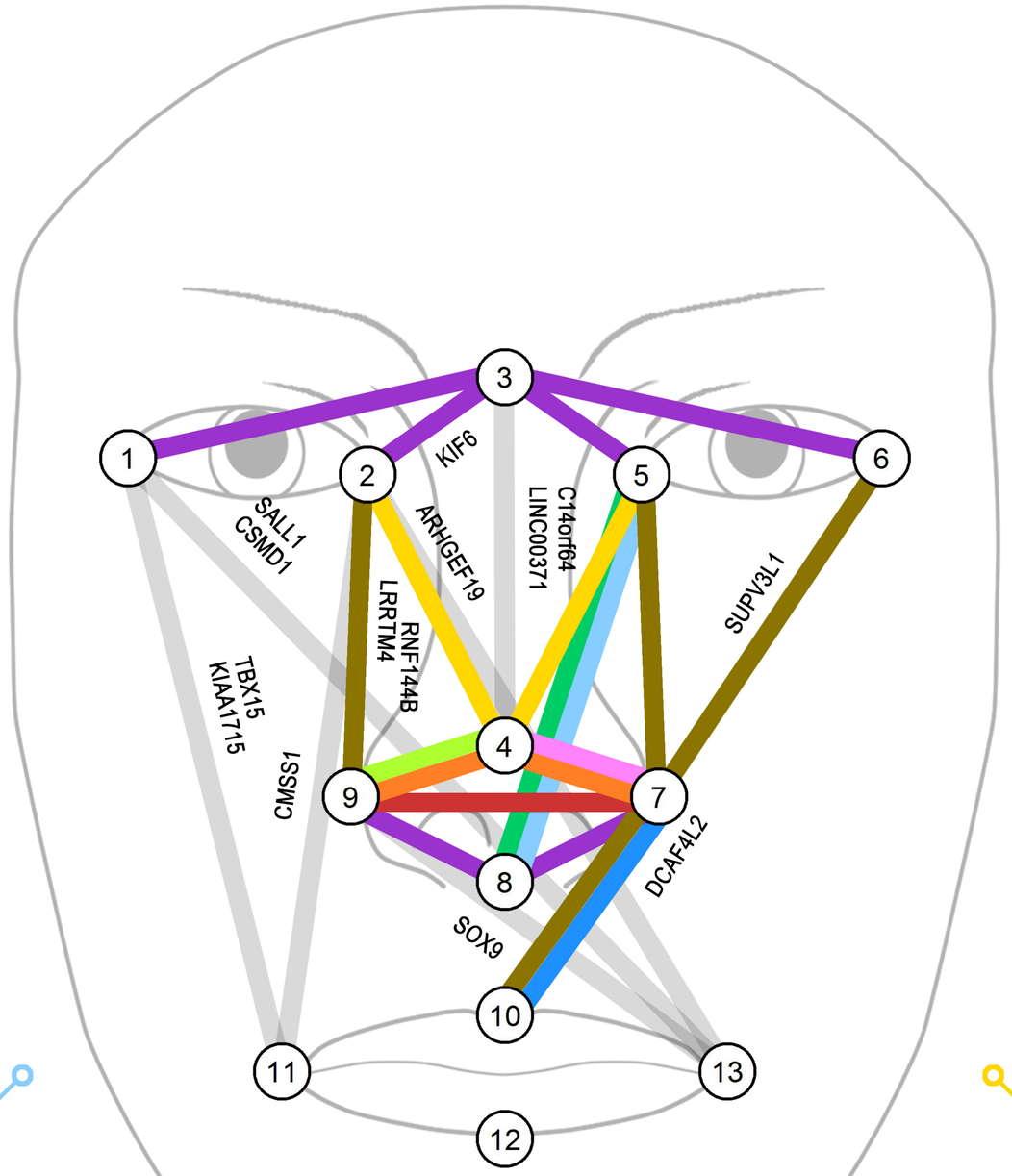
Zalc, A, Rattenbach, R, Aurade, F, Cadot, B, & Relaix, F. (2015). Pax3 and Pax7 play essential safeguard functions against environmental stress-induced birth defects. *Dev Cell, 33*(1), 56-66. doi:10.1016/j.devcel.2015.02.006

Zeng, H, Hoover, AN, & 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

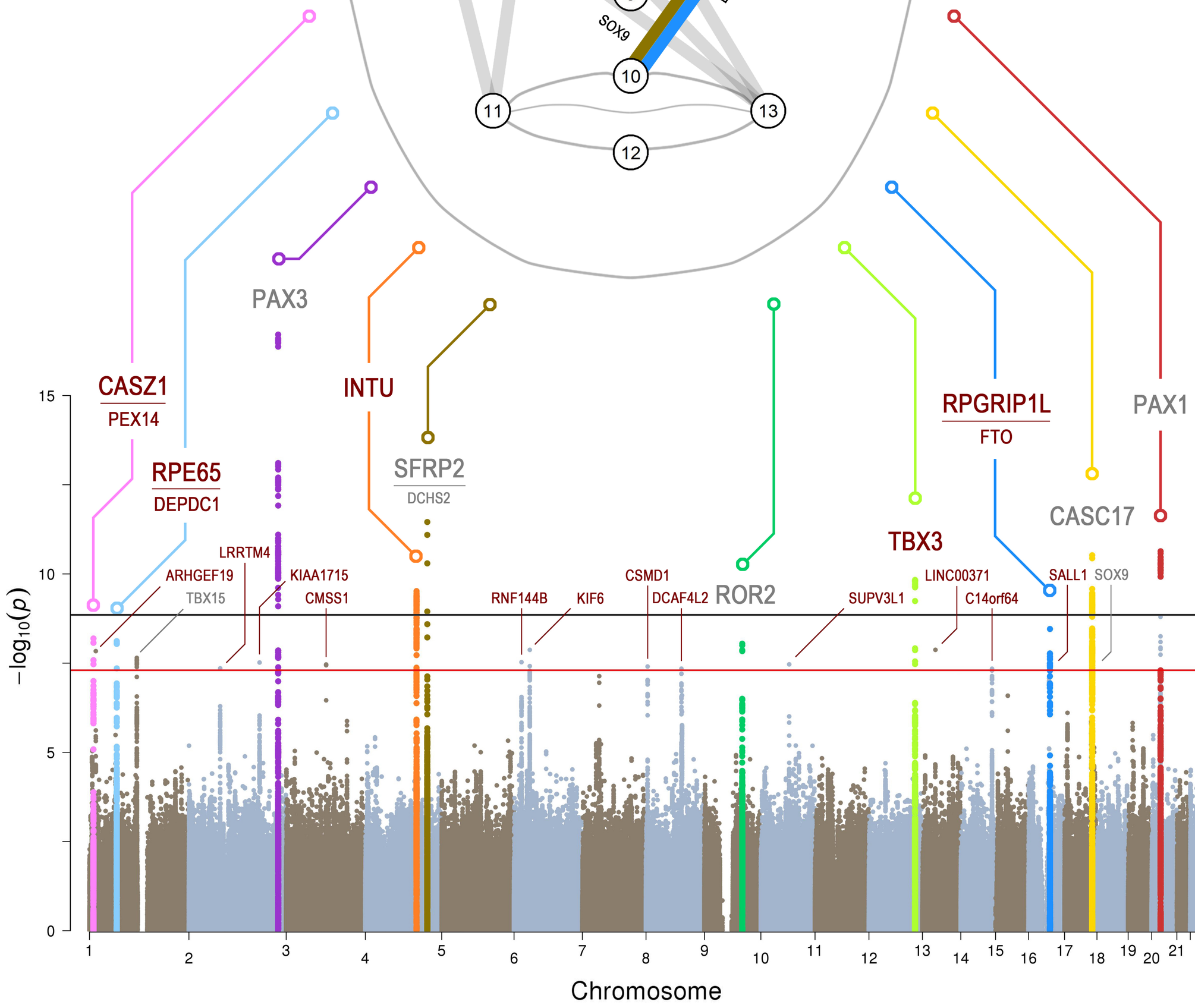
Zhou, X, & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat Genet, 44*(7), 821-824. doi:10.1038/ng.2310

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

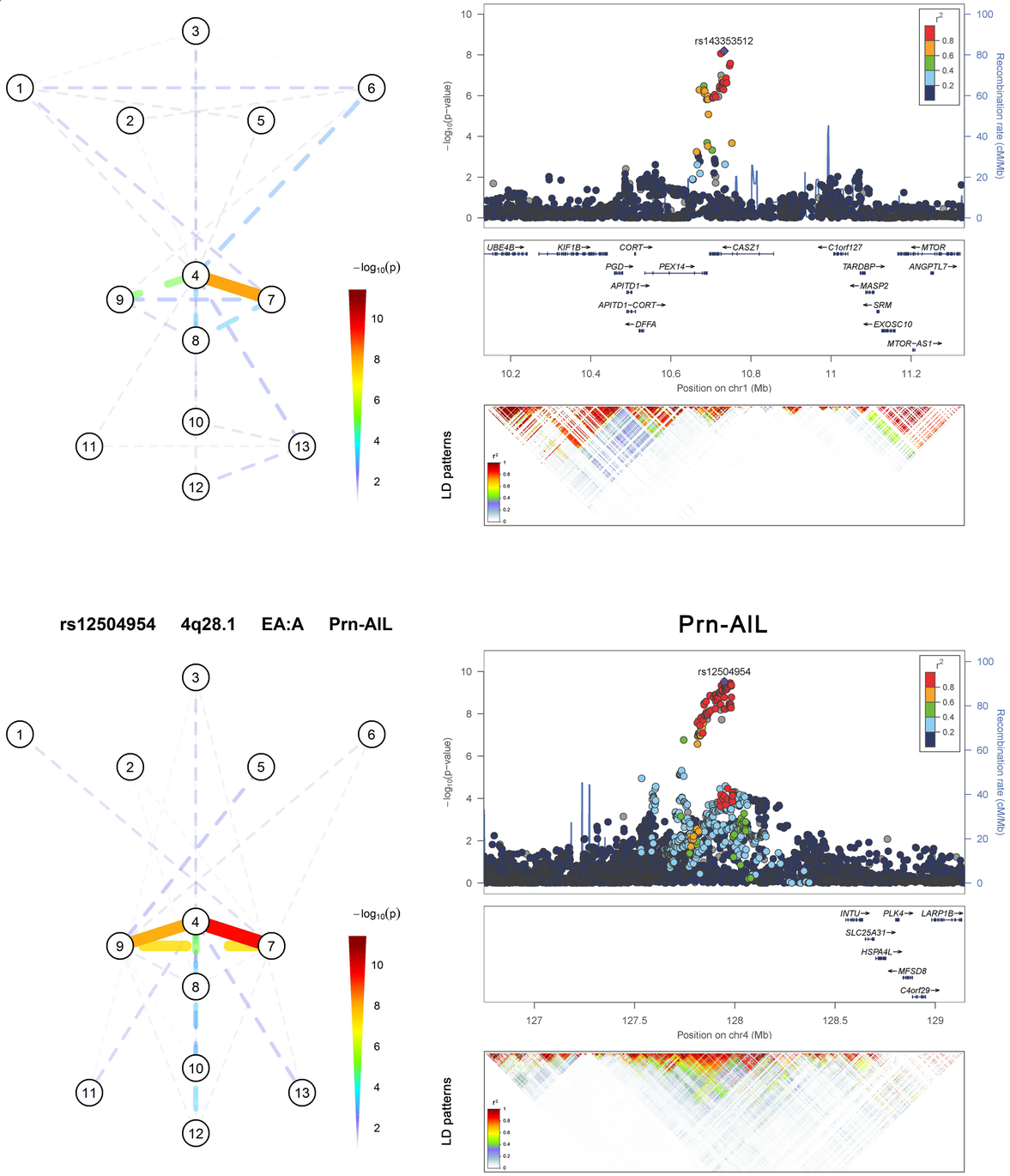




No	Abbreviation	Name
1	ExR	Right Exocanthion
2	EnR	Right Endocanthion
3	N	Nasion
4	Prn	Pronasale
5	EnL	Left Endocanthion
6	ExL	Left Exocanthion
7	AlL	Left Alare
8	Sn	Subnasale
9	AlR	Right Alare
10	Ls	Labliale superius
11	ChR	Right Cheilion
12	Li	Labiale inferius
13	ChL	Left Cheilion

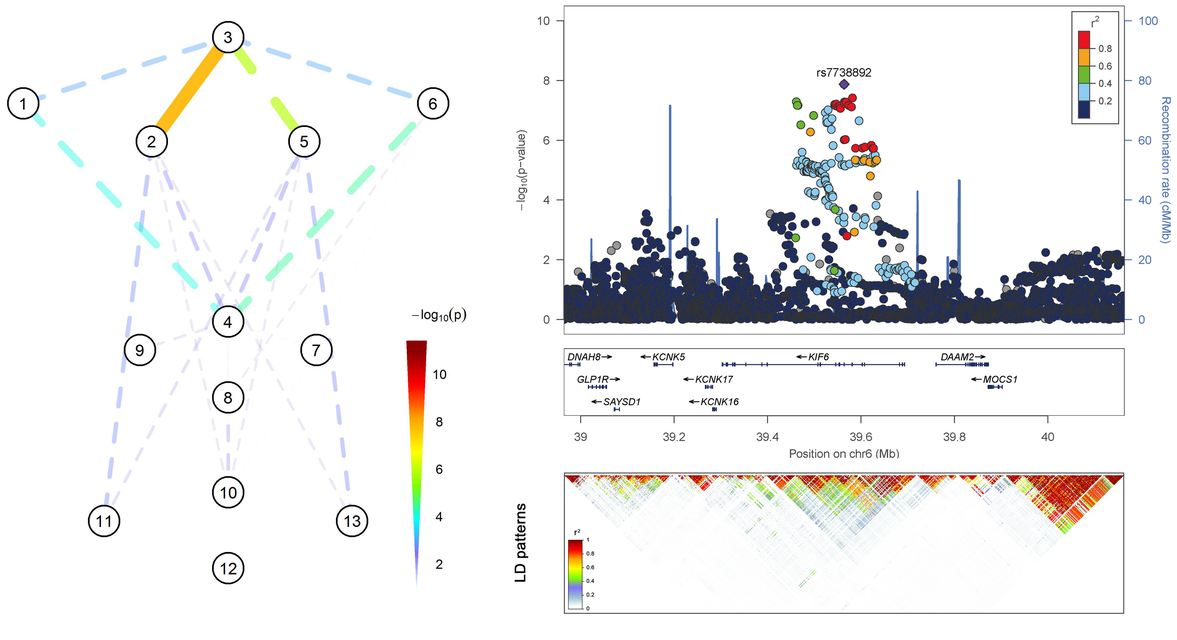


A rs143353512 1p36.22 EA:A Prn-AIL

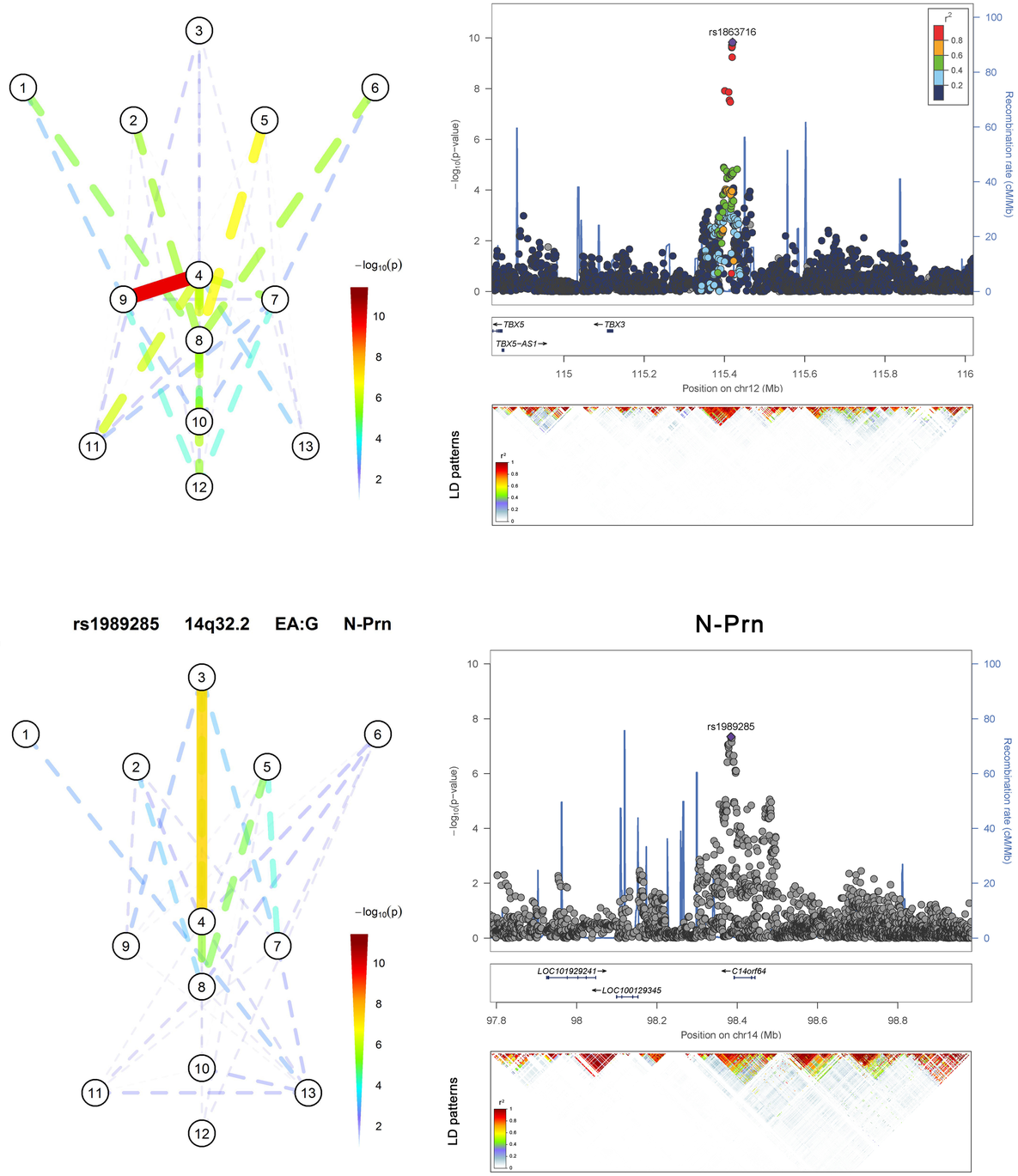


B rs12504954 4q28.1 EA:A Prn-AIL

C rs7738892 6p21.2 EA:T EnR-N

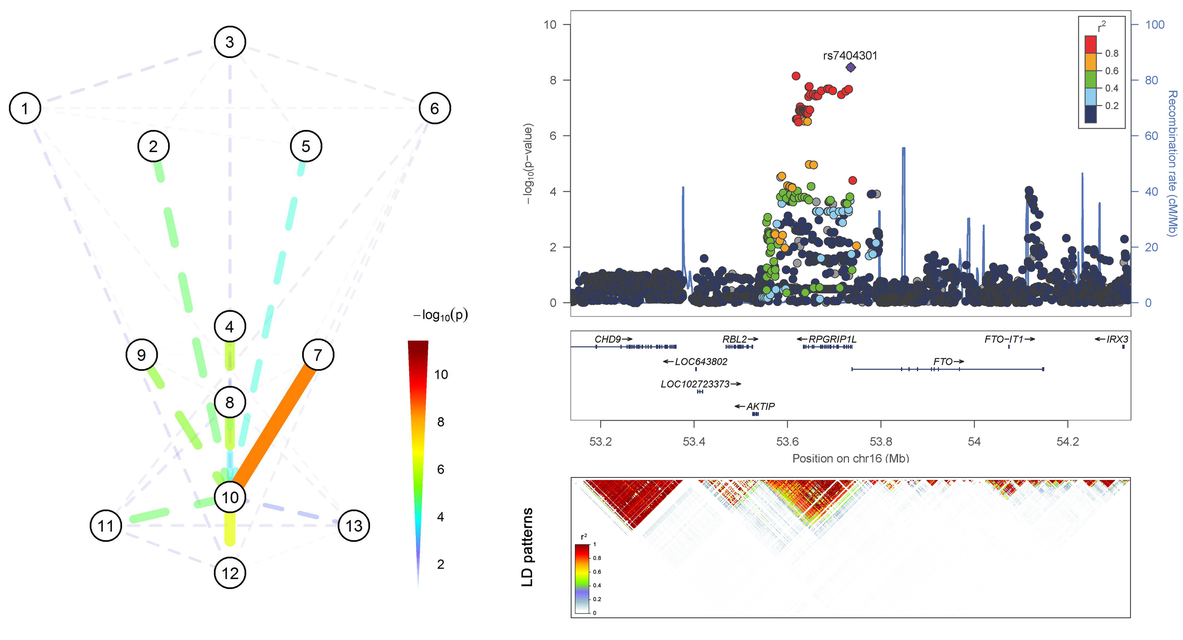


D rs1863716 12q24.21 EA:A AIR-Prn

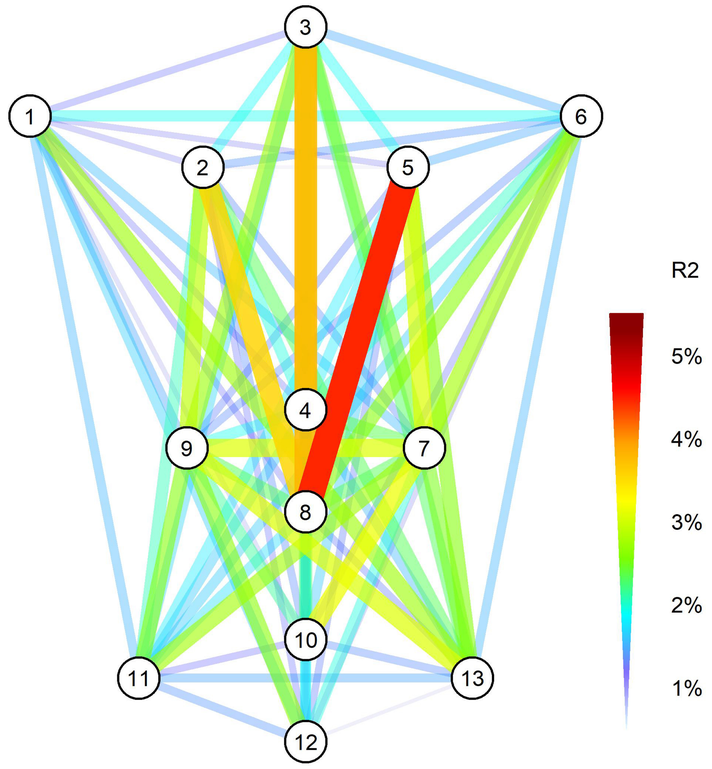


E rs1989285 14q32.2 EA:G N-Prn

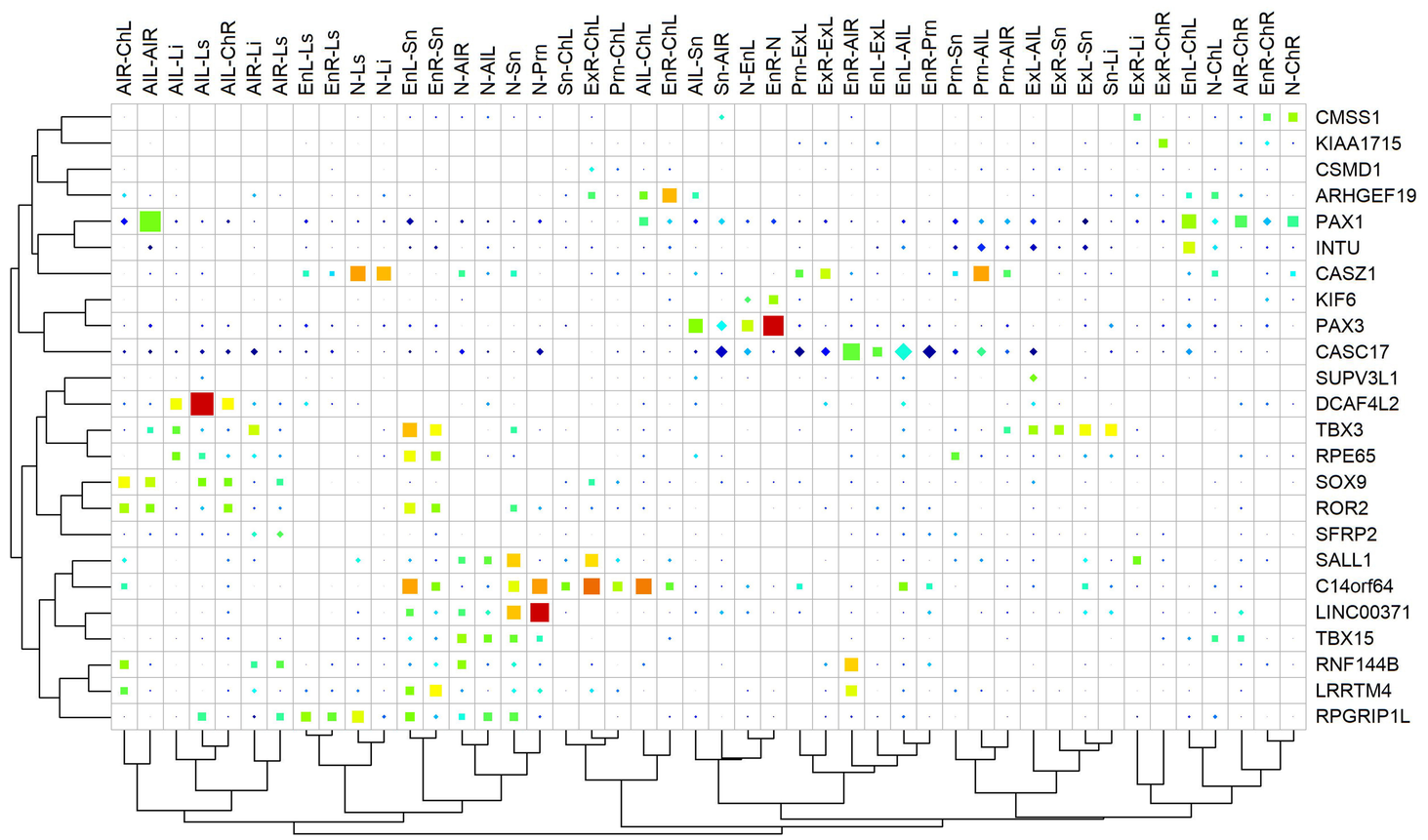
F rs7404301 16q12.2 EA:G AIL-Ls



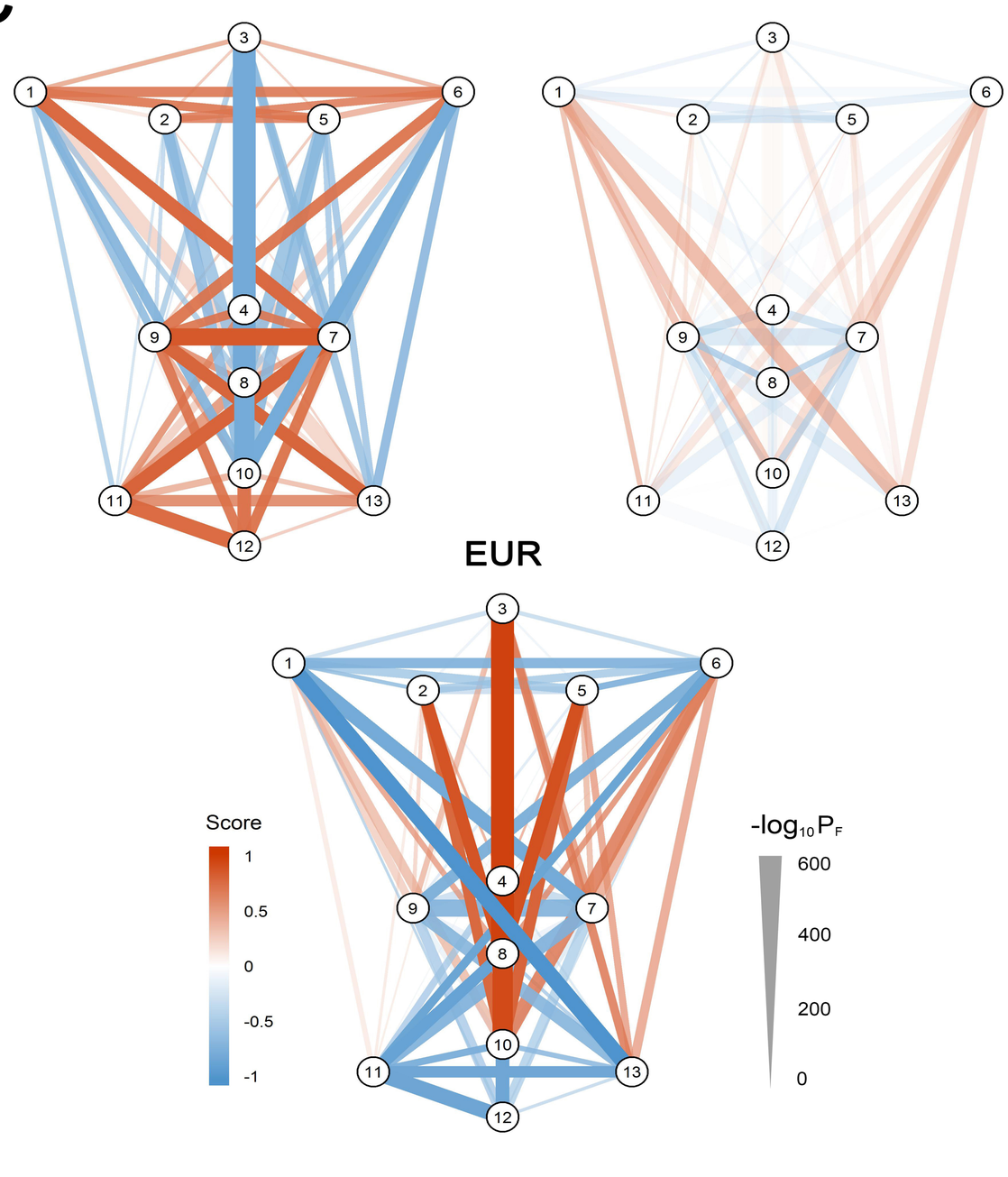
A


31 SNP R2
1
2
3
4
5
6
7
8
9
10
11
12
13
R2
5%
4%
3%
2%
1%

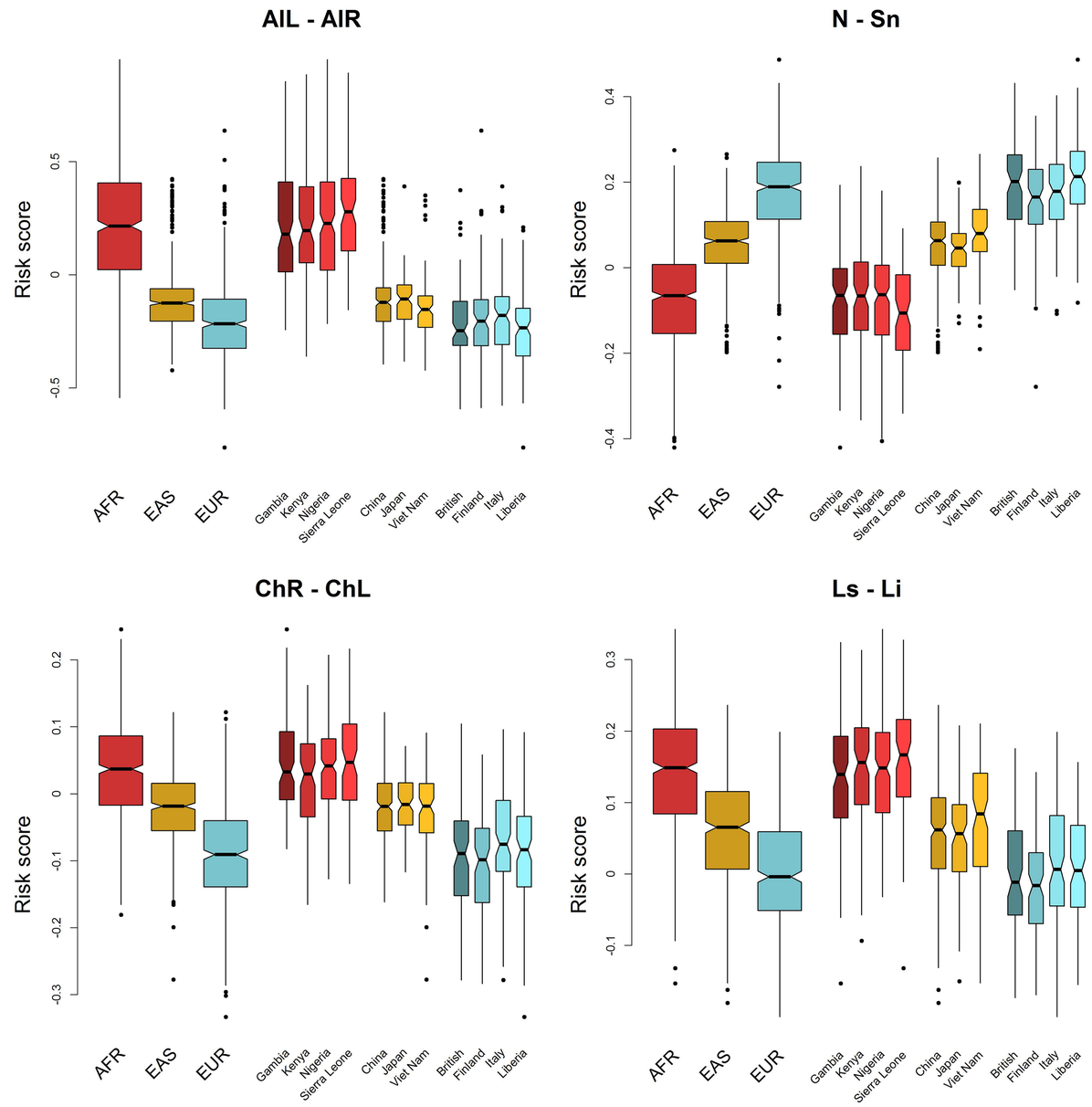
B


AIR-ChL
AIL-AIR
AIL-Li
AIL-Ls
AIL-ChR
AIR-Li
AIR-Ls
EnL-Ls
EnR-Ls
N-Ls
N-Li
EnL-Sn
EnR-Sn
N-AIR
N-AIL
N-Sn
N-Prn
Sn-ChL
ExR-ChL
Prn-ChL
AIL-ChL
EnR-ChL
AIL-Sn
Sn-AIR
N-EnL
EnR-N
Prn-ExL
ExR-ExL
EnR-AIR
EnL-ExL
EnL-AIL
EnR-Prn
Prn-Sn
Prn-AIL
Prn-AIR
ExL-AIL
ExR-Sn
ExL-Sn
Sn-Li
ExR-Li
ExR-ChR
EnL-ChL
N-ChL
AIR-ChR
EnR-ChR
N-ChR
CMSS1
KIAA1715
CSMD1
ARHGEF19
PAX1
INTU
CASZ1
KIF6
PAX3
CASC17
SUPV3L1
DCAF4L2
TBX3
RPE65
SOX9
ROR2
SFRP2
SALL1
C14orf64
LINC00371
TBX15
RNF144B
LRRTM4
RPGRIP1L
-log10(p)
5
4
3
2
1
R2 (%)
0.2
0.4
0.6
0.8
> 1

C


AFR
EAS
EUR
Score
1
0.5
0
-0.5
-1
-log10PF
600
400
200
0

D


AIL - AIR
N - Sn
ChR - ChL
Ls - Li
Risk score
AFR
EAS
EUR
Gambia
Kenya
Nigeria
Sierra Leone
China
Japan
Viet Nam
British
Finland
Italy
Liberia

A

CNCC *vs* non-CNCC

24 genes *vs* bg.

Normalized RNA-seq VST values

24 genes at the 24 face-associated regions

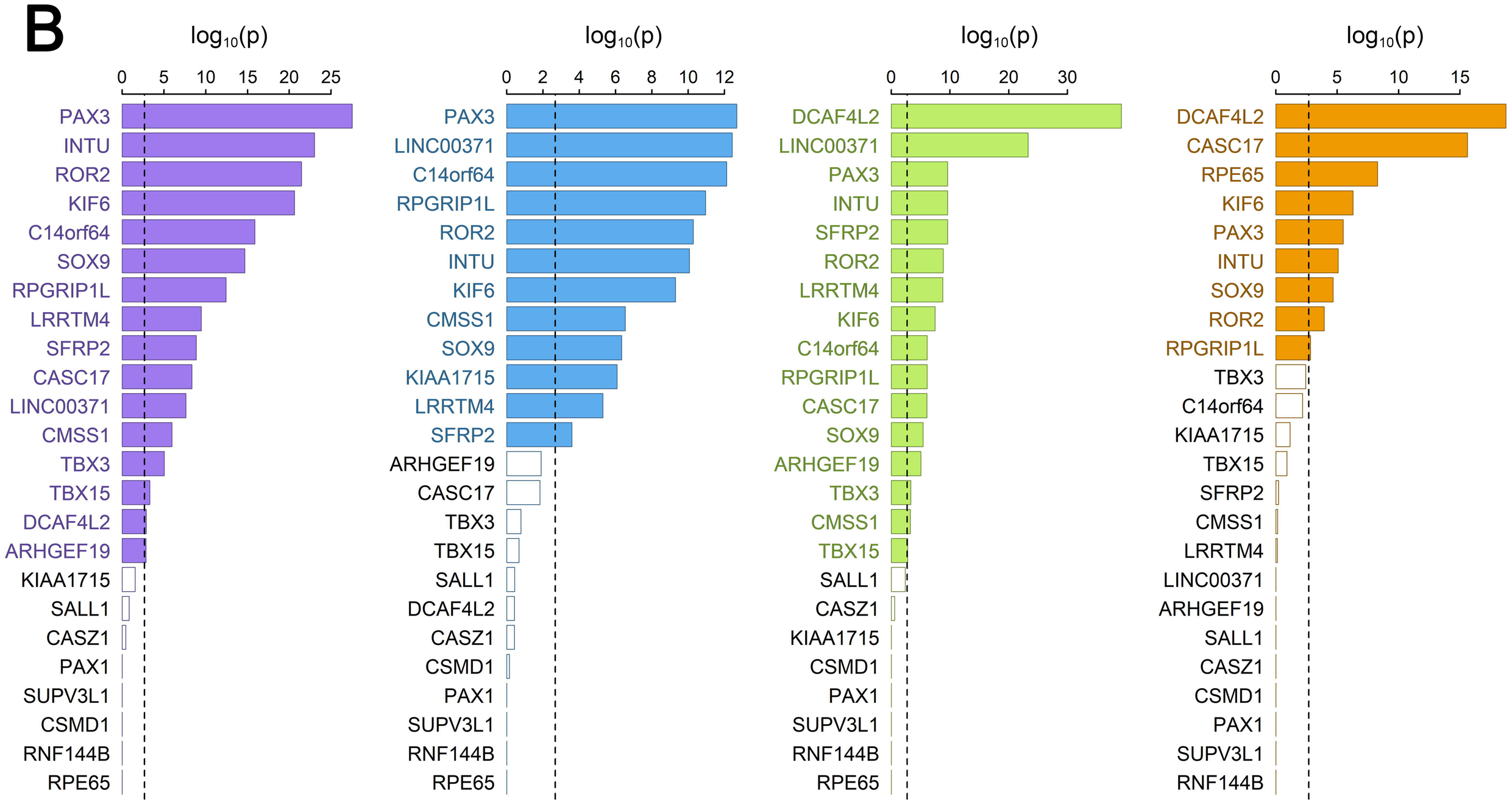
Background genes

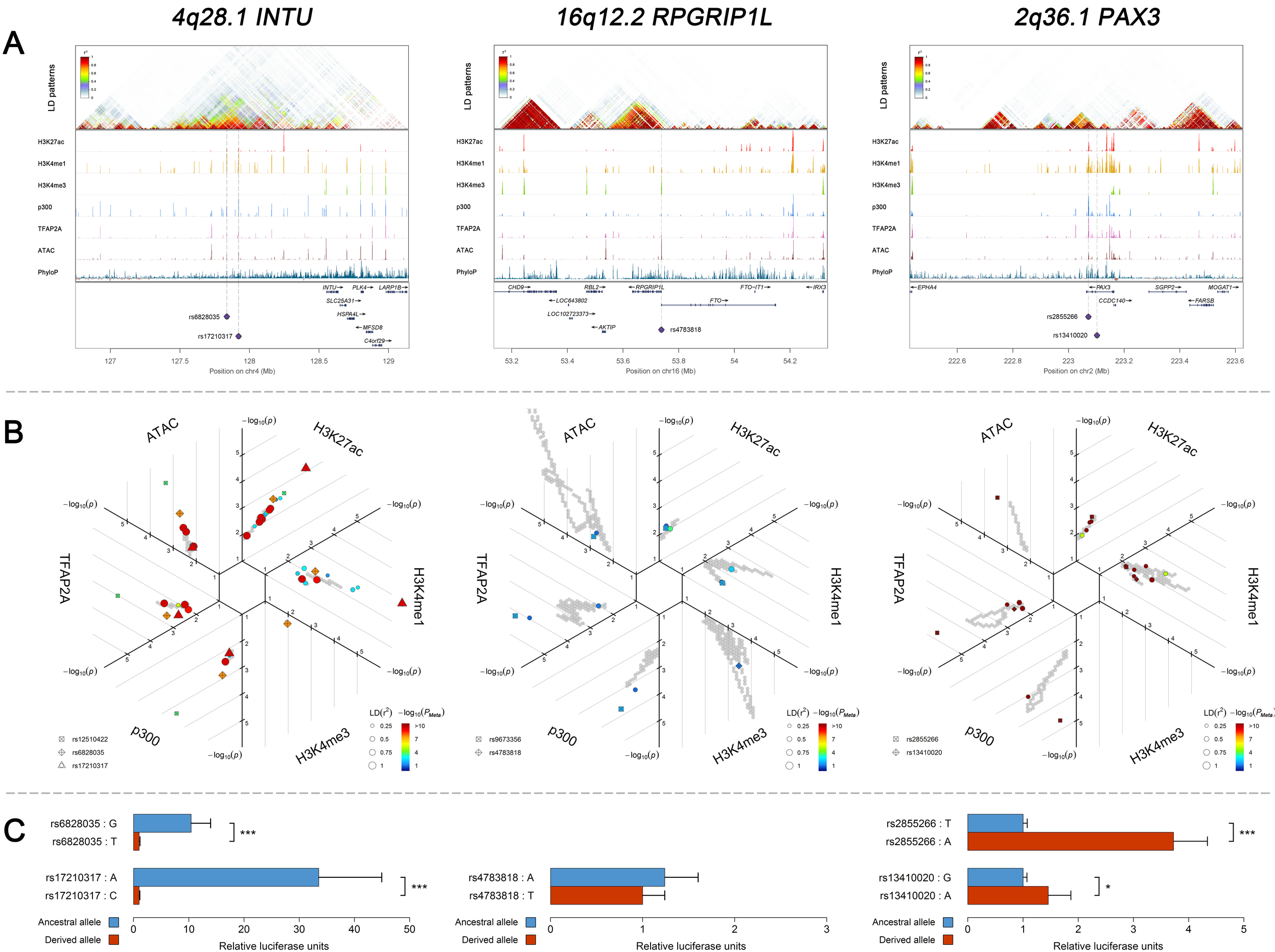
Tissue

Primary cell

Embryonic stem cell

B





4q28.1 INTU
16q12.2 RPGRIP1L
2q36.1 PAX3
A
LD patterns
H3K27ac
H3K4me1
H3K4me3
p300
TFAP2A
ATAC
PhyloP
INTU
PLK4
LARP1B
SLC25A31
HSPA4L
MFSD8
C4orf29
rs6828035
rs17210317
Position on chr4 (Mb)
CHD9
RBL2
RPGRIP1L
FTO-IT1
IRX3
LOC643802
LOC102723373
AKTIP
FTO
rs4783818
Position on chr16 (Mb)
EPHA4
PAX3
SGPP2
MOGAT1
CCDC140
FARSB
rs2855266
rs13410020
Position on chr2 (Mb)
B
ATAC
H3K27ac
TFAP2A
H3K4me1
p300
H3K4me3
$-\log_{10}(p)$
rs12510422
rs6828035
rs17210317
rs9673356
rs4783818
rs2855266
rs13410020
LD(r^2)
$-\log_{10}(P_{Meta})$
C
rs6828035 : G
rs6828035 : T
rs17210317 : A
rs17210317 : C
rs4783818 : A
rs4783818 : T
rs2855266 : T
rs2855266 : A
rs13410020 : G
rs13410020 : A
Ancestral allele
Derived allele
Relative luciferase units